BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 1, Number 3, May 1979

NONCONVEX MINIMIZATION PROBLEMS[1]

BY IVAR EKELAND

TABLE OF CONTENTS

I. The central result. The grandfather of it all is the celebrated 1961 theorem of Bishop and Phelps (see [7], [8]) that the set of continuous linear functionals on a Banach space E which attain their maximum on a prescribed closed convex bounded subset $X \subset E$ is norm-dense in E^*. The crux of the proof lies in introducing a certain convex cone in E, associating with it a partial ordering, and applying to the latter a transfinite induction argument (Zorn's lemma). This argument was later used in different settings by Brøndsted and Rockafellar (see [9]) and by F. Browder (see [11]). The various situations can be adequately summarized in a diagram:

This paper surveys my 1972 result on the variational principle and the various advances and applications which have been registered since. They stretch over a vast field of mathematics, from control theory to global analysis, and it is hoped that every mathematician will find something to enjoy. This should be possible, because the ideas involved are quite simple, and I have tried not to obscure them by too technical or detailed an exposition; the reader will be referred to the original papers for the more peripheral lemmas; received by the editors September 1, 1978.

AMS (MOS) subject classifications (1970). Primary 26A54, 26A96, 34H05, 35K55, 46B99, 47H10, 47H15, 49A10, 49B10, 58B20, 58C20, 93C15.

Key words and phrases. Fréchet derivative, variational problem, necessary conditions for optimality, convex functions, fixed point theorem, nonlinear semigroup, Pontryagin maximum principle, mean-value theorem, Asplund space, Hopf-Rinow theorem.

[1]Research supported by grants 67-9082 and 67-3990 from the National Research Council of Canada.

0002-9904/79/0000-0200/$08.75

	closed subsets	l.s.c. functions
convex	Bishop-Phelps	Brøndsted-Rockafellar
nonconvex	F. Browder	*

Note that the lower right-hand corner is empty. It was filled out in [19], [20], by adapting the Bishop-Phelps argument to lower semicontinuous nonconvex functions; as was pointed out to me in [27], transfinite induction is no longer needed. Here comes the result, with the most concise proof to date (from [4]):

THEOREM 1. *Let V be a complete metric space, and $F\colon V \to \mathbf{R} \cup \{+\infty\}$ a l.s.c. function, $\not\equiv +\infty$, bounded from below. Let $\varepsilon > 0$ be given, and a point $u \in V$ such that*

$$F(u) \leqslant \inf_V F + \varepsilon.$$

Then there exists some point $v \in V$ such that

$$F(v) \leqslant F(u), \tag{1}$$

$$d(u, v) \leqslant 1, \tag{2}$$

$$\forall w \neq v, \quad F(w) > F(v) - \varepsilon d(v, w). \quad \square \tag{3}$$

PROOF. Let us define inductively a sequence u_n, $n \in \mathbf{N}$, starting with $u_0 = u$. Suppose $u_n \in V$ is known. Now either:

(a) $\forall w \neq u_n$, $F(w) > F(u_n) - \varepsilon d(u_n, w)$. Then set $u_{n+1} = u_n$.

(b) $\exists w \neq u_n$: $F(w) \leqslant F(u_n) - \varepsilon d(u_n, w)$. Let S_n be the set of all such $w \in V$. Then choose $u_{n+1} \in S_n$ such that

$$F(u_{n+1}) - \inf_{S_n} F \leqslant \tfrac{1}{2}\Big[F(u_n) - \inf_{S_n} F\Big]. \tag{4}$$

I claim this is a Cauchy sequence. Indeed, if case (a) ever occurs, it is stationary, and if not, we have the inequalities

$$\varepsilon d(u_n, u_{n+1}) \leqslant F(u_n) - F(u_{n+1}), \quad \text{all } n \in \mathbf{N}.$$

Adding them up, we get

$$\varepsilon d(u_n, u_p) \leqslant F(u_n) - F(u_p), \quad \text{all } n \leqslant p. \tag{5}$$

The sequence $F(u_n)$ is decreasing and bounded from below (by $\inf_V F$), hence convergent, so the right-hand side goes to zero with (n, p) and (u_n) is a Cauchy sequence. Since the space V is complete, u_n converges to some $v \in V$.

I claim v satisfies properties (1), (2) and (3). Inequality (1) proceeds from the string of inequalities

$$F(u) \geqslant F(u_1) \geqslant \cdots \geqslant F(u_n) \geqslant F(u_{n+1}) \geqslant \cdots$$

and the fact that F is lower semicontinuous:

$$F(v) \leqslant \lim_n F(u_n).$$

Inequality (2) comes from taking $n = 0$ in (5):

$$\varepsilon d(u, u_p) \leqslant F(u) - F(u_p) \leqslant F(u) - \inf_V F$$
$$\leqslant \varepsilon \quad \text{by assumption.}$$

Letting $p \to \infty$, we get $d(u, v) \leqslant 1$.

If inequality (3) were not true, there would be some $w \neq v$ with $F(w) \leqslant F(v) - \varepsilon d(v, w)$. Letting $p \to \infty$ in equation (5), we get

$$F(w) \leqslant \lim F(u_p) \leqslant F(u_n) - \varepsilon d(u_n, w),$$

and hence $w \in S_n$ for all n. But relation (4) can be written as

$$2F(u_{n+1}) - F(u_n) \leqslant \inf_{S_n} F \leqslant F(w).$$

When $n \to \infty$, $F(u_n) \to l$, and this becomes $l \leqslant F(w)$. Since F is l.s.c., we also have $F(v) \leqslant l$. Finally, we get the inequality $F(v) \leqslant F(w)$, contradicting the definition of w. □

The Bishop-Phelps ordering argument may not be apparent any more, although it still underlies this proof. The reader who wants it spelt out is referred to the original proof (see [20] or [24]).

We immediately draw the obvious corollary:

THEOREM 1 *bis. Let V be a complete metric space, and F: $V \to \mathbf{R} \cup \{+\infty\}$ a l.s.c. function, $\not\equiv +\infty$, bounded from below. For any $\varepsilon > 0$, there is some point $v \in V$ with:*

$$F(v) \leqslant \inf_V F + \varepsilon, \tag{6}$$

$$\forall w \in V, F(w) \geqslant F(v) - \varepsilon d(v, w). \quad \square \tag{7}$$

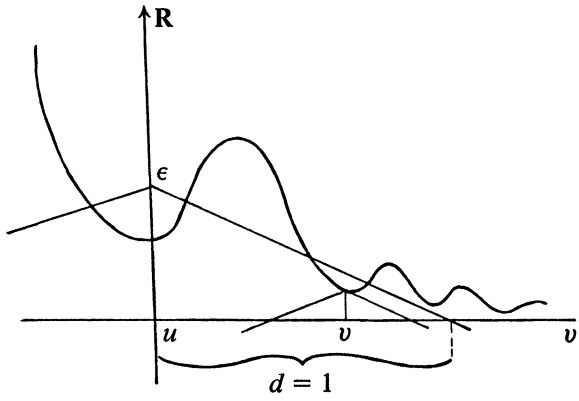
This relies on the fact that there always is some point u with $F(u) \leqslant \inf F + \varepsilon$. Inequality (6) then proceeds from (1) and (7) from (3).

Theorem 1 certainly is stronger than Theorem 1 bis. The main difference lies in inequality (2), which gives the whereabouts of point v in V, and which has no counterpart in Theorem 1 bis. In the sequel, we will refer to Theorem 1 as the strong statement, and to Theorem 1 bis as the weak statement. There is also a local statement, which is due to myself and Lebourg [23], and which will be stated as Theorem 1 ter. It starts with a definition:

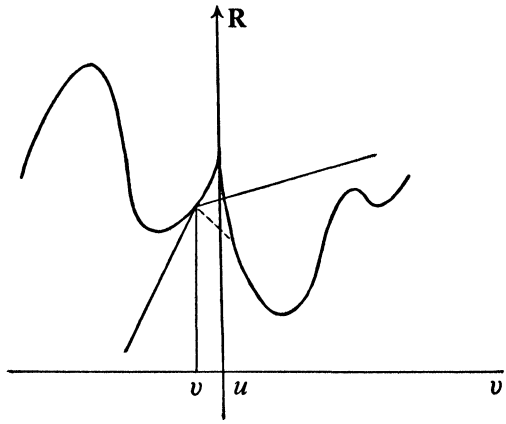
DEFINITION. Let V be a Banach space, and F a function from V to $\mathbf{R} \cup \{+\infty\}$, with $\operatorname{dom} F = \{v \in F | F(v) < +\infty\}$. It will be called ε-supported at v, with $\varepsilon > 0$ and $v \in \operatorname{dom} F$, if there exists $v^* \in V^*$ and $\eta > 0$ such that

$$\|w - v\| \leqslant \eta \Rightarrow F(w) \geqslant F(v) + \langle v^*, w - v \rangle - \varepsilon \|w - v\|. \quad \square \tag{8}$$

This can be regarded as a very weak and one-sided kind of differentiability; in fact, it can be proved (see [23]) that if both F and $-F$ are ε-supported at v for all $\varepsilon > 0$, then F is Fréchet-differentiable at v. On the other hand, it can also be regarded as a local and slanted version of inequality (7): it now holds only in some neighbourhood of v, and the linear term $\langle v^*, w - v \rangle$ tilts the vertical cone $\{(w, a) | a \geqslant F(v) - \varepsilon \|w - v\|\}$ in $V \times \mathbf{R}$.



THEOREM 1



THEOREM 1 ter

THEOREM 1 *ter. Let V be a Banach space, and assume there is a Fréchet-differentiable function ϕ: $V \to \mathbf{R}$ with $\phi(0) > 0$, and $\phi(w) \leqslant 0$ outside the unit ball. Let F: $V \to \mathbf{R} \cup \{+\infty\}$ be l.s.c. Then, for any $\varepsilon > 0$, the set of points where F is ε-supported is dense in* dom F. □

PROOF. Let $u \in V$ be given, and some $\alpha > 0$. We have to find in the ball $B(u; \alpha)$ some point v where F is ε-supported. Since F is l.s.c., we can assume that α has been chosen so small that F is bounded from below on $B(u; \alpha)$.

Define $\psi(w) = 1/\max[0, \phi((w - u)/\alpha)]$. It is clear that $u \in \text{dom}\, \psi \subset B(u; \alpha)$, and that ψ is Fréchet-differentiable on dom ψ. Applying Theorem 1 bis to $\psi + F$, we get a point v such that

$$\forall w \in V, \quad \psi(w) + F(w) \geqslant \psi(v) + F(v) - \varepsilon\|w - v\|.$$

It follows immediately that $\psi(v) < \infty$ (hence $v \in B(u; \alpha)$) and that $F(v) < \infty$ (hence $v \in \text{dom}\, F$). Moreover, since ψ is Fréchet-differentiable at v, there is some $v^* = -\psi'(v) \in V^*$ and some $\eta > 0$ such that

$$\|w - v\| \leqslant \eta \Rightarrow \psi(v) - \langle v^*, w - v\rangle + \varepsilon\|w - v\| \geqslant \psi(w).$$

Adding the last two inequalities, we get the desired result (8), with ε replaced by 2ε. □

Not all Banach spaces allow such a function ϕ. Hilbert spaces do (take $\phi(u) = 1 - (u, u)$), and L^p spaces for $1 < p < \infty$ (take $\phi(u) = 1 - \|u\|^p$), but L^1 and L^∞ do not. We refer to [23] for details.

The significance of these various statements will become clear as we proceed with the applications, which take up the rest of the paper. I have classified them according to which statement they use: weak (Theorem 1 bis), strong (Theorem 1), or local (Theorem 1 ter). Note that this does not coincide with the chronological order.

II. The weak statement.

A. FIXED-POINT THEOREMS.

A1. *J. Caristi.* In his 1976 paper [12], he proved a fixed point theorem which has aroused a considerable amount of interest, since it requires no continuity of the mapping under consideration.

THEOREM 2. [12]. *Let V be a complete metric space and f a self-map $V \to V$ satisfying*

$$\forall u \in V, \quad d(u, f(u)) \leqslant \phi(u) - \phi(f(u)) \tag{9}$$

where ϕ: $V \to \mathbf{R}$ is a prescribed l.s.c. function. Then f has a fixed point:

$$\exists v \in V: f(v) = v. \quad \square$$

PROOF. Apply Theorem 1 bis to the function ϕ, taking $\varepsilon = \frac{1}{2}$. We get some point $v \in V$ such that

$$\forall w \in V, \quad \phi(w) \geqslant \phi(v) - \tfrac{1}{2} d(v, w).$$

This should hold also for $w = f(v)$, yielding

$$\phi(v) - \phi(f(v)) \leqslant \tfrac{1}{2} d(v, f(v)).$$

Writing $u = v$ in inequality (9), we get

$$d(v, f(v)) \leqslant \phi(v) - \phi(f(v)).$$

Comparing the last two inequalities, we get $d(v, f(v)) \leqslant \frac{1}{2} d(v, f(v))$, and hence $d(v, f(v)) = 0$. □

This argument readily extends to multi-valued mappings [30]: if we associate with every $u \in V$ a closed nonempty subset $\Gamma(u) \subset V$ in such a way that

$$\forall u \in V, \forall w \in \Gamma(u), \quad d(u, w) \leqslant \phi(u) - \phi(w), \tag{10}$$

then some point $v \in V$ can be found with $\Gamma(v) = \{v\}$.

These results are best understood in the framework of dynamical systems. The l.s.c. function ϕ is called the entropy, and the system is seen as going from state x_n at time n to state $x_{n+1} = f(x_n)$ (or $x_{n+1} \in \Gamma(x_n)$) at time $n + 1$ (see [2]). Inequality (9), or (10), implies that $\phi(x_n) > \phi(x_{n+1})$ unless $x_n = x_{n+1}$: the entropy decreases until the system reaches a stationary state. Such systems are "dissipative" in a very strong way.

A2. *F. Clarke.* Aside from his main contributions, which presently will be dealt with, he proved in 1976 a fixed point theorem of another kind. It is an

extension of Banach's contraction principle; for the sake of convenience, it will be stated for closed convex subsets of Banach spaces. The self-map f: $V \to V$ is assumed not only to be continuous, but also to be a weak directional contraction in the following sense:

$$\exists \sigma < 1: \forall u \in V, \exists t \in (0, 1]: \|f(u_t) - f(u)\| \leqslant \sigma \|u_t - u\| \qquad (11)$$

where $u_t = tf(u) + (1 - t)u$ describes the line segment from u to $f(u)$ as t runs from 0 to 1.

THEOREM 3 [18]. *Let V be a closed convex subset of a Banach space. Any continuous map f: $V \to V$ satisfying condition* (11) *has a fixed point*

$$\exists v \in V: f(v) = v. \quad \square$$

PROOF. Apply Theorem 1 bis to the function $F(w) = \|w - f(w)\|$ with $0 < \varepsilon < 1 - \sigma$. Since f is continuous, so is F. We get some point $v \in V$ such that

$$\forall w \in V, \quad \|w - f(w)\| \geqslant \|v - f(v)\| - \varepsilon \|v - w\|.$$

It is assumed that some $t \in (0, 1]$ can be found with $\|f(v_t) - f(v)\| \leqslant \sigma t \|f(v) - v\|$. Writing $w = v_t$ into the preceding inequality, we get

$$\begin{aligned} \|v - f(v)\| &\leqslant \|v_t - f(v_t)\| + \varepsilon \|v - v_t\| \\ &\leqslant \|v_t - f(v)\| + \|f(v) - f(v_t)\| + \varepsilon \|v - v_t\| \\ &\leqslant \|v_t - f(v)\| + \sigma t \|f(v) - v\| + \varepsilon t \|f(v) - v\|. \end{aligned}$$

Moreover, since v_t belongs to the line segment $[v, f(v)]$, we have

$$\begin{aligned} \|v - f(v)\| &= \|v - v_t\| + \|v_t - f(v)\| \\ &= t\|f(v) - v\| + \|v_t - f(v)\|. \end{aligned}$$

The preceding inequality thus boils down to

$$t\|f(v) - v\| \leqslant (\sigma + \varepsilon) t \|f(v) - v\|, \quad \text{with } t > 0.$$

Dividing by t, we get $\|f(v) - v\| \leqslant (\sigma + \varepsilon)\|f(v) - v\|$. Since $\sigma + \varepsilon < 1$, this means that $f(v) = v$, and the result is proved. $\square$

The reader will note that this proof extends to multi-valued mappings Γ: $V \to V$, with *ad hoc* assumptions. He will note also that if condition (11) holds with $t = 1$, the sequence $f^n(u)$ is Cauchy, and converges to some fixed point, as in the traditional proof of existence.

On the other hand, note that neither Theorem 2 nor Theorem 3 guarantee uniqueness; indeed, the former applies to the identity map of V, and as to the latter, a counter-example is provided in [18].

B. NONLINEAR SEMIGROUPS.

H. Brézis and F. Browder have extended Theorem 1 bis to a general principle on ordered sets (see [10]):

THEOREM 4 [10]. *Let $(X, \gtrsim)$ be an ordered set such that any totally ordered sequence in X has a minorant*:

$$x_n \gtrsim x_{n+1} \Rightarrow \exists y \in X: x_n \gtrsim y, \quad \textit{all } n \in \mathbf{N}. \qquad (12)$$

Let $\phi: X \to \mathbf{R} \cup \{+\infty\}$ *be an increasing function on* X, *bounded from below*:

$$x \gtrsim y \Rightarrow \phi(x) \geqslant \phi(y), \tag{13}$$

$$\exists C: \phi(x) \geqslant C, \quad \text{all } x \in X. \tag{14}$$

Then there exists $y \in X$ *such that*

$$y \gtrsim z \Rightarrow \phi(y) = \phi(z). \quad \square \tag{15}$$

PROOF. Start from any point x_0, and build a sequence (x_n) by induction as follows. If x_n is known, either it satisfies property (15), or it does not. In the first case we are done; in the latter, since ϕ is nondecreasing, we can find x_{n+1} such that

$$x_{n+1} \in S_n = \{x \in X | x_n \gtrsim x\},$$

$$\phi(x_{n+1}) - \inf_{S_n} \phi \leqslant \tfrac{1}{2}\Big[\phi(x_n) - \inf_{S_n} \phi\Big]. \tag{16}$$

Using assumption (12), we find some point y with $x_n \gtrsim y$ for all n. The sequence $\phi(x_n)$ is decreasing and bounded from below, hence convergent, and $\phi(y) \leqslant \lim \phi(x_n)$ by (13). I claim that the point y satisfies (15).

If it did not, there would be some point $z \in X$ with $y \gtrsim z$ and $\phi(y) > \phi(z)$. Then z would belong to all the S_n, and we would write (16) as

$$\phi(x_{n+1}) - \tfrac{1}{2}\phi(x_n) \leqslant \tfrac{1}{2}\phi(z).$$

Letting $n \to \infty$ yields $\lim \phi(x_n) \leqslant \phi(z)$, which contradicts $\phi(z) < \phi(y) \leqslant \lim \phi(x_n)$. $\square$

Theorem 4 can be regarded as a constructive version of Zorn's lemma: condition (12) then means that $(X, \gtrsim)$ is "countably inductive". Theorem 1 is the special case where $X = V$, $\phi = F$, and

$$u \gtrsim v \Leftrightarrow F(u) - \varepsilon d(u, v) \geqslant F(v).$$

Condition (12) has to be checked, and the conclusion (15) is that y is minimal: $y \gtrsim z \Rightarrow \phi(y) = \phi(z) \Rightarrow y = z$.

The originality of the Brézis-Browder result is twofold: it enables us to use other ordering relations than the preceding one; and it can be used in reverse, to show that some function ϕ on an ordered space is not bounded from below. For instance, we have:

COROLLARY 5 **[10]**. *Let* $(X, \gtrsim)$ *be an ordered set and* $\phi: X \to \mathbf{R}$ *an increasing function such that*:

(16) *for any decreasing sequence* (x_n) *with* $\phi(x_n)$ *bounded from below, there exists some* y *such that* $x_n \gtrsim y$ *for all* n *and* $\phi(x_n) \to \phi(y)$.

(17) *for any* $x \in X$ *and any* $\varepsilon > 0$, *there is some* $z \in X$ *such that* $x \gtrsim z$ *and* $\phi(z) \geqslant \phi(x) - \varepsilon$.

Then, for each $x \in X$, *the set* $S(x) = \{y \in X | x \gtrsim y\}$ *is mapped by* ϕ *onto* $(-\infty, \phi(x)]$. $\square$

PROOF. Take any $a < \phi(x)$, and consider the set

$$X_a = \{z \in S(x) | \phi(z) \geqslant a\}.$$

Condition (12) holds on $(X_a, \gtrsim)$; indeed, let (x_n) be a decreasing sequence

in X_a. Then the constant a bounds all the $\phi(x_n)$ from below; by assumption (16), there exists $y \in X$ such that $x_n \gtrsim y$, and $\phi(y) = \lim \phi(x_n) \geqslant a$, so that y also belongs to X_a.

Applying Theorem 4, we get a point $y \in X_a$ such that $y \gtrsim z$ and $z \in X_a$ imply $\phi(y) = \phi(z)$. Now if $\phi(y) > a$, assumption (17) would enable us to find some $z \in X$ with $y \gtrsim z$ and $\phi(z) \in (\phi(y), a)$, hence $z \in X_a$. This is a contradiction, which proves that y solves the equation $\phi(y) = a$. □

Now for the applications. Let V be a complete metric space and S a (nonlinear) semigroup on V. In other words, for each $t \geqslant 0$ we have a self-map $S(t)$: $V \to V$, with $S(0) = \mathrm{Id}$ and $S(t_1 + t_2) = S(t_1) \circ S(t_2)$ for all t_1, $t_2 \geqslant 0$. It is assumed that S is a continuous semigroup of contractions:

$$\forall u \in V, \quad \text{the map } t \to S(t)u \text{ is continuous on } [0, \infty) \tag{18}$$

$$\forall t \geqslant 0, \forall u, v \in V, \quad d(S(t)u, S(t)v) \leqslant d(u, v). \tag{19}$$

Brézis and Browder associate with S and any $L \geqslant 0$ an ordering on $V \times [0, \infty)$. It is defined as follows:

$$(u, a) \gtrsim (v, b) \Leftrightarrow b \geqslant a \text{ and } d(S(b - a)u, v) \leqslant L(b - a). \tag{20}$$

Imagine in V a bubble whose center lies at u at time a and moves along with the semigroup $S(t)$ and whose radius is zero at time a and grows at the rate L. Relation (20) simply means that this bubble contains v at time b. It is easy to check that this relation is reflexive and antisymmetric. It is also transitive; indeed, if $(u, a) \gtrsim (v, b) \gtrsim (w, c)$, we have

$$\begin{aligned} d(S(c - a)u, w) &\leqslant d(S(c - a)u, S(c - b)v) + d(S(c - b)v, w) \\ &= d(S(c - b) \circ S(b - a)u, S(c - b)v) + d(S(c - b)v, w) \\ &\leqslant d(S(b - a)u, v) + d(S(c - b)v, w) \\ &\leqslant L(b - a) + L(c - b) = L(c - a). \end{aligned}$$

The function ϕ: $V \times [0, \infty) \to \mathbf{R}$ is defined by

$$\phi(u, a) = -a.$$

It clearly is increasing. We will apply Theorem 4 and Corollary 5 on product sets $X = F \times [0, \infty)$, with F a closed subset of V. For this, we shall want to know that $\gtrsim$ and ϕ satisfy condition (16) on X, which is true, but nontrivial; we refer to Lemma 1 of **[10]** for the proof. We now state:

THEOREM 6 **[10]**. *Let F be a closed subset of V and $C \geqslant 0$ be such that*

$$\liminf_{t \to 0} \frac{1}{t} d(S(t)u, F) \leqslant C, \quad \textit{all } u \in F. \tag{21}$$

Then $d(S(t)u, F) \leqslant Ct$ for all $u \in F$ and $t \geqslant 0$.

PROOF. Apply Corollary 5 to $X = F \times [0, \infty)$, with $\gtrsim$ and ϕ defined as above, and $L > C$. Condition (16) has been seen to hold. As for (17), pick any $(u, a) \in X$ and $\varepsilon > 0$. By assumption, some $t \in (0, \varepsilon)$ can be found with: $d(S(t)u, F) < Lt$. Taking any $v \in F$ with $d(S(t)u, v) < Lt$, and setting $b = a + t$, we get $(u, a) \gtrsim (v, b)$ and $\phi(v, b) > \phi(u, a) - \varepsilon$. This is condition (17).

It follows from Corollary 5 (setting $x = (u, 0)$) that for any $u \in F$ and $t \geqslant 0$ some $v \in F$ can be found with $(u, 0) \gtrsim (v, t)$, that is, $d(S(t)u, v) \leqslant Lt$. Taking the infimum over v in F, we get $d(S(t)u, F) \leqslant Lt$. Since this is to hold for all $L > C$, we get the desired result. □

Taking $C = 0$ yields Martin's theorem (see [29]): the set F is invariant under the semigroup S. Condition (21) then can be understood as stating that all trajectories leave F tangentially (if at all).

THEOREM 7 [10]. *Let F be a closed subset of V, and set*

$$c = \liminf_{t\to\infty} \frac{1}{t}\, d(S(t)u, F). \tag{22}$$

Then c does not depend on $u \in F$, and

$$\sup_{u\in F} \inf_{t>0} \frac{1}{t} d(S(t)u, F) = c. \square \tag{23}$$

PROOF. Since $S(t)$ is a contraction, we have

$$\left|d(S(t)u, F) - d(S(t)v, F)\right| \leqslant d(S(t)u, S(t)v) = d(u, v).$$

Dividing by t and letting $t \to \infty$, we see that c does not depend on u. It follows immediately from the definitions that

$$\sup_{u\in F} \inf_{t>0} \frac{1}{t} d(S(t)u, F) \leqslant c. \tag{24}$$

Call the left-hand side l. Take any $L > l$. On the set $X = F \times [0, \infty)$ we define the ordering $\gtrsim$ by (20) and the function $\phi(u, a) = -a$. I claim there is no $y \in X$ which satisfies condition (15).

Indeed, for any $(u, a) \in F \times [0, \infty)$, there is some $t > 0$ such that $d(S(t)u, F) < Lt$. As before, we take some $v \in F$ with $d(S(t)u, v) < Lt$, and setting $b = a + t$, we get $(u, a) \gtrsim (v, b)$ and $\phi(v, b) = \phi(u, a) - t$. Thus no point in X satisfies condition (15).

For any $(u, a) \in X$, set $(v, b) \in S(u, a)$ iff $(u, a) \gtrsim (v, b)$. Now if the function ϕ were bounded from below on the set $S(u, a)$, we could apply Theorem 4 to this set, and get a point (v, b) satisfying condition (15) in $S(u, a)$, and hence in X. Since this is impossible, the function ϕ is unbounded on all sets $S(u, a)$. In other words, for any $u \in F$, a sequence (v_n, t_n) can be found such that

$$v_n \in F, \quad t_n \to \infty \quad \text{and} \quad (u, 0) \gtrsim (v_n, t_n).$$

The latter means that $d(S(t_n)u, v_n) \leqslant Lt_n$. Since $v_n \in F$ and $t_n \to \infty$, this implies that

$$\liminf_{t\to\infty} \frac{1}{t} d(S(t)u, F) \leqslant L.$$

We have proved that $L > l$ implies $L \geqslant c$, and hence $l \geqslant c$, which concludes the proof. □

Theorems 6 and 7 are converse to each other in a certain sense. To gain a better understanding, let us imagine $S(t)u$ as the position at time t of an escapee fleeing a prison F which he has left at time 0 and point $u \in \text{bdy } F$.

The rate $d(S(t)u, F)/t$ will be called the mean escape speed; when $t \to 0$ we get the normal exit speed, and when $t \to \infty$ we get the asymptotic mean speed, which does not depend on the starting point anymore. Theorem 6 states that if all the normal exit speeds are less than C, then so are all the mean escape speeds (and hence the asymptotic mean speed). Theorem 7 states that if the asymptotic mean speed is c, for any $\varepsilon > 0$ some trajectory can be found along which the mean escape speed (and hence the normal exit speed) is always greater than $c - \varepsilon$. Hence the relationship $c \leqslant C$.

We refer to [**10**] for similar results when the contraction property (19) is replaced by

$$\exists \omega \in \mathbf{R}: d(S(t)u, S(t)v) \leqslant e^{\omega t} d(u, v), \quad \text{all } t \geqslant 0, u, v \in V.$$

C. Optimization and control.

This really was the motivation which led me to prove Theorem 1. Let us start with the following simple result. Recall that a function $F: V \to \mathbf{R}$ is called Gâteaux-differentiable if at every point $v \in V$ there is a continuous linear functional $F'(v) \in V^*$ such that

$$\forall u \in V, \quad \lim_{t \to 0_+} \frac{1}{t} \left[F(v + tu) - F(v) \right] = \langle F'(v), u \rangle.$$

Theorem 8 [**20**]. *Let V be a Banach space and F a Gâteaux-differentiable function on V, bounded from below. For any $\varepsilon > 0$, there is a point $v \in V$ such that*

$$F(v) \leqslant \inf_V F + \varepsilon, \tag{25}$$

$$\|F'(v)\|^* \leqslant \varepsilon. \square \tag{26}$$

Proof. Apply Theorem 1 bis to F: there is a point $v \in V$ such that

$$\forall w \in V, \quad F(w) \geqslant F(v) - \varepsilon \|w - v\|.$$

Take any $u \in V$, and set $w = v + tu$ in the preceding inequality, with $t > 0$. This yields

$$\frac{1}{t} \left[F(v + tu) - F(v) \right] \leqslant -\varepsilon \|u\|.$$

Letting $t \to 0_+$, we get $\langle F'(v), u \rangle \geqslant -\varepsilon \|u\|$. Taking the infimum of both sides over all $u \in V$ with $\|u\| = 1$, we get $-\|F'(v)\|^* \geqslant -\varepsilon$, the desired inequality. $\square$

Any point v which minimizes F over V satisfies (25) and (26) with $\varepsilon = 0$. On the other hand, there might not be any such point: the usual conditions ensuring the existence of a minimum are quite stringent (F should be convex, have bounded level sets, and V should be reflexive). What Theorem 8 does is, even in the absence of an exact minimum, to provide us with points which almost minimize F and almost satisfy the first-order necessary conditions. In other words, the equations $F(v) = \inf F$ and $F'(v) = 0$ can be satisfied to any prescribed accuracy $\varepsilon > 0$.

This obviously has wide applications to nonconvex variational problems, including nonlinear Euler-Lagrange equations. For instance, let Ω be a

bounded open subset of $\mathbf{R}^n$, and f: $\Omega \times \mathbf{R}^n \to [0, \infty)$ a C^1 function such that

$$\forall (x, \xi) \in \Omega \times \mathbf{R}^n, \quad |f'_\xi(x, \xi)| \leqslant a + b|\xi|^{p-1}$$

with a, b and $p \in [1, \infty)$ given constants. Then the functional

$$F(u) = \int_\Omega f(x, \operatorname{grad} u(x))\, dx$$

is positive Gâteaux-differentiable and well defined on the Banach space

$$W_0^{1,p}(\Omega) = \{u \in L^p(\Omega) \mid \partial u / \partial x_i \in L^p(\Omega),\ 1 \leqslant i \leqslant n,\ u = 0 \text{ on bdy } \Omega\}.$$

Denote by $W^{-1,p}(\Omega)$ the dual space of $W_0^{1,p}(\Omega)$, and apply Theorem 8: for any $\varepsilon > 0$, there is some $v \in W_0^{1,p}(\Omega)$ such that

$$\|\operatorname{div} f'_\xi(x, \operatorname{grad} v(v))\|_{W^{-1,p}} \leqslant \varepsilon.$$

The reader is referred to [20] or [24] for such situations. Here we will proceed several steps further, directly to control theory. Let me first recall the traditional setting of Pontryagin's minimum principle. Consider the differential equation

$$\begin{aligned} \frac{dx}{dt}(t) &= f(t, x(t), u(t)), \\ x(0) &= x_0 \in \mathbf{R}^n \end{aligned} \tag{27}$$

on the prescribed time interval $[0, T]$. The vector $x(t)$ describes the state of the system, $u(t)$ is the control at time t and belongs to some prescribed compact metrizable set K. The function f: $[0, T] \times \mathbf{R}^n \times K \to \mathbf{R}^n$ is continuous, and

(a) the $\partial f / \partial x_i$, $1 \leqslant i \leqslant n$, are well defined and continuous

(b) $\langle x, f(t, x, u)\rangle \leqslant c(1 + |x|^2)$ for some constant c.

Let a measurable control u: $[0, T] \to K$ be given. Condition (b) ensures that any solution to equation (27) extends to the whole of $[0, T]$, and condition (a) that any such solution is unique. So the endpoint $x(T)$ is well defined once the control u is chosen.

Let g: $\mathbf{R}^n \to \mathbf{R}$ be a differentiable function. The problem is to find some measurable control v such that the corresponding trajectory x minimizes $g(x(T))$ among all solutions of (27).

THEOREM 9 [20]. *For every $\varepsilon > 0$, there exists a measurable control v, the corresponding trajectory being y, such that*

$$g(y(T)) \leqslant \inf g(x(T)) + \varepsilon, \quad \textit{and} \tag{28}$$

$$\langle f(t, y(t), v(t)), p(t)\rangle \leqslant \min_{u \in K} \langle f(t, y(t), u), p(t)\rangle + \varepsilon \tag{29}$$

for almost every $t \in [0, T]$. Here p is the solution of the linear differential system:

$$\begin{aligned} \frac{dp_i}{dt}(t) &= -\sum_{j=1}^{n} \frac{\partial f_j}{\partial x_i}(t, y(t), v(t)) p_j(t), \\ p(T) &= g'(y(T)). \quad \square \end{aligned}$$

Whenever (28) holds with $\varepsilon = 0$ (i.e. when v is an optimal control), then (29) also holds with $\varepsilon = 0$ (i.e. v satisfies the first-order necessary conditions). This is essentially the statement of Pontryagin's minimum principle. However, there is no guarantee that such an optimal control will exist, except in special cases (see [**24**]), whereas a control v satisfying (28) and (29) can be found for any $\varepsilon > 0$.

The proof relies on Theorem 1 bis, but requires a little ingenuity. Define V as the set of all measurable controls u: $[0, T] \to K$, with the following metric:

$$d(u_1, u_2) = \text{meas}\{t \in [0, T] | u_1(t) \neq u_2(t)\}.$$

LEMMA 10. *V is a complete metric space.*

PROOF. Let us first check that d is a distance. For any u_1, u_2, u_3 in V, we have

$$\{t|u_1(t) \neq u_3(t)\} \subset \{t|u_1(t) \neq u_2(t)\} \cup \{t|u_2(t) \neq u_3(t)\}$$

and hence $d(u_1, u_3) \leqslant d(u_1, u_2) + d(u_2, u_3)$, the triangle inequality.

I claim that V is complete. Indeed, if (u_n) is a Cauchy sequence in V, a subsequence (u_{n_k}) can be found such that $d(u_{n_k}, u_{n_{k+1}}) \leqslant 2^{-k}$. Setting

$$A_k = \bigcup_{p \geqslant k} \{t|u_{n_p}(t) \neq u_{n_{p+1}}(t)\}$$

we have

$$\text{meas } A_k \leqslant \sum_{p=k}^{\infty} 2^{-p} = 2^{1-k}, \quad \text{and} \quad A_k \supset A_{k+1}.$$

Defining $\bar{u} \in V$ by $\bar{u}(t) = u_{n_k}(t)$ whenever $t \notin A_k$, we see that the subsequence (u_{n_k}) converges to $\bar{u}$. Since the full sequence (u_n) has the Cauchy property, it converges also. □

We now proceed with the proof of Theorem 9:

PROOF. Consider the function F: $u \to g(x(T))$, where $u \in V$, and x is the corresponding solution of (27). It is easily seen to be continuous and bounded from below (details in [**20**]). Applying Theorem 1 bis, we get some $v \in V$ such that

$$F(v) \leqslant \inf_V F + \varepsilon,$$

$$\forall u \in V, \quad F(u) \geqslant F(v) - \varepsilon d(u, v).$$

The first inequality is just (28). The second one will be analysed further, by taking into account the equation $dy/dt = f(t, y(t), v(t))$, which holds almost everywhere on $[0, T]$. Take any $t_0 > 0$ where it holds, any $k_0 \in K$, and define $u_\tau \in V$ for $\tau \geqslant 0$ as follows:

$$u_\tau(t) = k_0 \qquad \text{if } t \in [0, T] \cap (t_0 - \tau, t_0),$$

$$u_\tau(t) = v(t) \quad \text{if } t \notin [0, T] \cap (t_0 - \tau, t_0).$$

Clearly, $d(u_\tau, v) = \tau$ when τ is small enough. Denote by x_τ the trajectory associated with u_τ, so that $u_0 = v$ and $x_0 = y$. Writing $u = u_\tau$, and $F(u_\tau) = g(x_\tau(T))$, in the second inequality, we get

$$\forall \tau \geqslant 0, \quad g(x_\tau(T)) - g(y(T)) \geqslant -\varepsilon\tau$$

and hence $dg(x_\tau(T))/d\tau|_{\tau=0} \geqslant -\varepsilon$. But the left-hand side can be computed to be (see [**20**, Lemma 7.4]; it is a classical argument using the linearized equation and its resolvent), provided t_0 is a Lebesgue point for v:

$$\frac{d}{d\tau}\, g(x_\tau(T))\Big|_{\tau=0} = \langle f(t_0, y(t_0), k_0) - f(t_0, y(t_0), v(t_0)), p(t_0)\rangle.$$

Finally, we get

$$\langle f(t_0, y(t_0), k_0) - f(t_0, y(t_0), v(t_0)), p(t_0)\rangle \geqslant -\varepsilon,$$

which is the desired result, since k_0 is any point of K and t_0 almost any point of $[0, T]$. □

To conclude, let me point out that the pseudo-solutions given by Theorem 9, in the absence of any exact solution, may well be unsatisfactory. For instance, when ε is very small, they may become so irregular that they cannot be implemented in any practical way. Typically, when $\varepsilon \to 0$, the corresponding pseudo-solutions converge in a very large space $\mathcal{V} \supset V$; elements of $\mathcal{V}$ are called "relaxed" or "chattering" controls (see [**24**]).

D. Existence of solutions.

The setting is the same as in the preceding section, but this time we are interested in exact solutions, no longer in approximate ones, however good they may be. Sufficient conditions for the existence of minimizers are quite stringent; the most general one is the condition (C) of Palais and Smale, which is fundamental to global analysis (see [**33**], [**38**]). I state it for Banach spaces, although it really belongs to Banach manifolds:

Theorem [**20**]. *Let F be a Gâteaux-differentiable function on a Banach space V, bounded from below and satisfying the following condition*:

> *whenever $F'(u_n) \to 0$ in V^* and $F(u_n)$ is bounded, then either $F'(u_n) = 0$ for some n or the sequence u_n has a cluster point in V.*

Then the function F attains its minimum:

$$\exists v \in V\colon F(v) = \inf_{u \in V} F(u) \text{ and } F'(v) = 0. \square$$

Proof. By Theorem 1 bis, there is a sequence u_n in V such that $F'(u_n) \to 0$ and $F(u_n) \to \inf_V F$. Now, either we can extract a subsequence (still denoted by u_n) such that $F'(u_n) \neq 0$ for all n, or $F'(u_n) = 0$ for all but a finite number of n.

In the former case, the u_n have a cluster point v towards which a subsequence converges. By continuity, $F(v) = \lim_n F(u_n) = \inf_V F$, the desired result.

In the latter case, denote by S the set of points $v \in V$ where F' vanishes. For each n, there is a C^1 path in S connecting u_n to some point v_n on the boundary of S; by the mean value theorem, $F(u_n) = F(v_n)$. Since v_n is on the boundary of S, there is some point $w_n \notin S$ such that $|F(v_n) - F(w_n)| \leqslant 1/n$. Clearly $F'(w_n) \neq 0$ and $F(w_n) \to \inf_V F$: we are back in the preceding case. □

Another type of problem relates to the equation $F'(v) = v^*$. The following easy result shows that it can be solved for many right-hand sides, using very few assumptions on F.

THEOREM [**20**]. *Let F be a Gâteaux-differentiable function on a Banach space V, and ϕ a continuous function on $[0, \infty)$, such that*

$$\forall v \in V, \quad F(v) \geqslant \phi(\|v\|),$$
$$\phi(t)/t \to \infty \quad \text{when } t \to \infty.$$

Then $F'(V)$ is dense in V^*. □

PROOF. Let any $v^* \in V^*$, and $\varepsilon > 0$, be given. Consider the function $G(v) = F(v) - \langle v^*, v\rangle$; it is l.s.c. and bounded from below. By Theorem 1 bis, there is some point v in V such that

$$\|G'(v)\|^* = \|F'(v) - v^*\|^* \leqslant \varepsilon.$$

This is the desired result. □

III. The full statement. The results I am going to describe now rely on the full statement of Theorem 1; the estimate (2) is going to play a crucial role. Note that it can be made as precise as we need, by the simple device of endowing V with the new distance λd. For any $\lambda > 0$, this is a *bona fide* distance, and V is still complete, so that we can apply Theorem 1 to the metric space $(V, \lambda d)$. The assumptions are unchanged, so is relation (1), but inequalities (2) and (3) become

$$d(u, v) \leqslant 1/\lambda, \tag{30}$$
$$\forall w \neq v, \quad F(w) > F(v) - \varepsilon\lambda d(v, w). \tag{31}$$

In other words, there is a trade-off. The greater λ is, the better you know the position of v, but the less interesting it becomes: you can't both have v very close to u, and $F'(v)$ very close to zero. A conciliatory choice is $\lambda = \varepsilon^{-1/2}$. Let me restate Theorem 1 in that case:

COROLLARY 11. *Let V be a complete metric space, and F: $V \to \mathbf{R} \cup \{+\infty\}$ a l.s.c. function, $\not\equiv +\infty$. If $u \in V$ is such that*:

$$F(u) \leqslant \inf_V F + \varepsilon, \tag{32}$$

then there exists $v \in V$ such that

$$F(v) \leqslant F(u), \tag{33}$$
$$d(u, v) \leqslant \sqrt{\varepsilon}\,, \tag{34}$$
$$w \neq v, \quad F(w) > F(v) - \sqrt{\varepsilon}\; d(v, w). \quad \square \tag{35}$$

I now proceed to the applications.

A. MATHEMATICAL PROGRAMMING.

Frank Clarke has developed a powerful method for finding necessary conditions for optimality. It has evolved from two basic ingredients:

(i) the observation that inequality (35) can be written as:

$$\forall w \neq v, \quad F(w) + \sqrt{\varepsilon}\ d(v, w) > F(v), \tag{36}$$

and understood as saying that the point v minimizes a new function G, defined by $G(w) = F(w) + \sqrt{\varepsilon}\ d(v, w)$, which is a slight perturbation of the original function F.

(ii) the new calculus developed for locally Lipschitzian functions on Banach spaces (see [13], [17]). With any such function f and any point $x \in E$ is associated a weakly compact convex subset $\partial f(x)$ of E^*, the generalized gradient of f at x. It plays for locally Lipschitzian functions the same role as the gradient for continuously differentiable functions. Indeed, similar rules of calculus hold, and $\partial f(x)$ boils down to $\{f'(x)\}$ when f is C^1.

First I will show Clarke's method at work in a very simple setting; this kind of result foreshadows §IV.

PROPOSITION 12. *Let F be a l.s.c. function on a Banach space V. Assume that for every $\varepsilon > 0$, a Gâteaux-differentiable function F_ε can be found such that*

$$F_\varepsilon \leqslant F \quad \text{and} \quad \inf F_\varepsilon \geqslant \inf F - \varepsilon, \tag{36}$$

$$F_\varepsilon'(v) \to \phi(u) \quad \text{when } \varepsilon \to 0 \text{ and } v \to u. \tag{37}$$

Then any point $u \in V$ which minimizes F also satisfies the equation

$$\phi(u) = 0 \quad \text{in } V^*. \ \square \tag{38}$$

PROOF. Assume u minimizes F on V. It follows from inequalities (36) that

$$F_\varepsilon(u) \leqslant \inf_V F_\varepsilon + \varepsilon.$$

By Corollary 11, some point $v_\varepsilon \in V$ can be found such that $\|u - v_\varepsilon\| \leqslant \sqrt{\varepsilon}$, and v_ε minimizes the new functional $G_\varepsilon(w) = F_\varepsilon(w) + \sqrt{\varepsilon}\ \|v_\varepsilon - w\|$ over V. Now F_ε is differentiable at v_ε, and the function $w \to \|v_\varepsilon - w\|$, although not differentiable at v_ε, has $+1$ as a directional derivative in every direction. It follows easily that v_ε must satisfy the condition

$$F_\varepsilon'(v_\varepsilon) + \sqrt{\varepsilon}\ B^* \ni 0, \tag{39}$$

with B^* the unit ball of V^*. Now let ε go to zero; then v_ε converges to u (here is where the estimate $d(u, v_\varepsilon) \leqslant \sqrt{\varepsilon}$ is so crucial), and the left-hand side of (39) converges to $\phi(u)$ because of (37). Hence the result. $\square$

Note that I did not assume F itself to be differentiable; there is no equation $F'(u) = 0$, but $\phi(u) = 0$ takes its place. In other words, this procedure enables us to find first-order conditions, when there are none to be derived by standard differentiation procedures.

In real-life optimization and control, the situation is more complicated. Implementing this procedure will typically involve nondifferentiable (but locally Lipschitzian) functions, and require the corresponding generalized calculus. Clarke himself has used his method to get Fritz-John optimality conditions for nonconvex problems in mathematical programming (see [15]). We now turn to this piece of work.

Consider the problem of minimizing $f(x)$ subject to the following equality and inequality constraints:

$$g_i(x) = 0, \qquad 1 \leqslant i \leqslant r,$$
$$h_j(x) \leqslant 0, \qquad 1 \leqslant j \leqslant k,$$

where f, the g_i and the h_j are locally Lipschitzian functions on $\mathbf{R}^n$.

THEOREM [15]. *If $\bar{x} \in \mathbf{R}^n$ minimizes f subject to the preceding constraints, there exists real numbers λ, μ_i, ν_j, not all zero, with $\nu_j \geqslant 0$ all j, and $\lambda \geqslant 0$, such that*

$$\lambda \partial f(\bar{x}) + \sum_{i=1}^{r} \mu_i \partial g_i(\bar{x}) + \sum_{j=1}^{k} \nu_j \partial h_j(\bar{x}) \ni 0.$$

If all functions are continuously differentiable, this relation becomes an equation: $\lambda f'(\bar{x}) + \sum \mu_i g_i'(\bar{x}) + \sum \nu_j h_j'(\bar{x}) = 0$. If further regularity conditions are met (the $g_i'(\bar{x})$ and the $h_j'(\bar{x})$ for $h_j(\bar{x}) = 0$ linearly independent), then $\lambda \neq 0$; dividing by λ, we get the usual Lagrange multiplier rule. □

In this type of result, it is usually fairly easy to come up with a set of numbers λ, μ_i, ν_j, satisfying the above relation. The difficulty lies in proving that they are not all zero. This has to be done, of course otherwise this relation would become trivial and useless.

One way to start is to notice that $\bar{x}$ minimizes the function:

$$F(x) = \max_{i,j} \{ f(x) - f(\bar{x}), |g_i(x)|, h_j(x) \}$$

which is locally Lipschitzian (but not differentiable, even if all the functions involved in the maximization are C^1). Using generalized calculus, we write that 0 should belong to the generalized gradient of F at $\bar{x}$:

$$0 \in \partial F(\bar{x}).$$

The right-hand side can be seen to the convex hull of $\partial f(\bar{x})$, $\partial |g_i(\bar{x})|$, and $\partial h_j(\bar{x})$ for $h_j(\bar{x}) = 0$. This means that nonnegative numbers λ, μ_i, ν_j can be found so that

$$\lambda \partial f(\bar{x}) + \sum \mu_i \partial |g_i(\bar{x})| + \sum \nu_j \partial h_j(\bar{x}) = 0,$$
$$\lambda + \sum \mu_i + \sum \nu_j = 1.$$

It would seem that we have reached our goal. Alas, it's farther away than ever! The reason is that $\partial |g_i(\bar{x})|$ is to be expressed as $\bigcup_{|t| \leqslant 1} t \partial g_i(x)$, so that the relation we have just proved comes down as

$$\lambda \partial f(\bar{x}) + \sum \mu_i t_i \partial g_i(\bar{x}) + \sum \nu_j \partial h_j(\bar{x}) = 0,$$
$$\lambda + \sum \mu_i + \sum \nu_j = 1, \qquad -1 \leqslant t_i \leqslant 1.$$

The trivial solution now appears by setting $\lambda = 0$, $\nu_j = 0$, $\mu_i \neq 0$ but $t_i = 0$; we cannot avoid it this way. Something different must be done. It will be to consider the function

$$F_\varepsilon(x) = \max_{i,j} \{ f(x) - f(\bar{x}) + \varepsilon, |g_i(x)|, h_j(x) \}$$

for $\varepsilon > 0$. There is no reason why $\bar{x}$ should minimize F_ε anymore; but we

clearly have:

$$F_\varepsilon(\bar{x}) = \varepsilon \leqslant \inf_x F_\varepsilon(x) + \varepsilon.$$

Apply Corollary 11: there exists $x_\varepsilon \in \mathbf{R}^n$ such that $\|\bar{x} - x_\varepsilon\| \leqslant \sqrt{\varepsilon}$ and x_ε minimizes $G_\varepsilon(x) = F_\varepsilon(x) + \sqrt{\varepsilon}\,\|x - x_\varepsilon\|$ over $\mathbf{R}^n$. This implies

$$0 \in \partial G_\varepsilon(x_\varepsilon) = \partial F_\varepsilon(x_\varepsilon) + \sqrt{\varepsilon}\ B$$

with B the unit ball of $\mathbf{R}^n$. As above, $\partial F_\varepsilon(x_\varepsilon)$ is computed to be the convex hull of the generalized gradients at x_ε of some functions f, $|g_i|$, h_j–the ones whose value at x_ε is exactly $F_\varepsilon(x_\varepsilon)$. But $F_\varepsilon(x_\varepsilon) > 0$, otherwise x_ε would satisfy all the constraints and $f(x_\varepsilon)$ would be strictly less than $f(\bar{x})$, which is impossible since $\bar{x}$ is optimal. It follows that we will have to consider $\partial|g_i(x_\varepsilon)|$ only for those i which satisfy $g_i(x_\varepsilon) \neq 0$. But the map $t \mapsto |t|$ is C^1 on $\mathbf{R} \setminus \{0\}$, with derivative sgn t, so that $\partial|g_i(x_\varepsilon)| = [\operatorname{sgn} g_i(x_\varepsilon)]\partial g_i(x_\varepsilon)$ by the chain rule. Finally, we have

$$\lambda_\varepsilon \partial f(x_\varepsilon) + \sum \mu_{\varepsilon,i}[\operatorname{sgn} g_i(x_\varepsilon)]\partial g_i(x_\varepsilon) + \sum \nu_{\varepsilon,j}\partial h_j(x_\varepsilon) \ni 0,$$

$$\lambda_\varepsilon \geqslant 0, \quad \mu_{\varepsilon,i} \geqslant 0, \quad \nu_{\varepsilon,j} \geqslant 0, \quad \lambda_\varepsilon + \sum \mu_{\varepsilon,i} + \sum \nu_{\varepsilon,j} = 1.$$

We now let $\varepsilon \to 0$. By compactness, $(\lambda_\varepsilon, \mu_{\varepsilon,i}, \nu_{\varepsilon,j})$ has a cluster point in the $(k + r)$-dimensional simplex, say (λ, μ_i, ν_j). Let $u_\varepsilon \in \partial f(x_\varepsilon)$, $v_\varepsilon \in \partial g_i(x_\varepsilon)$, $w_\varepsilon \in \partial h_j(x_\varepsilon)$ be such that $\lambda_\varepsilon u_\varepsilon + \sum \mu_{\varepsilon,i} v_{\varepsilon,i} + \sum \nu_{\varepsilon,j} w_{\varepsilon,j} = 0$. From known properties of the generalized gradients it follows that $(u_\varepsilon, v_{\varepsilon,i}, w_{\varepsilon,j})$ has a cluster point (u, v_i, w_i) with $u \in \partial f(\bar{x})$, $v_i \in \partial g_i(\bar{x})$ and $w_j \in \partial h_j(\bar{x})$. In the limit, we have

$$\lambda u + \sum (\pm \mu_i) v_i + \sum \nu_j w_j = 0,$$

$$\lambda \geqslant 0, \quad \mu_i \geqslant 0, \quad \nu_j \geqslant 0, \quad \lambda + \sum \mu_i + \sum \nu_j = 1$$

which is the desired result.

B. CONTROL THEORY.

We have already met control theory in §II.C: recall Theorem 9 and its setting. In this case, the "easy" minimum principle states that if (28) holds with $\varepsilon = 0$ (i.e. the control v is optimal), then so does (29) for almost every time t. This is quite easy to prove, using elementary methods in differential equations (see [34]). But if we now add a constraint on the final state, that $x(T)$ should belong to a prescribed subset C of $\mathbf{R}^n$, and minimize only over such controls as satisfy this condition, then finding necessary conditions becomes an excruciatingly difficult task. This "hard" statement was given by Pontryagin et al (see [35]), but their proof left the main difficulty untouched. Subsequent proofs (see [34], [32]) have dealt with it by such devices as fixed-point theorems, and are quite complicated.

Clarke's proof certainly is the simplest; it relies only on the "easy" statement and Theorem 1. I cannot describe it without getting into technicalities, but I think the following (closely related) statement captures the essence of the argument. For the time being, the assumptions on f are as in §II.C.

PROPOSITION 13 [14]. *Consider the differential equation*

$$dx/dt = f(t, x(t), u(t)), \qquad x(0) = x_0, 0 \leqslant t \leqslant T, \tag{40}$$

and let $A \subset \mathbf{R}^n$ be the attainable set, i.e. the set of all feasible final states $x(T)$ (there exists a measurable control u: $[0, T] \to K$ such that equation (40) is satisfied).

If $x(T)$ belongs to the boundary of A, there exists a nonvanishing function p: $[0, T] \to \mathbf{R}^n$ such that

$$\frac{dp_i}{dt} = -\sum_{j=1}^{n} \frac{\partial f_j}{\partial x_i}(t, x(t), u(t))p_j(t), \tag{41}$$

$$\langle p(t), f(t, x(t), u(t))\rangle = \min_{k \in K} \langle p(t), f(t, x(t), k)\rangle. \quad \square \tag{42}$$

PROOF. Before we start, note that there are no regularity properties at all to be expected from A, except that it be compact and connected. Note also that if $x(T)$ minimizes over A some differentiable function g, then we simply apply the "easy" minimum principle, to get a function p satisfying (41), (42), and $p(T) = -g'(x(T))$. Provided $g'(x(T)) \neq 0$, this will ensure that p is nonvanishing, for it is the solution of a linear differential equation, and if it vanishes somewhere, it vanishes everywhere.

The case where $x(T)$ minimizes no function g on A such that $g'(x(T)) \neq 0$ is the tough one. To break it, Clarke uses the fact that $x(T)$ lies on the boundary of A: for any $\varepsilon > 0$, there is in $\mathbf{R}^k$ some point $\xi_\varepsilon \notin A$ such that $\|\xi_\varepsilon - x(T)\| \leqslant \varepsilon$. Consider the set V of all measurable controls u: $[0, T] \to K$, endowed with the same distance d as in Lemma 10:

$$d(u_1, u_2) = \text{meas}\{t \in [0, T] | u_1(t) \neq u_2(t)\}, \tag{43}$$

and the functional F_ε: $V \to \mathbf{R}$ defined by

$$F_\varepsilon(w) = \|\xi_\varepsilon - z(T)\|$$

where z is the trajectory associated with the control w.

We know that V is a complete metric space, and that F_ε is continuous. Moreover,

$$F_\varepsilon(x(T)) \leqslant \varepsilon \leqslant \inf_{w \in V} F_\varepsilon(z(T)) + \varepsilon.$$

The stage is set for Corollary 11: there is some measurable control $v_\varepsilon \in V$, the associated final state being $y_\varepsilon(T) \in A$, such that

$$\text{meas}\{t \in [0, T] | u(t) \neq v_\varepsilon(t)\} \leqslant \sqrt{\varepsilon}\,, \tag{44}$$

$$v_\varepsilon \text{ minimizes } G_\varepsilon(w) = \|\xi_\varepsilon - z(T)\| + \sqrt{\varepsilon}\ \text{meas}\{w \neq v_\varepsilon\}. \tag{45}$$

We now transform (45) into a *bona fide* control problem. Define a function f_0: $[0, T] \times K \to \mathbf{R}$ by

$$f_0(t, k) = \begin{cases} \sqrt{\varepsilon} & \text{if } k \neq v_\varepsilon(t), \\ 0 & \text{if } k = v_\varepsilon(t). \end{cases}$$

The following control problem:

$$\frac{dz}{dt} = f(t, z(t), w(t)), \qquad z(0) = x_0,$$
$$\frac{dz_0}{dt} = f_0(t, w(t)), \qquad z_0(0) = 0,$$
$$\text{minimize } \|\xi_\varepsilon - z(T)\| + z_0(T)$$

has v_ε as an optimal solution; this is exactly statement (45). We now use the "easy" minimum principle to give necessary conditions which the control v_ε satisfies: there exists a function $(p^\varepsilon, p_0^\varepsilon)$: $[0, T] \to \mathbf{R} \times \mathbf{R}^n$ such that

$$\frac{dp_i^\varepsilon}{dt} = -\sum_{i=1}^{n} \frac{\partial f_j}{\partial x_i}(t, y_\varepsilon(t), v_\varepsilon(t)) p_j^\varepsilon(t), \tag{46}$$

$$dp_0^\varepsilon/dt = 0, \tag{47}$$

$$p^\varepsilon(T) = (\xi_\varepsilon - y_\varepsilon(T))/\|\xi_\varepsilon - y_\varepsilon(T)\|, \tag{48}$$

$$p_0^\varepsilon(T) = 1, \tag{49}$$

$$\langle f(t, y_\varepsilon(t), v_\varepsilon(t)), p^\varepsilon(t)\rangle + f_0(t, v_\varepsilon(t)) p_0^\varepsilon(t)$$
$$= \min_{k \in K} \{\langle f(t, y_\varepsilon(t), k), p^\varepsilon(t)\rangle + f_0(t, k) p_0^\varepsilon(t)\}. \tag{50}$$

From (47) and (49), we get $p_0^\varepsilon(t) = 1$ for all t. Since ξ_ε does not belong to A, the final condition (48) implies that $\|p^\varepsilon(T)\| = 1$. Letting $\varepsilon \to 0$ in the system (46)–(50), and taking into account the equation $dy_\varepsilon/dt = f(t, y_\varepsilon(t), v_\varepsilon(t))$ and the relation (44), we get the system (40)–(42). Moreover, $\|p(T)\| = \|p^\varepsilon(T)\| = 1$, so that the function p indeed vanishes nowhere. □

Note a geometrical consequence: whenever $x(T)$ belongs to the boundary of the attainable set A, the intermediate velocities $dx(t)/dt$ belong to the boundary of the set $f(t, x(t), k)$ of feasible velocities. The main interest, however, lies elsewhere.

This proof will work with precious little assumptions on f, essentially that it be locally Lipschitzian with respect to x, and Borelian with respect to (t, u). It can be slightly adapted to prove the "hard" minimum principle, with the weakest assumptions to date:

THEOREM 14 **[14]**. *Let C_0 and C_1 be closed subsets of $\mathbf{R}^n$, and consider the control problem*:

$$dx/dt = f(t, x(t), u(t)), \qquad 0 \leqslant t \leqslant T,$$
$$x(0) \in C_0, \qquad x(T) \in C_1,$$
$$\textit{minimize } g(x(T)).$$

If the control u: $[0, T] \to K$ is optimal, there exists a function p: $[0, T] \to K$ such that

$$\frac{dp_i}{dt} = -\sum_{i=1}^{n} \frac{\partial f_j}{\partial x_i}(t, x(t), u(t)) p_j(t),$$

$p(0)$ is normal to C_0 at $x(0)$,

$p(T)$ is the projection of $g'(x(T))$ on the tangent cone to C_1 at $x(T)$,

$$\langle f(t, x(t), u(t)), p(t)\rangle = \min_{k \in K} \langle f(t, x(t), k), p(t)\rangle \quad a.e. \ \square$$

Ironically enough, this "hard" minimum principle is the same as the "easy" one, except for the boundary conditions on p. Note that these ensure that p nowhere vanishes (and hence that the last equation is meaningful) provided that $g'(x(T))$ does not belong to the normal cone to C_1 at $x(T)$. For the precise meaning of "normals" (always outwards) and "tangents" to a closed subset of $\mathbf{R}^n$, the reader is referred to [13].

The proof consists in transforming the final condition $x(T) \in C_1$ into an initial one by enlarging the control system to dimension $(2n + 1)$, and then applying a result similar to Proposition 13 when the initial condition $x(0) = x_0$ is replaced by $x(0) \in C_0$ (see [14]).

Clarke has adapted his argument to the case where the controlled differential equation

$$\frac{dx}{dt}(t) = f(t, x(t), u(t)), \qquad u(t) \in K,$$

is replaced by the differential inclusion

$$\frac{dx}{dt}(t) \in E(t, x(t)),$$

where $E(t, x)$ is a nonempty compact subset of $\mathbf{R}^n$, depending measurably on t and continuously x. Obviously, the first formulation falls within the second (simply set $f(t, x, K) = E(t, x)$). On the other hand, differential inclusions cannot always be restated as controlled differential equations (except in the case when $E(t, x)$ is convex; see [39]). So the second formulation is the more general.

In the paper [16], Proposition 13 and Theorem 14 are restated and reproved in this new framework, yielding an improved version of the minimum principle. The method is essentially the same as before, and I will not go any further into it.

C. CALCULUS.

If V is a Banach space and f: $V \to \mathbf{R}$ is locally Lipschitzian, recall that $\partial f(u) \subset V^*$ is the generalized gradient of f at u. In the following, S denotes a closed subset of V, $N_S(u)$ the (outward) normal cone to S at u, and $d(u, S)$ the distance from u to S. They are related by

$$N_S(u) = \bigcup_{\lambda > 0} \lambda \partial d(u, S), \quad \text{all } u \in S.$$

A. Ioffe has proved the following very interesting result:

THEOREM 15. *Let E be the closed subset of S defined by the equation $f(w) = 0$. Then, for any $u \in S \setminus E$ and any $a \in (0, 1]$, there exists some $v \in S$ such that*

$$\|v - u\| \leqslant a d(u, E),$$

$$\exists v_1^* \in \partial |f(v)| \text{ and } v_2^* \in N_S(v) \text{ with } \|v_1^* + v_2^*\|^* \leqslant |f(u)|/a d(u, E).$$

PROOF. Consider the function $F: V \to \mathbf{R} \cup \{+\infty\}$ defined by

$$F(x) = \begin{cases} |f(x)| & \text{if } x \in S, \\ +\infty & \text{if } x \notin S. \end{cases}$$

It is l.s.c. since f is continuous and S is closed. We then apply Theorem 1, using on V the new norm $|||w||| = \|w\|/ad(u, E)$ and setting $\varepsilon = |f(u)|$. In other words, we use formula (30) and (31) with $\lambda = ad(u, E)$ and $\varepsilon = |f(u)|$. We get a point v such that:

$$\|v - u\| \leqslant ad(u, E),$$

$$\forall w \in V, \quad F(w) \geqslant F(v) - \|v - w\| \, |f(u)|/ad(u, E).$$

The last inequality implies that $F(v)$ is finite, so that $v \in S$ and $|f(v)| = F(v)$ (simply take $w \in S$, so that the left-hand side is finite). We write it again as

$$\forall w \in S, \quad |f(w)| + \|v - w\| \, |f(u)|/ad(u, E) \geqslant |f(v)|.$$

The left-hand side $g(w)$ attains its minimum on S for $w = v$. In this setting, the Lagrange multiplier rule becomes: $\partial g(v) + N_S(v) \ni 0$, which means that:

$$\partial |f(v)| + N_S(v) + B^* \frac{|f(u)|}{ad(u, E)} \ni 0$$

where B^* is the unit ball in V^*. This is the desired result. □

I understand Ioffe's theorem as an n-dimensional version of the mean value theorem. To see why, consider the following corollary:

COROLLARY 16. *Let g be a C^1 function on $\mathbf{R}^n$. For any point u in $\mathbf{R}^n$, denote by $E[u]$ the level set $\{w \mid g(w) = g(u)\}$. Then, for any two points u_1 and u_2, there is a point v such that*

$$|u_1 - v| \leqslant d(u_1, E[u_2]),$$

$$\|g'(v)\| \leqslant \frac{|g(u_1) - g(u_2)|}{d(u_1, E[u_2])}. \quad \square$$

PROOF. Simply apply Theorem 15 to $f(u) = g(u) - g(u_2)$ and $S = \mathbf{R}^n$. For any $a \in (0, 1]$, one gets some v_a with

$$\|u_1 - v_a\| \leqslant ad(u_1, E[u_2]),$$

$$\exists v_a^* \in \partial |f(v_a)| : \|v_a^*\| \leqslant \frac{|g(u_1) - g(u_2)|}{ad(u_1, E[u_2])}.$$

If $a < 1$, the first equation implies that $v_a \notin E[u_2]$, so that $f(v_a) \neq 0$, and $\partial |f(v_a)| = \pm f'(v_a) = \pm g'(v_a)$. When a goes to 1, the sequence v_a has a cluster point $v \in \mathbf{R}^n$ because it stays in the ball with radius $d(u_1, E[u_2])$ centered at u_1, and $g'(v_a)$ has $g'(v)$ as a cluster point since g' is continuous. The desired result follows immediately. □

Ioffe himself gives a different twist to his theorem; his paper is based on the following corollary:

COROLLARY 17 [26]. *Let E be the closed subset of S defined by the equation $f(w) = 0$, and take $w \in E$. Assume there are $\varepsilon > 0$ and $c > 0$ such that, for all $v \in S \setminus E$ with $\|v - w\| \leqslant \varepsilon$:*

$$v_1^* \in \partial|f(v)| \text{ and } v_2^* \in N_S(v) \Rightarrow \|v_1^* + v_2^*\|^* \geqslant c.$$

Then, for all $u \in S$ with $\|u - w\| \leqslant \varepsilon/2$:

$$d(u, E) \leqslant |f(u)|/c. \quad \square$$

PROOF. Theorem 15 gives us for every $a \in (0, 1)$ a point $v_a \in S$, (which does not belong to E since $\|v_a - u\| < d(u, E)$) such that

$$\|v_a - w\| \leqslant \|v_a - u\| + \|u - w\| \leqslant ad(u, E) + \|u - w\|$$
$$\leqslant (a + 1)\|u - w\| \leqslant (a + 1)\varepsilon/2 \leqslant \varepsilon$$

and $v_1^* \in \partial|f(v_a)|$ and $v_2^* \in N_S(v_a)$ such that

$$d(u, E) \leqslant |f(u)|/a\|v_1^* + v_2^*\|^*.$$

By assumption, the right-hand side is less than $|f(u)|/ac$. Letting a go to 1, we get the desired result. $\square$

Corollary 17 has important applications to optimization theory. Note that its scope is broader than is apparent at first glance: the set E defined by the single equation $f(w) = 0$ may be of codimension higher than one. Take for instance a set E originally written as $\{w | g_i(w) = 0, 1 \leqslant i \leqslant p\}$, the functions g_i being C^1, and the $g_i'(w)$, $1 \leqslant i \leqslant p$, being linearly independent along E. Then E can also be written as $\{w | f(w) = 0\}$, with $f(w) = \max_i |g_i(w)|$; then f would be locally Lipschitzian, and $\partial f(w)$ would be bounded away from zero on $N \setminus E$, with N some neighbourhood of E, and one could apply Corollary 17. On the other hand, if one tried to represent E as $\{w | g(w) = 0\}$ with g a C^1 function, one would have to use something like $g(w) = \sum_{i=1}^n g_i(w)^2$, and $g'(w)$ would not be bounded away from zero on N.

COROLLARY 18 ([25], [37]). *Let V and W be Banach spaces, and $A: V \to W$ be a linear continuous mapping with closed range. Let v_i^*, $1 \leqslant i \leqslant p$, be given in V^*, and consider the set:*

$$E = \{v \in V | Av = 0 \text{ and } \langle v, v_i^* \rangle \leqslant 0, 1 \leqslant i \leqslant n\}.$$

Then there is some $k > 0$ such that

$$d(v, E) \leqslant k\left(\|Av\| + \sum_{i=1}^n \langle v, v_i^* \rangle_+\right). \quad \square$$

Here $t_+ = \max\{0, t\}$, for $t \in \mathbf{R}$. We shall prove this result under the added assumption that V and W are finite-dimensional, that A is onto, and that there is some point v_0 in V with $Av_0 = 0$ and $\langle v_0, v_i^* \rangle < 0$ for all i. The general case is proved along the same lines.

PROOF. Consider the function $f: V \to \mathbf{R}$ defined by

$$f(v) = \|Av\| + \sum_{i=1}^n \langle v, v_i^* \rangle_+.$$

Then $E = \{v \in V | f(v) = 0\}$. Note that f is a sublinear function and E a convex cone, so that the inequality $d(v, E) \leqslant f(v)$ will hold everywhere if it is true in a neighbourhood of the origin. By Corollary 17, with $S = V$, it will be sufficient to show that $\partial f(v)$ is bounded away from zero when $|f(v)| \neq 0$.

It is easily seen that

$$\partial f(v) = \begin{cases} {}^tA\dfrac{Av}{\|Av\|} + \displaystyle\sum_{i=1}^{n} I(\langle v_i^*, v\rangle)v_i^* & \text{if } Av \neq 0, \\ {}^tAB + \displaystyle\sum_{i=1}^{n} I(\langle v_i^*, v\rangle)v_i^* & \text{if } Av = 0. \end{cases}$$

Here B is the unit ball of W, with boundary S and $I(t)$, for any $t \in \mathbf{R}$, is the interval defined by

$$I(t) = \begin{cases} \{0\} & \text{if } t < 0, \\ [0, 1] & \text{if } t = 0, \\ \{1\} & \text{if } t > 0. \end{cases}$$

In each case, any $v^* \in \partial f(v)$ can be expressed as

$$\partial f(v) = \lambda_0 {}^tAw^* + \sum_{i=1}^{n} \lambda_i v_i^*$$

with $\|w^*\| = 1$ and $0 \leqslant \lambda_i \leqslant 1$ for $i = 0, \ldots, n$. Moreover, if $f(v) \neq 0$, at least one of the λ_i is equal to 1.

Consider the $(n + 1)$ cube $\Pi_{i=0}^{n}[0, 1]$, and let C be its boundary. Let ϕ: $S \times C \to V$ be defined by $\phi(w^*, \lambda) = \lambda_0 {}^tAw^* + \Sigma_{i=1}^{n} \lambda_i v_i^*$. We have just proved that

$$f(v) \neq 0 \Rightarrow \partial f(v) \subset \phi(B \times C).$$

It is clear that $S \times C$ is compact, and so is $\phi(S \times C)$. It cannot contain the origin. Indeed, if $0 = \phi(w^*, \lambda)$ we have

$$\langle 0, v_0\rangle = \lambda_0 \langle w^*, Av_0\rangle + \sum_{i=1}^{n} \lambda_i \langle v_i^*, v_0\rangle$$

$$0 = \sum_{i=1}^{n} \lambda_i \langle v_i^*, v_0\rangle.$$

Since $\langle v_i^*, v_0\rangle > 0$ for all $i \geqslant 1$, this implies all the λ_i are zero for $i \geqslant 1$, and hence $\lambda_0 = 1$. Then $\phi(w^*, \lambda) = {}^tAw^*$ which cannot be zero since $\|w^*\| = 1$ and A is onto. So $\phi(w^*, \lambda) = 0$ leads to a contradiction.

It follows that $\phi(S \times C)$ is bounded away from zero, and so is $\partial f(v)$ for $f(v) \neq 0$. □

IV. The local statement. In this last type of results, Theorem 1 ter is used to find "generic" properties. Recall that, if V is a complete metric space, a subset $U \subset V$ is called *residual* if it is a countable intersection of open dense subsets, and a property $P(u)$ is called *generic* if it holds at all points u of a residual subset. Because of the Baire category theorem, whenever properties

$P_1(u), \ldots, P_n(u), \ldots, n \in \mathbf{N}$, are generic, then property $\bigwedge_{i=1}^{\infty} P_n(u)$ is generic too.

Now if V is endowed with a σ-algebra $\mathfrak{A}$ and a positive measure μ, a property $P(u)$ will be said to hold *almost everywhere* if there is a subset $N \in \mathfrak{A}$ with $\mu(N) = 0$ and $u \notin N \Rightarrow P(u)$. This notion may be more familiar; but the point is that many spaces, especially infinite-dimensional spaces, cannot be fitted with such $\mathfrak{A}$ and μ in any meaningful way. But they still are complete metric spaces, so that the category approach will succeed, whereas the measure theoretic one will not. This is of considerable practical importance, since most function spaces come under this heading.

A. Geometry of Banach spaces.

With Lebourg (see [23]), we have investigated point-wise infima of families of smooth functions. The setting is as follows; let X be any set, and V be a Banach space; let f be a real function on $X \times V$; define a function F on V by

$$F(v) = \inf_{x \in X} f(x, v). \tag{51}$$

Assuming $F(v) \not\equiv -\infty$ everywhere, the first question is: if the functions $v \to f(x, v)$ are all C^1, say, what differentiability properties would F have? Note that smoothness with respect to x is obviously irrelevant. Note also that F need not be differentiable everywhere, as examples with $X = \{1, 2\}$ and $V = \mathbf{R}$ will show. What can be hoped, though, is that conditions can be found under which F will be differentiable generically.

Theorem 19 [23]. *Assume that there exists a Fréchet-differentiable function ϕ: $V \to \mathbf{R}$ with $\phi(0) > 0$ and $\phi(w) \leqslant 0$ outside the unit ball. Assume moreover that, for any $u \in V$, we have $F(u) > -\infty$ and we can find $\eta > 0$ and $\theta > 0$ so that*

$$\text{the set } \{ f'_v(x, v) \mid \|v - u\| \leqslant \eta, f(x, v) \leqslant F(u) + \theta \} \text{ is bounded in } V^*, \tag{52}$$

$$\text{the set of functions } \{ v \to f'_v(x, v) \mid f(x, u) \leqslant F(u) + \theta \} \text{ is equicontinuous at } u \text{ in } V. \tag{53}$$

Then there is a residual set $G \subset V$ at every point u of which F is Fréchet-differentiable. Moreover for any $u \in G$, we have the property

$$f(x_n, u) \to F(u) \Rightarrow f'_v(x_n, u) \to F'(u). \quad \square \tag{54}$$

The existence of such a "smooth bump function" ϕ is a geometric property of the Banach space. It will be true, for instance, whenever V is finite-dimensional, or when it admits an equivalent norm which is Fréchet-differentiable. It means essentially that there should be sufficiently many Fréchet-differentiable functions on V.

Conditions (52) and (53) describe the actual assumptions on the family $v \to f(x, v)$, $x \in X$. I now proceed with the proof, referring to [23] for fuller details. With any $\varepsilon > 0$, we associate the set G_ε of all $u \in V$ such that there exists $\theta > 0$ and $\eta > 0$ with

$$\left.\begin{aligned}\|v_1 - u\| &\leqslant \eta, f(x_1, u) \leqslant F(u) + \theta \\ \|v_2 - u\| &\leqslant \eta, f(x_2, u) \leqslant F(u) + \theta\end{aligned}\right\} \Rightarrow \|f_v'(v_1, x_1) - f_v'(v_2, x_2)\|^* \leqslant \varepsilon. \tag{55}$$

The proof now proceeds in three steps:

Step 1. G_ε is open. Fix u in G_ε. By assumption (52), we can find η_0, θ_0 and k so that $\|f_v'(x, v)\|^* \leqslant k$ whenever $\|v - u\| \leqslant \eta_0$ and $f(x, v) \leqslant F(u) + \theta_0$. Moreover, the function F is upper semicontinuous at u, since it is a pointwise infimum of continuous functions. From there, it is not difficult to find $\eta_1 < \eta_0$ and $\theta_1 < \theta_0$ so that:

$$|f(x, v) - f(x, w)| \leqslant k\|v - w\|,$$
$$|F(v) - F(w)| \leqslant k\|v - w\|$$

whenever $\|v - u\| \leqslant \eta_1$, $\|w - u\| \leqslant \eta_1$, and $f(x, v) \leqslant F(v) + \theta_1$, $f(x, w) \leqslant F(w) + \theta_1$.

Now let η and θ be given with $\eta < \eta_1$ and $\theta < \theta_1$. Take $\eta' = \min(\eta/2, \theta/3k)$ and $\theta' = \theta/3$. Take any v such that $\|v - u\| \leqslant \eta'$. Then the relation $\|v_i - v\| \leqslant \eta'$ certainly implies $\|v_i - v\| \leqslant \eta/2$ (with $i = 1, 2$). The inequality $f(x_i, v) \leqslant F(v) + \theta'$ also will imply $f(x_i, u) \leqslant F(u) + \theta$. From condition (55) will follow $\|f_v'(x_1, v_1) - f_v'(x_2, v_2)\|^* \leqslant \varepsilon$. This proves that any v such that $\|v - u\| \leqslant \eta'$ belongs to G_ε.

Step 2. G_ε is dense. Since F is locally Lipschitzian, it is continuous. The Banach space V has the required function ϕ, so that we can apply Theorem 1 ter: the set S_ε of points where F is $\varepsilon/8$-supported is dense in V.

I claim S_ε is included in G_ε. Indeed, let u belong to S_ε. There exists some $\eta_0 > 0$ and $u^* \in V^*$ such that

$$\|u - v\| \leqslant \eta_1 \Rightarrow F(v) \geqslant F(u) + \langle u^*, v - u\rangle - \varepsilon\|v - u\|/8. \tag{56}$$

Start now on another tack. From the equicontinuity relation (53) it follows that some $\eta_1 < \eta_0$ and some $\theta > 0$ can be found such that

$$\|f_v'(x, u) - f_v'(x, v)\|^* \leqslant \varepsilon/8 \tag{57}$$

whenever $\|u - v\| \leqslant \eta_1$ and $f(x, u) \leqslant F(u) + \theta$. Integrating this relation along the line segment with endpoints u and v, we get

$$|f(x, v) - f(x, u) - \langle f_v'(x, u), v - u\rangle| \leqslant \varepsilon\|v - u\|/8. \tag{58}$$

We can always restrict θ to be less than $\varepsilon\eta_1/8$, so that this inequality becomes

$$F(v) \leqslant F(u) + \langle f_v'(x, u), v - u\rangle + \varepsilon\eta_1/4. \tag{59}$$

We now add (56) and (59). We get, for any v with $\|u - v\| \leqslant \eta_1$ and any x with $f(x, u) \leqslant F.(u) + \varepsilon\eta_1/8$:

$$\langle u^* - f_v'(x, u), v - u\rangle \leqslant 3\varepsilon\eta_1/8.$$

It follows immediately that $\|u^* - f_v'(x, u)\|^* \leqslant 3\varepsilon/8$. In view of inequality (57) and the triangle inequality, we have $\|u^* - f_v'(x, v)\|^* \leqslant \varepsilon/2$. The set of all $f_v'(x, v)$, for $\|u - v\| \leqslant \eta_1$ and $f(x, u) \leqslant F(u) + \varepsilon\eta_1/8$, is contained in the

ball of radius $\varepsilon/2$ centered at u^*. So its diameter must be less than ε, which is exactly condition (55). Hence u does belong to G_ε.

Step 3. $G = \bigcap_{n\in\mathbf{N}} G_{1/n}$ has the desired properties. G, as defined, certainly is a residual subset of V. Moreover, if u belongs to G, condition (55) is going to hold for any $\varepsilon > 0$. This means that, for any sequence (x_n, v_n) in $X \times V$ such that $v_n \to u$ in V and $f(x_n, u) \to F(u)$ in $\mathbf{R}$, the sequence $f'_v(x_n, v_n)$ is going to be Cauchy in V^*, and hence will converge to a limit u^*.

That u^* is the Fréchet-derivative of F at u now comes easily. On the one hand, use formula (58) with η_1 and θ so small that $\|f'_v(x, v) - u^*\|^* \leqslant \varepsilon/8$ for $\|v - u\| \leqslant \eta_1$ and $f(x, u) \leqslant F(u) + \theta$. Setting $x = x_n$, we get,

$$f(x_n, v) \leqslant f(x_n, u) + \langle u^*, v - u\rangle + \varepsilon\|v - u\|/4$$

and letting n go to infinity, this becomes, for $\|v - u\| \leqslant \eta_1$:

$$F(v) \leqslant F(u) + \langle u^*, v - u\rangle + \varepsilon\|v - u\|/4. \tag{60}$$

On the other hand, formula (58) also implies

$$f(x, v) \geqslant f(x, u) + \langle u^*, v - u\rangle - \varepsilon\|v - u\|/4 \tag{61}$$

whenever $\|u - v\| \leqslant \eta_1$ and $f(x, u) \leqslant F(u) + \theta$. Take $\eta_2 = \min(\eta_1, \theta/3k)$ with k the Lipschitz constant around u. If $\|u - v\| \leqslant \eta_2$ and $f(x, v) \leqslant F(v) + \theta/3$, we also will have $f(x, u) \leqslant F(u) + \theta$, so that inequality (61) will hold. Writing it at every point of a sequence y_n in X such that $f(y_n, v) \to F(v)$, we get, for $\|v - u\| \leqslant \eta_2$:

$$F(v) \geqslant F(u) + \langle u^*, v - u\rangle - \varepsilon\|v - u\|/4. \tag{62}$$

Comparing (60) and (62) yields the desired result. □

A first application of this result is to convex functions. Let us say that V is an Asplund space if every continuous convex function on V is Fréchet-differentiable at every point of a residual subset. It is a long-standing conjecture (see [1]) that V is an Asplund space if and only if there is an equivalent norm on V which is Fréchet-differentiable off the origin. Theorem 19 proves the "if" part; the "only if" part is still an open question.

COROLLARY 20 [23]. *Every Banach space with a Fréchet-differentiable norm (off the origin) is an Asplund space.*

PROOF. It follows from the Hahn-Banach theorem in $V \times \mathbf{R}$, that every continuous convex function F on V is the point-wise supremum of a family of affine functionals. In other words, there exists a function F^*: $V^* \to \mathbf{R} \cup \{+\infty\}$ (which turns out to be convex and l.s.c.) such that

$$\forall v \in V, \quad F(v) = \sup\{\langle v, v^*\rangle - F^*(v^*) | v^* \in V^*\}.$$

It can be proved that, for every point $u \in V$, there is a neighbourhood U of u and a bounded subset X of V^* such that

$$\forall v \in U, \quad F(v) = \sup\{\langle v, v^*\rangle - F^*(v^*) | v^* \in X\}$$

(this implies for instance that F is locally Lipschitzian, a well-known fact). One now applies Theorem 19, with $f(v^*, v) = F^*(v^*) - \langle v, v^*\rangle$. Checking assumptions (52) and (53) with $f'_v(v^*, v) = -v^* \in X$ is perfectly straight-forward, and the result follows. □

Another application is to closest points and farthest points in a closed subset of a Banach space. These results were first proved by Edelstein ([5], [6]).

COROLLARY 21. *Let V be a Banach space. Assume V and its dual V^* are uniformly convex. Let X be a bounded and closed subset of V. Then there is a residual set G in V such that*

(a) *for any $u \in G$, there is a single point x in X such that $\|u - x\|$ is minimum (x is the "projection" of u on X).*

(b) *for any $u \in G$, there is a single point y in X such that $\|u - y\|$ is maximum (y is farthest from u in X).*

PROOF. For the sake of simplicity, assume V is a Hilbert space. We then apply Theorem 19 to the function

$$f(x, u) = \|u - x\|^2.$$

Note that $F(u)$ then is just $d(u, X)^2$. We have: $f_v'(x, v) = 2(v - x)$. Assumptions (52) and (53) are easily checked, so that conclusion (54) holds at all points u of a residual subset G_+:

$$\|u - x_n\|^2 \to d(u, X)^2 \Rightarrow 2(u - x_n) \to F'(u).$$

This can also be written as follows:

$$\|u - x_n\| \to d(u, X) \Rightarrow x_n \to u - \tfrac{1}{2}F'(u).$$

Since X is closed, the limit $x = u - \frac{1}{2}F'(u)$ also belongs to X, so it is the closest point to u in X. There couldn't be another one, x' say, for the sequence x_n defined by $x_{2p} = x$ and $x_{2p+1} = x'$ would be minimizing ($\|u - x_n\| = d(u, X)$ all n) and yet wouldn't be Cauchy. So property (a) holds for $G = G_+$.

Similarly, applying Theorem 19 to the function $f(x, u) = -\|u - x\|^2$, one gets a residual set G_- on which property (b) holds. Finally, both properties (a) and (b) will hold on the residual subset $G = G_+ \cap G_-$. □

Recent progress in this direction has been achieved by Lebourg (see [28]) and Borwein (see [3]).

B. GLOBAL ANALYSIS.

From now on V will not be a Banach space anymore, but a Riemannian manifold. This means that V is a C^∞ manifold, possibly infinite-dimensional, modelled on some Hilbert space H, and that for every $u \in V$ we are given on the tangent space $TM_u \simeq H$ a scalar product $\langle \cdot, \cdot \rangle_p$, depending smoothly on u, with $\| \cdot \|_u$ defining an equivalent norm on H. The length of any C^1 path c: $[0, T] \to V$ is defined as

$$L(c) = \int_0^T \left\| \frac{dc}{dt}(t) \right\|_{c(t)} dt.$$

It is easily seen to depend only on the curve $c([0, T])$ itself, and not on the way the point $c(t)$ moves along it; namely, if ϕ: $[0, 1] \to [0, T]$ is a diffeomorphism, $L(c \circ \phi) = L(c)$. For that reason, it is always possible to take $T = 1$, which we will do systematically.

The Riemannian structure defines on M a metric which is compatible with its manifold topology; namely, given any two points u and v in the same component of M, denote by $\mathcal{C}_u^v$ the set of all C^1 paths from u to v, and define the distance $d(u, v)$ as the infimum of the lengths of all such paths:

$$\mathcal{C}_u^v = \{c \in C^1([0, 1], V) | c(0) = u, c(1) = v\},$$

$$d(u, v) = \inf\{L(c) | c \in \mathcal{C}_u^v\}.$$

Any path $c \in \mathcal{C}_u^v$ with smallest length is called a minimal geodesic. A basic question in the geometry of Riemannian manifolds is whether there is a minimal geodesic between two points, and whether it is unique.

The local problem is easy to answer. It follows from the definition of the Riemannian structure, and from the local existence and uniqueness results of classical analysis (theorems on implicit functions, or solutions to differential equations) that every point $u \in V$ has a neighbourhood U such that between any two points v_1 and v_2 in U there is a unique minimal geodesic. Moreover, there is some $\delta > 0$ and a local chart $\exp_u: \delta B \to U$ (where B is the unit ball in H) such that

$$\|\xi\| \leqslant \delta \Rightarrow d(u, \exp_u \xi) = \|\xi\|. \tag{63}$$

The global problem is much more difficult. In finite dimensions, the existence question is solved by the Hopf-Rinow theorem: if V is complete in the d-metric, and connected, there is at least one minimal geodesic between any two points. Using Sard's theorem, it can also be shown that for any fixed $u \in V$, the property that there is a unique minimal geodesic between u and v holds for almost every v in V (see [**31**] for all this).

The Hopf-Rinow theorem is known to be false in the infinite-dimensional case (neither can Sard's theorem be used). Indeed, consider in l^2 the infinite-dimensional ellipsoid $\Sigma_{n=1}^{\infty} x_n^2/a_n^2 = 1$, with a_n (half-length of the nth axis) $= 1/n$ for instance; the coordinates x_n, $n \in \mathbf{N}$, are taken with respect to the orthonormal basis e_n, $n \in \mathbf{N}$. Cutting the ellipsoid by the 2-plane generated by e_0 and e_n, we find a path c_n between the North pole e_0 and the South pole $-e_0$. Clearly c_n converges to the line segment c_∞ between e_0 and $-e_0$, defined by $c_\infty(t) = (1 - 2t)e_0$. We have $L(c_\infty) = 2 = \inf_n L(c_n)$, so the distance from e_0 to $-e_0$ on the ellipsoid is 2. However, the only path with length 2 between e_0 and $-e_0$ is c_∞, which does not lie on the ellipsoid.

In 1976, I proved that existence, as well as uniqueness, are generic properties in the infinite-dimensional case. In view of the above, this result seems to be optimal.

THEOREM 22 [**21**]. *Let V be a (possibly infinite-dimensional) Riemannian manifold, complete in the d-metric, and connected. Let any $u \in V$ be given. Then there is a residual subset of V, every point v of which can be joined to u by a unique minimal geodesic.* □

This is proved by using the local statement, Theorem 1 ter, with minor modifications to suit the new, nonlinear, setting. A function $F: V \to \mathbf{R}$ will be called ε-supported at $v \in V$ if there exist $\eta > 0$ and $\zeta \in TV_u$ such that

$$\|\xi\|_v \leqslant \eta \Rightarrow F(\exp_v \xi) \geqslant F(v) + \langle \zeta, \xi \rangle_v - \varepsilon \|\xi\|_v. \tag{64}$$

The local statement now is that, given $\varepsilon > 0$, any l.s.c. function is ε-supported on a dense subset of V (see [21] for details). This is the one we are going to use for proving Theorem 22.

It is not difficult to get a feeling for this proof. Let u be fixed in V. Given another point $v \in V$, we seek a minimal geodesic between v and u. Now such a path has to solve a second-order differential equation, the so-called geodesics equation, which is simply the Euler-Lagrange equation associated with the path integral $\frac{1}{2}\int_0^T \|dc(t)\,dt\|^2_{c(t)}\,dt$. Any solution starting from v is fully known if its tangent vector at v is given (because $c(0) = v$ and $dc(0)/dt = \xi$ yields a Cauchy problem). The fact that the d-metric is complete implies that such a solution will be defined on the whole time interval $[0, \infty)$.

It now is a question of aiming: how does one find ξ in TV_v so that the solution to the geodesics equation, with $c(0) = v$ and $dc(0)/dt = \xi$, reaches a given point u? If the function $F(w) = d(w, u)$ is differentiable at v, it is clear enough how to aim: one should set $\xi = \operatorname{grad}_v F$. But there is no guarantee this will ever happen, except in a small neighbourhood of u. On the other hand, we know that F is going to be ε-supported on a dense subset of V, and this certainly restricts the possibilities for ξ. Indeed, if F is ε-supported at v, the vectors ζ satisfying (64) build up a closed cone $C_\varepsilon(v)$, in which the direction ξ obviously is to be chosen. The smaller ε is, the smaller the cone, and the better the aim; if $\varepsilon = 0$, the cone becomes a half-line, and there is only one possible direction for ξ.

The proof goes by seeking a residual subset G of V on which F is ε-supported for all $\varepsilon > 0$; if v belongs to G, the cone $\bigcap_{\varepsilon>0} C_\varepsilon(v)$ reduces to a half-line, and one aims in that direction. As in the preceding section (IV A), density and openness are proved separately. There are still many technicalities to be overcome, and the set G is described in a more careful way than I have just stated. However, I will be content with that; let me just give the first step in the proof.

LEMMA 23. *Assume the function $F(w) = \|w - u\|$ is ε-supported at $v \neq u$, for some $\varepsilon > 0$. Then, for any $\theta > 4\sqrt{\varepsilon}$, some $\delta > 0$ can be found such that*

$$\forall \eta \in (0, \delta), \exists \alpha > 0: \textit{diameter } C_v^\eta(\alpha) \leqslant \theta\eta,$$

where the set $C_v^\eta(\alpha)$ is defined by

$$C_v^\eta(\alpha) = \{c(t) \mid c \in \mathcal{C}_v^u,\ L(c) \leqslant d(u, v) + \alpha,\ d(v, c(t)) = \eta\}. \quad \square$$

Before giving the proof of this lemma, I want to explain its meaning. Imagine c as a piece of string connecting u and v. If this string were perfectly taut, we would have $L(c) = d(u, v)$. We cannot achieve this, only $L(c) \leqslant d(u, v) + \alpha$; the smaller $\alpha > 0$, the tauter the string. Now $C_v^\eta(\alpha)$ is just the set of points where such taut strings cut a small sphere of radius η around v. The relation $\operatorname{diam} C_v^\eta(\alpha) \leqslant \theta\eta$ means that this set is seen from v under an angle θ. The lemma can be understood as saying that if two strings are taut enough, they start out in about the same direction. Now to the proof.

PROOF. Take $\beta > 0$ so small that $\theta > 4(\varepsilon + \beta/2)^{1/2}$. Using formula (64), we find $\delta > 0$ and $\zeta \in TV_v$ such that

$$\|\xi\|_v \leqslant \delta \Rightarrow d(u, \exp_v \xi) \geqslant d(u, v) + \langle \zeta, \xi \rangle_v - \varepsilon\|\xi\|_v.$$

First, we apply the triangle inequality to the left-hand side, and use formula (63). This yields

$$\|\xi\|_v \leqslant \delta \Rightarrow (1+\varepsilon)\|\xi\|_v \geqslant \langle \zeta, \xi \rangle_v,$$

and hence

$$\|\zeta\|_v \leqslant 1+\varepsilon.$$

On the other hand, if we take any $\eta \in (0, \delta)$, any path c with $L(c) \leqslant d(u, v) + \beta\eta$, and any time t such that $d(v, c(t)) = \eta$, we have

$$\begin{aligned} d(u, c(t)) &\leqslant L(c) - \int_0^t \left\| \frac{dc}{ds}(s) \right\|_{c(s)} ds \\ &\leqslant d(u, v) + \beta\eta - d(v, c(t)). \end{aligned}$$

Writing this inequality into the first one, by setting $c(t) = \exp_v \xi(t)$, we get

$$-(1-\varepsilon)\|\xi(t)\|_v + \beta\eta \geqslant \langle \xi(t), \zeta \rangle_v.$$

Dividing by η throughout, this becomes

$$\langle \zeta, \xi(t)/\eta \rangle \leqslant -1 + \varepsilon + \beta.$$

Dividing by $\|\zeta\|_v$, which is less than $1+\varepsilon$, we get

$$\left\langle -\frac{\zeta}{\|\zeta\|_v}, \frac{\xi(t)}{\eta} \right\rangle \geqslant \frac{1-\varepsilon-\beta}{1+\varepsilon}.$$

Since both $-\zeta/\|\zeta\|_v$ and $\xi(t)/\eta$ are unit vectors, this implies that their distance is less than $[2(2\varepsilon+\beta)]^{1/2}$. By formula (63) the distance between the points $\exp_v(-\zeta\eta/\|\zeta\|_v)$ and $\exp_v \xi(t)$ is at most $2\eta(\varepsilon + \beta/2)^{1/2}$, which is less than $\theta/2$. Since the first one is fixed, and the second one can be anywhere in $C_v^\eta(\alpha)$, the desired result follows. □

V. BY WAY OF CONCLUSION.

I hope to have shown how fruitful Theorem 1 can be in the study of minimization problems. Proving the existence of optimal solutions, or characterizing them by first-order necessary conditions, both fall within its scope.

It is my hope that it will be found relevant to the study of more general variational problems, where one does not seek a minimum, but some kind of saddle-point. A typical instance of such problems is the search for periodic solutions to Hamiltonian equations, with the recent work of Rabinowitz to show how difficult it is (see [36]).

A starting point would be to find conditions under which an unbounded functional F on a Banach space V has a critical point, i.e. some point v where $F'(v) = 0$. All I have to offer in that direction is the following result (Newton's method revisited):

PROPOSITION 25. *Let V be a Hilbert space, and $F: V \to \mathbf{R}$ a C^2 functional. Set $S = \{u \in V | F'(u) = 0\}$, and assume $F''(u)F'(u)/\|F'(u)\|$ is bounded away from zero on $V \setminus S$. Then S is nonempty.* □

PROOF. Assume S is empty. Then the functional $G(u) = \|F'(u)\|$ is C^1 on V, with derivative:

$$G'(u) = F''(u)F'(u)/\|F'(u)\|.$$

By Theorem 1 bis, there is a sequence u_n in V such that $G(u_n) \to \inf G$ and $\|G'(u_n)\| \to 0$. This contradicts the fact that $F''(u)F'(u)/\|F'(u)\|$ is bounded away from zero on V. □

BIBLIOGRAPHY

1. E. Asplund, *Fréchet-differentiability of convex functions*, Acta Math. **121** (1968), 31–47.

2. J. Aubin and J. Siegel, *Fixed points and stationary points of dissipative multivalued maps*, M.R. report 7712, University of Southern California, Los Angeles, 1977.

3. J. Borwein, *Weak local supportability and applications to approximation* (preprint), Dalhousie University, Halifax, September 1977.

4. M. Crandall, 1976, personal communication.

5. M. Edelstein, *Farthest points of sets in uniformly convex Banach spaces*, Israel J. Math. **4** (1966), 171–176.

6. _____, *On nearest points of sets in uniformly convex Banach spaces*, J. London Math. Soc. **43** (1968), 375–377.

7. E. Bishop and R. R. Phelps, *A proof that all Banach spaces are subreflexive*, Bull. Amer. Math. Soc. **67** (1961), 97–98.

8. _____, *The support functional of a convex set*, Convexity (Klee, ed.), Proc. Sympos. Pure Math., vol. 7, Amer. Math. Soc., Providence, R. I., 1963, pp. 27–35.

9. A. Brøndsted and R. T. Rockafellar, *On the subdifferentiability of convex functions*, Proc. Amer. Math. Soc. **16** (1965), 605–611.

10. H. Brézis and F. Browder, *A general ordering principle in nonlinear functional analysis*, Advances in Math. **21** (1976), 355–364.

11. F. Browder, *Normal solvability for nonlinear mappings into Banach spaces*, Bull. Amer. Math. Soc. **79** (1973), 328–350.

12. J. Caristi, *Fixed point theorems for mappings satisfying inwardness conditions*, Trans. Amer. Math. Soc. **215** (1976), 241–251.

13. F. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc. **205** (1975), 247–262.

14. _____, *The maximum principle under minimum hypothesis*, SIAM J. Control and Optimization **14** (1976), 1078–1091.

15. _____, *A new approach to Lagrange multipliers*, Math. Operations Res. **1** (1976), 165–174.

16. _____, *Necessary conditions for a general control problem*, Calculus of Variations and Optimal Control Theory (Russell, ed.), Academic Press, New York, 1976, pp. 257–278.

17. _____, *Generalized gradients of Lipschitz functionals*, MRC Technical Report #1687, University of Wisconsin, Madison, Wis., August 1976.

18. _____, *Pointwise contraction criteria for the existence of fixed points*, MRC Technical Report #1658, July 1976, University of Wisconsin, Madison, Wis.; Bull. Canad. Math. Soc. (to appear).

19. I. Ekeland, *Sur les problèmes variationnels*, C. R. Acad. Sci. Paris **275** (1972), 1057–1059; **276** (1973), 1347–1348.

20. _____, *On the variational principle*, J. Math. Anal. Appl. **47** (1974), 324–353.

21. _____, *The Hopf-Rinow theorem in infinite dimension*, MRC Technical Report #1692, August 1976, University of Wisconsin, Madison, Wis.; J. Differential Geometry (to appear).

22. _____, *Le théorème de Hopf-Rinow en dimension infinie*, C. R. Acad. Sci. Paris **284** (1977), 149–150.

23. I. Ekeland and G. Lebourg, *Generic Frechet-differentiability and perturbed optimization problems in Banach spaces*, Trans. Amer. Math. Soc. **224** (1976), 193–216.

24. I. Ekeland and R. Temam, *Convex analysis and variational problems*, North-Holland, Amsterdam, 1976.

25. A. Hoffman, *On approximate solutions of systems of linear inequalities*, J. Res. Nat. Bureau Standards **49** (1952), 263–265.

26. A. Ioffe, *Regular points of Lipschitz mappings*, Trans. Amer. Math. Soc. (to appear).

27. J. Lasry, personal communication, 1972.

28. G. Lebourg, *Problèmes d'optimisation perturbés dans les espaces de Banach*, preprint, CEREMADE, Université Paris-Dauphine, 1978.

29. R. Martin, *Invariant sets for evolution systems*, Proc. International Conference on Differential Equations (Antosiewicz, ed.), Academic Press, New York, 1975.

30. M. Maschler and B. Peleg, *Stable sets and stable points of set-valued dynamic systems*, SIAM J. Control **14** (1976), 985–995.

31. J. Milnor, *Morse theory*, Ann. of Math. Studies, Princeton Univ. Press, Princeton, N. J., 1963.

32. L. Neustadt, *Optimization: a theory of necessary conditions*, Princeton Univ. Press, Princeton, N. J., 1976.

33. R. Palais, *Morse theory on Hilbert manifolds*, Topology **2** (1963), 299–340.

34. R. Pallu de la Barriere, *Optimal control theory*, Saunders, Philadelphia, Penn., 1967.

35. L. Pontryagin, V. Boltyanskiĭ, R. Gamkrelidze and E. Mischenko, *The mathematical theory of optimal processes*, Wiley, New York, 1962.

36. P. Rabinowitz, *Periodic solutions of hamiltonian systems*, MRC Technical Report #1783, August 1977.

37. S. Robinson, *Regularity and stability for convex multivalued functions*, Math. Operations Res. (1976), 130–143.

38. R. Palais and S. Smale, *A generalized Morse theory*, Bull. Amer. Math. Soc. **70** (1964), 165–170.

39. I. Ekeland and M. Valadier, *Representation of set-valued mappings*, J. Math. Anal. Appl. **35** (1971), 621–629.

DÉPARTEMENT DE MATHEMATIQUES, UNIVERSITÉ PARIS-DAUPHINE, PARIS, FRANCE

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER V6T 1W5, BRITISH COLUMBIA, CANADA