MAXIMA OF SMOOTH FAMILIES III:

MORSE-SARD THEOREM

Y. Yomdin

Max-Planck-Institut für
Mathematik
Gottfried-Claren-Straße 26

5300 Bonn 3
Federal Republic of Germany

Ben Gurion University
of the Negev
Beer-Sheva 84105

ISRAEL

MPI/SFB 84 - 38

Abstract.

Let f be a Lipschitzian function. For $\gamma \geq 0$ the point x is called a γ-critical point of f, if the distance between the generalized gradient $\partial f(x)$ and 0 is at most γ. The values of f at γ-critical points form the set of γ-critical values of f. For functions f, representable as $f(x) = \max_t h(x, t)$, where f (x, t) is sufficiently smooth in both variables, we give the upper bound for the "size" of the set of γ-critical values of f in terms of γ and the uniform bounds of derivatives of h. In particular, for $\gamma = 0$ we obtain the extension of the Morse-Sard theorem to the class of maximum functions.

1. Introduction

The Morse-Sard theorem (see e. g. [6]), which asserts, that the set of critical values of a sufficiently smooth mapping has measure zero, is an important tool in smooth analysis. It is widely used in study of critical points of smooth functions and, in particular, in smooth optimization. Also in many questions of a nonsmooth analysis and optimization it is important to have the similar techniques.

However, some essential problems arise in attempts to extend the Morse-Sard theorem to the nonsmooth case. First of all, the differentiability assumptions in this theorem are sharp, and hence, it is simply false for the classes of nonsmooth functions, usually considered in optimization (except of the case where the dimensions of the source and the image coincide, see [4], proposition 1.2.). E. g. for any $n \geq 2$ one can easily construct a Lipschitzian (and even C^{n-1})-function of n variables, whose critical values cover the whole interval (see [9]). Thus, the best one can hope is to find a subclass of the class of Lipschitzian functions, which is sufficiently big to contain nonsmooth functions, important in applications, and, on the other hand, which is small enough to allow the extension of the Morse-Sard theorem.

In this paper we consider one such a subclass, consisting of the functions, representable as maxima (minima) of smooth families. The intensive study of this class was started recently (see [5], [7] , [11] , [12]). The results obtained show, from one side, that this class is appropriate in many questions of nonsmooth analysis and optimization; from the other side, reach techniques from smooth analysis can be used in its investigation.

Our main result is a general theorem about the "size" of the set of critical (and "almost-critical") values of functions, representable as maxima of smooth families. This theorem gives, in particular, the conditions, under which critical values have measure zero (in terms of smoothness of underlying families, but not of maxima functions themselves).

In fact, our result is much stronger than the usual Morse-Sard theorem (even in a smooth case). The main advantage is that we consider not only exactly critical, but also "almost critical" points and values; this makes the results "stable" with respect to inaccuracy of initial data and computations. Another important point is that, instead of the Lebesgue measure, we use the metric entropy, which turns to be much more effective in applications (see sections 2 and 4 below for a detailed discussion of these questions).

The possibility to extend the Morse-Sard theorem gives an additional evidence for the importance of the class of maxima functions of smooth families. It turns out that also the Morse theory (including the existence, density and stability of Morse functions, normal forms for critical points and, of course, the usual topological conclusions) can be extended to maximum functions. We hope to publish these results seperately.

The author would like to thank J.-B. Hirriart-Urruty for providing references [3] and [4] and the Max-Planck-Institute für Mathematik, where this paper was written, for kind hospitality.

2. Definitions and statement of main results

The critical point of a differentiable function is a point, where the gradient is zero. Now let $f: D \to R$, where D is a bounded domain in R^n, be a Lipschitzian function, i. e. $| f(x) - f(y)| \leq K \|x - y\|$ for some K and any $x,y \in D$.

Definition 2.1. (See [1]). The generalized gradient $\partial f(x)$ of f at $x \in D$ is the convex hull in R^n of all the $v \in R^n$ of the form

$$v = \lim_{x_i \to x} \operatorname{grad} f(x_i),$$

where $x_i \in D$ converge to x and f is differentiable at x_i for each i.

(Note that f, as a Lipschitzian function, is differentiable almost everywhere in D).

Definition 2.2. The point $x \in D$ is called critical point of f, if $\partial f(x)$ contains zero. The values of f at critical points are called critical values.

This definition agrees with the usual one for differentiable functions and it is rather natural also in a general case. E. g. if $x \in D$ is regular (i. e. non-critical) for a Lipschitzian function f, then in some new Lipschitzian coordinate system $(y_1, \ldots, y_n)$ near x, f can be written as $f(y_1, \ldots, y_n) = c + y_1$.

However, in numerically-oriented applications this notion is not very convenient. Indeed, if the accuracy of intial data and computations is finite, we can never say if the grad $f(x)$ is exactly zero ($\partial f(x)$ exactly

contains zero). Namely for the Morse-Sard theorem this difficulty is crucial: both the statement and the conclusion of this theorem are meaningless in the framework of approximate computations.
This justifies the following definition:

Definition 2.3. For $\gamma \geq 0$ the point $x \in D$ is called a γ-critical point of a Lipschitzian function $f : D \to R$, if the distance between $\partial f(x)$ and 0 is at most γ. The set of γ-critical points of f is denoted by $\Sigma(f,\gamma)$ and the set of γ-critical values $f(\Sigma(f,\gamma)) \subset R$ is denoted by $\Delta(f,\gamma)$.

For $\gamma = 0$ the γ-critical points and values coincide with the usual ones. But if the accuracy of our computations is $h > 0$, we are forced to work with γ-critical points and values, with $\gamma > h > 0$.

Our main theorem gives the upper bound for the "size" of the set $\Delta(f,\gamma)$; Before we specify the notion of "size", we should mention an additional disadvantage of the usual Morse-Sard theorem, which makes difficult its "effective" applications: the usage of the Lebesgue (or the Hausforff) measure. Indeed, the fact that the Lebesgue measure of the set A is small, does not allow to find effectively the points out of this set.

Fortunately, the usage of another geometric invariant, much more "effective" than the Lebesgue measure, is not only desirable, but also very natural in questions, concerning the geometry of critical values (see [10]).

Definition 2.4. For a bounded $A \subset R$ and for any $\varepsilon > 0$ define $M(\varepsilon, A)$ as the minimal number of intervals of length ε, covering A. $H(\varepsilon, A) = \log_2 M(\varepsilon, A)$

is called the ε-entropy of the set A (See [2], [8]).

We give some properties and applications of the ε-entropy (concerning mostly the "effective" finding of points out of the set with small entropy) in section 4. Here we consider only two examples: for $A = [0, 1]$, $M(\varepsilon, A) \sim {}^1/\varepsilon$, and for $A = \{1, \frac{1}{2}, \ldots, \frac{1}{n}, \ldots\}$, $M(\varepsilon, A) \sim {}^1/\sqrt{\varepsilon}$. In general, if for some $A \subset [0, 1]$ we obtain $M(\varepsilon, A) \ll \frac{1}{\varepsilon}$, we can conclude not only that A is "small" but also that it is "sparce" in [0, 1].

Now we define the class of nonsmooth functions we work with. These are the functions representable as $f(x) = \max_t h(x, t)$, where h depend smoothly on x and t. To simplify the notations and to obtain explicit constants we assume below, that the domain of definition of f is the closed unit ball B^n in R^n, and the domain of the parameter t is the closed unit ball B^m in R^m.

Definition 2.5. For n, m, k = 1, 2, ... and $C \geq 0$ let $S^k_{n,m}(C)$ denote the set of functions f, defined on the unit ball $B^n \subset R^n$ and representable as

$$f(x) = \max_{t \in B^m} h(x, t) \quad , \quad x \in B^n$$

where $h : B^n \times B^m \to R$ is a k times continuously differentiable function and the uniform norm of the k - th derivative of h does not exceed C. Let $S^k_{n,m} = \bigcup_{C > 0} S^k_{n,m}(C)$.

One can easily show, that $S^k_{n,m}$ coincides with the set of functions on B^n, representable as $f(x) = \max_{t \in T} h(x, t)$, where T^m is any compact m-dimensional manifold and $h : B^n \times T^m \to R$ a C^k-smooth function.

Clearly, all the functions in $S^k_{n,\,m}$ are Lipschitzian.
Now we can formulate the main result of this paper:

Theorem 2.6. There exist constants A_o, A_1, B, depending only on n, m and k, with the following properties:

Let for some $C \geq 0$, $f \in S^k_{n,\,m}$ (C), and let for some $\gamma \geq 0$, Δ (f, γ) be the set of γ-critical values of f.
Then for any $\varepsilon > 0$,

$$M(\varepsilon, \Delta(f, \gamma)) \leq \begin{cases} A_o + A_1 \gamma \cdot \frac{1}{\varepsilon}, & \varepsilon \geq BC \\ A_o \left(\frac{BC}{\varepsilon}\right)^{\frac{N}{k}} + A_1 \gamma\left(\frac{1}{\varepsilon}\right)\left(\frac{BC}{\varepsilon}\right)^{\frac{N-1}{k}}, & \varepsilon \leq BC, \end{cases}$$

where $N = (m + 2)(n + 1) - 2$.

We give some consequences and applications of theorem 2.6 in section 4 below. Now, to clarify the structure of the inequalities, consider only the special case $\gamma = 0$. Then the second term in the bounds for $M(\varepsilon, \Delta(f,0))$. disappears, and for small ε we obtain

$$M(\varepsilon, \Delta(f,0) \leq K \left(\frac{1}{\varepsilon}\right)^{\frac{N}{k}},$$ where the constant K depends on n, m, k and C.

If, in addition, we assume that the smoothness k is greater than N , we obtain that the number of intervals of length ε , covering Δ (f, 0), grows slowler than $^1/\varepsilon$ as $\varepsilon \to 0$. Hence the measure of Δ (f, 0) is zero, and we obtain the following:

Corollary 2.7. (The Morse-Sard theorem for maximum functions)
Let $f \in S^k_{n,m}$ with $k > (m + 2)(n + 1) - 2$.

Then the set of critical values Δ (f, 0) of f has the Lebesgue measure zero.

3. Proof of theorem 2.6.

We reduce the study of the γ-critical values of $f(x) = \max_t h(x, t)$ to the study of γ-critical values of some auxiliary smooth function.

Let Δ^n be the standard n-dimensional simplex $\Delta^n = \{ (\lambda_o, \ldots, \lambda_n), \lambda_i \geq 0, \lambda_o + \ldots + \lambda_n = 1\}$.

Consider the function

$$\phi : B^n \times \underbrace{B^m \times \ldots \times B^m}_{n+1} \times \Delta^n \to R,$$

defined as follows:

$$\phi (x, t_o, \ldots, t_n, \gamma_o, \ldots, \gamma_n) = \sum_{i=o}^{n} \gamma_i h (x, t_i).$$

(Since $f \in S^k_{n, m}$ (C), we fix some representation of f as

$$f(x) = \max_{t \in B^m} h(x, t)$$

with $h : B^n \times B^m \to R$ k times continuously differentiable and with the uniform norm of the k-th derivative of h not exceeding C).

Clearly, ϕ is k times continuously differentiable and the uniform norm of the k-th derivative of ϕ does not exceed those of h, and hence, is at most C.

<u>Lemma 3.1.</u> For any $\gamma \geq 0$ each γ-critical value of f is also a γ-critical value of ϕ (or of the restriction of ϕ on the boundary of $B^n \times B^m \times \ldots \times B^m \times \Delta^n$).

<u>Proof.</u> First of all, we need a precise description of the generalized gradient of a maximum function. Let $x_o \in B^n$. Denote by $T(x_o)$ the closed set in B^m,

consisting of those $t \in B^m$, for which

$$h(x_o, t) = f(x_o) = \max_{t \in B^m} h(x_o, t)$$

<u>Lemma 3.2.</u> ([1] , theorem 2.1). The generalized gradient $\partial f(x_o)$ coincides with the convex hull of all the vectors $\text{grad}_x h(x_o, t)$, $t \in T(x_o)$.

Now let $x_o \in B^n$ be a γ-critical point of f. By definition 2.3, we can find the vector $v_o \in \partial f(x_o)$ with $\| v_o \| \leq \gamma$. By lemma 3.2, v_o belongs to the convex hull of all the vectors of the form $\text{grad}_x h(x_o, t)$, $t \in T(x_o)$, and by the Caratheodory theorem, v_o belongs in fact to the convex hull of at most $n + 1$ such vectors.

Hence, we can find the points $\bar{t}_o, \ldots, \bar{t}_n \in T(x_o)$ and the nonnegative numbers $\bar{\lambda}_o, \ldots, \bar{\lambda}_n$, $\bar{\lambda}_o + \ldots + \bar{\lambda}_n = 1$, such that

$$v_o = \sum_{i=o}^{n} \bar{\lambda}_i \, \text{grad}_x h(x_o, \bar{t}_i) .$$

Consider the point $\bar{z} = (x_o, \bar{t}_o, \ldots, \bar{t}_n, \bar{\lambda}_o, \ldots, \bar{\lambda}_n) \in U = B^n \times B^m \times \ldots \times B^m \times$
Assume in addition, that the points t_i, $i = o, \ldots, n$, belong to the interior of the ball B^m.

<u>Lemma 3.3.</u>

1. $\phi(\bar{z}) = f(x_o)$
2. $\bar{z}$ is a γ-critical point of ϕ

<u>Proof.</u> By definition of $T(x_o)$, for any $t \in T(x_o)$, $h(x_o, t) = f(x_o)$. hence, $\phi(z) = \sum_{i=o}^{n} \lambda_i h(x_o, \bar{t}_i) = f(x_o) \sum_{i=o}^{n} \lambda_i = f(x_o)$.

Furthermore, we have the following identities for the first derivatives of ϕ at $\bar{z}$:

1. $d_{t_i} \phi(\bar{z}) = 0, i = 0, \ldots, n$
2. $d_\lambda \phi(\bar{z}) = 0$
3. $\text{grad}_x \phi(\bar{z}) = v_o$

Indeed, at each $\bar{t}_i$, $i = 0, \ldots, n$, the function $h(x_o, t)$ attains its maximal value on B^m and since, by assumptions, $\bar{t}_i$ is an inner point of B^m, $d_t h(x_o, \bar{t}_i) = 0, i = 0, \ldots, n$. In turn, $d_{t_i} \phi(\bar{z}) = \bar{\lambda}_i d_t h(x_o, \bar{t}_i) = 0$.

The identity $d_\lambda \phi(\bar{z}) = 0$ follows immediately from the fact that $h(x_o, \bar{t}_i) = f(x_o)$, $i = o, \ldots, n$.

Finally, $\text{grad}_x \phi(\bar{z}) = \sum_{i=o}^{n} \bar{\lambda}_i \text{grad}_x h(x_o, \bar{t}_i) = v_o$ by the choice of $\bar{\lambda}_i, \bar{t}_i$.

Thus $\text{grad} \phi(\bar{z}) = \text{grad}_x \phi(\bar{z}) = v_o$ and hence $||\text{grad} \phi(\bar{z})|| = ||v_o|| \leq \gamma$, and $\bar{z}$ is a γ-critical point of ϕ. Lemma 3.3 is proved.

If some of the points $\bar{t}_i$ belong to the boundary of B^m, we obtain in the same way that $\bar{z}$ is a γ-critical point of the restriction of ϕ on the boundary of U.

Since by construction $f(x_o) = \phi(\bar{z})$, lemma 3.1 follows.

Now we apply to the function ϕ theorem 1.1 [10], which gives an upper bound for the ε-entropy of the γ-critical values of differentiable functions. It is sufficient to consider only the γ-critical values of ϕ on the interior of U, the γ-critical values of the restriction of ϕ on ∂U can be treated exactly in the same way, and satisfy even better restrictions.

The following form of theorem 1.1 [10], which follows immediately from the original general one, is appropriate in our situation:

Theorem 3.4. Let $\Psi : B_r^N \to R$ be a k times continuously differentiable function, defined on the ball B_r^N of radius r in R^N. Denote by $R_k(\Psi)$ the number

$$R_k(\Psi) = \sup_{z \in B_r^N} \| d^k \Psi(z) \| \cdot r^k .$$

Let for $\gamma \geq 0$, $\Delta(\Psi, \gamma)$ be the set of γ-critical values of Ψ. Then for any $\varepsilon > 0$

$$M(\varepsilon, \Delta(\Psi, \gamma)) \leq \begin{cases} \bar{A}_o + \bar{A}_1 \gamma (\frac{r}{\varepsilon}), & \varepsilon \geq R_k(\Psi) \\ \bar{A}_o \left[\frac{R_k(\Psi)}{\varepsilon} \right]^{\frac{N}{k}} + \bar{A}_1 \gamma (\frac{r}{\varepsilon}) \left[\frac{R_k(\Psi)}{\varepsilon} \right]^{\frac{N-1}{k}}, & \varepsilon \leq R_k(\Psi), \end{cases}$$

where the constants $\bar{A}_o$, $\bar{A}_1$ depend only on N and k.

We want to apply this theorem to our function ϕ. First of all, the domain of ϕ is contained in the ball of radius K_1 in R^N, where $N = (m + 2)(n + 1) - 2$ and K_1 depends only on n and m. By the standard extension results, ϕ can be extended to the k-smooth function $\bar{\phi}$ on this ball $B_{K_1}^N$, with the uniform norm of the k-th derivative, not exceeding $K_2 C$ (where K_2 also depends only on n and m).

For the constant $R_k(\bar{\phi})$ we therefore obtain:

$R_k(\bar{\phi}) \leq K_2 C K_1^k = BC$, where $B = K_2 K_1^k$. Theorem 3.4 now gives:

$$M(\varepsilon, \Delta(\phi, \gamma)) \leq M(\varepsilon, \Delta(\bar{\phi}, \gamma)) \leq$$

$$\leq \begin{cases} \bar{A}_o + \bar{A}_1 \gamma (\frac{K_1}{\varepsilon}), & \varepsilon \geq BC, \\ \bar{A}_o (\frac{BC}{\varepsilon})^{\frac{N}{k}} + \bar{A}_1 (\frac{K_1}{\varepsilon}) (\frac{BC}{\varepsilon})^{\frac{N-1}{k}}, & \varepsilon \leq BC \end{cases} .$$

Combaining these estimates with the similar ones for the γ-critical values of ϕ on the boundary of U, and using lemma 3.1, we complete the proof of theorem 2.6.

4. Some consequences

In section 2 we used theorem 2.6 in a special case $\gamma = 0$ to derive the Morse-Sard theorem for maximum functions. In general, we can play with the values of parameters γ and ε in theorem 2.6 in order to obtain the required information about the set of almost critical values.

First of all, to simplify computations, we always assume below that $\varepsilon \leq BC$, and hence we can use only the second inequality of theorem 2.6.

Corollary 4.1. For any $f \in S^k_{n,\, m}$ (C), $\gamma \geq 0$ and $\varepsilon \leq BC$,

$$M (\varepsilon, \Delta (f, \gamma)) \leq C_1 (\frac{1}{\varepsilon})^{\frac{N}{k}} + C_2 \gamma (\frac{1}{\varepsilon})^{1 + \frac{N-1}{k}}$$

where the constants C_1 and C_2 depend only on n, m, k and C .

We also assume below, that $k > N$. Then the first term in the inequality of corollary 4.1 grows slowler than $\frac{1}{\varepsilon}$ as $\varepsilon \to 0$. The degree of $\frac{1}{\varepsilon}$ in the second term is $1 + \frac{N-1}{k} > 1$, but it contains the factor γ. Hence, taking γ positive, but sufficiently small, we still can obtain nontrivial bounds for $M (\varepsilon, \Delta (f, \gamma))$.

Let us obtain in this way the upper bound for the Lebesgue measure m of $\Delta (f, \gamma)$.

Corollary 4.2. For any $f \in S^k_{n,\, m}$ (C) and $\gamma \leq (BC)^{1-\frac{1}{k}}$

$$m\,(\Delta\,(f,\,\gamma)) \leq C_3\,\gamma^{\frac{k-N}{k-1}}\,.$$

where C_3 depends only on n, m, k, C.

Proof. Substitute in expression of corollary 4.1.

$\varepsilon = \gamma^{\frac{k}{k-1}}$. We obtain:

$$m\,(\,\Delta\,(f,\,\gamma)) \leq \varepsilon \cdot M\,(\varepsilon,\,\Delta\,(\,f,\,\gamma)) \leq$$

$$\leq \gamma^{\frac{k}{k-1}}\,[C_1\,(\tfrac{1}{\gamma})^{\frac{k}{k-1}\frac{N}{k}} + C_2\,\gamma\,(\tfrac{1}{\gamma})^{\frac{k}{k-1}(1+\frac{N-1}{k})}] =$$

$$= (C_1 + C_2)\,\gamma^{\frac{k-N}{k-1}} = C_3\,\gamma^{\frac{k-N}{k-1}}\,.$$

Since we assume $k > N$, the degree of γ in the inequality of corollary 4.2 is positive. Hence, taking γ sufficiently small, we obtain an arbitrarily small upper bound for $m\,(\,\Delta(f,\,\gamma))$. In particular, for $\gamma = 0$, we obtain once more the Morse-Sard theorem.

We mentioned already, that the information concerning the upper bounds for the Lebesgue measure, is not "effective". But the upper bounds for the ε-entropy of the set allow to find points out of this set very effectively. In fact, such points exist in any sufficiently dense regular net in R. We prove first one simple geometric lemma, which clarify the properties of the ε-entropy.

<u>Lemma 4.3.</u> Consider any regular 3ε-net x_i in R, $x_i = x_o + 3i\varepsilon$, $i = \ldots -1.\ 0,\ 1,\ \ldots$. Denote by U_i the open 2ε-interval, centered at x_i. Then any bounded set $A \subset R$ intersects at most $M(\varepsilon, A)$ intervals U_i.

<u>Proof.</u> Assume that A intersects intervals $U_{i_1}, \ldots, U_{i_l}$. Fix for any $j = 1, \ldots, l$ the point $z_j \in U_{i_j} \cap A$. Then $| z_{j_1} - z_{j_2} | > \varepsilon$ for any $j_1 \neq j_2$, and hence

$$M(\varepsilon, A) \geq M(\varepsilon, \{z_1, \ldots, z_l\}) = l$$

Combaining this lemma with corollary 4.1, we can find, for any given interval of the length L in R, such a small $\gamma_o > 0$ that for any $\gamma \leq \gamma_o$, there are points in our interval, not belonging to $\Delta(f, \gamma)$. Moreover, their whole neighborhoods of a fixed size do not intersect $\Delta(f, \gamma)$, and these points can be found in any sufficiently dense net. These last properties are especially important in computations with finite accuracy. To present these properties precisely, it is convenient to give the following definition:

<u>Definition 4.4.</u> For f as above and $\gamma \geq 0$ any point $c \in R \setminus \Delta(f, \gamma)$ is called γ-regular.

In other words, $c \in R$ is a γ-regular value of f, if at each point $x \in B^n$, with $f(x) = c$, the distance between $\partial f(x)$ and 0 is at least γ.

<u>Theorem 4.5.</u> Let $f \in S^k_{n,m}(C)$, $k > N$, and let $L > 0$ be given. Denote by $\gamma_o = \gamma_o(C, L)$ the min $[(L/6C_3)^{\frac{k-1}{k-N}}, (BC)^{1-\frac{1}{k}}]$, and let $\varepsilon_o = \gamma_o^{\frac{k}{k-1}}$.

Consider some regular $3\varepsilon_o$-net x_i, $x_i = x_o + 3i\varepsilon_o$, $i = \ldots - 1, 0, 1, \ldots$.

Then in any interval $V \supset R$ of length L there is at least one point x_j of our net, such that the whole ε_o-neighborhood of x_j consists of γ_o-regular values of f.

<u>Proof.</u> By corollary 4.1 we have

$$M(\varepsilon_o, \Delta(f, \gamma_o)) \leq C_1 \left(\frac{1}{\gamma_o}\right)^{\frac{k}{k-1} \cdot \frac{N}{k}} + C_2 \left(\frac{1}{\gamma_o}\right)^{\frac{k}{k-1}\left(1 + \frac{N-1}{k}\right) - 1} =$$

$$= (C_1 + C_2) \left(\frac{1}{\gamma_o}\right)^{\frac{N}{k-1}} = C_3 \left(\frac{1}{\gamma_o}\right)^{\frac{N}{k-1}}$$

Hence, the set $\Delta(f,\gamma_o)$ intersects at most $C_3 \left(\frac{1}{\gamma_o}\right)^{\frac{N}{k-1}}$ among the ε_o-neighborhood of points x_i, $i = \ldots -1, 0, 1, \ldots$, by lemma 4.3.

On the other hand, any interval $V \supset R$ of length L contains at least

$\frac{L}{3\varepsilon_o} - 1 > \frac{L}{6\varepsilon_o} = \frac{L}{6} \left(\frac{1}{\gamma_o}\right)^{\frac{k}{k-1}}$ points x_i, and since, by definition,

$\gamma_o \leq (L/6C_3)^{\frac{k-1}{k-N}}$, we have $\frac{L}{6} \left(\frac{1}{\gamma_o}\right)^{\frac{k}{k-1}} \geq C_3 \left(\frac{1}{\gamma_o}\right)^{\frac{N}{k-1}}$

Theorem is proved.

Taking smaller γ_o and ε_o, we can guarantee that an arbitrarily big part of the points of our net in any interval of length L are γ_o-regular values of f, together with their ε_o-neighborhoods.

We conclude with some remarks:

1. Considering the asymptotic behavior of the ε-entropy of critical values, as $\varepsilon \to 0$, one can obtain, by theorem 2.6, the upper bound for the entropy dimension of the set of critical values (which is, roughly, the highest degree of $\frac{1}{\varepsilon}$ in the asymptotic expression for $M(\varepsilon, \Delta(f, 0))$. Compare [10] and [12].)

2. The restrictions on the geometry of critical and almost-critical values, given by theorem 2.6, form a set of necessary conditions for the representability of a given function as a maximum of a smooth family. Using these conditions, and especially the upper bounds for the entropy dimension, one can easily construct explicit examples of "nice" functions, not representable as maxima of "too smooth" families. Compare [12], theorem 3.1. (where only "smooth" critical points are considered).

3. The results of this paper can be extended in several directions. First of all, we can consider functions, representable as differences of functions from $S^k_{n,m}$ (such an extension is natural and important, see e. g. [5], [7]). Secondly, some information on the geometry of critical values can be obtained for minimax functions of smooth families. Finally, we can consider the mappings into R^q (instead of functions, i. e. mappings into R), whose coordinates belong to $S^k_{n,m}$.

However, the proper setting of these results requires a deeper insight into the structure of maximum functions, in particular, some additional information on the behavior of these functions under usual arithmetic and analytic operations.

4. There is an important class of Lipschitzian functions, closely related to the maxima of smooth families. It consists of functions, representable as the difference of two convex ones (dc-functions; see e. g. [5], [7], [3]). It turns out, that for dc-functions of one and two variables the Morse-Sard theorem is still valid (see [3]). However, for d. c. functions of 3 and more variables the critical values can cover the whole interval.

R E F E R E N C E S

1. F. M. Clarke, Generalized gradients and applications, Trans. Amer. Math. Soc., 205 (1975), 247-262.
2. A. N. Kolmogorov, V. M. Tihomirov, ε-entropy and ε-capacity of sets in functional spaces. Usp. Mat. Nank 2 (86), 14 (1959), 3-86, Amer. Math. Soc. Transl. (2), 17 (1961), 277-364.
3. E. M. Landis, On functions, representable as the difference of two convex functions, Dokl. Akad. Nank SSSR, 80, 1 (1951), 9-11.
4. F. Mignot, Centrole dans les inêquations variationelles elliptiques, I. Funct. Anal,. 22, 2 (1976), 130-185.
5. R. T. Rockafellar, Favorable classes of Lipschitz continuous functions in subgradient optimization, in "Progress in nondifferentiable optimization", E. Nurminski, ed., Pergamon Press, 1981.
6. A. Sard, The measure of critical values of differentiable maps, Bull. Amer. Math. Soc., 48 (1942), 883-890.
7. A. Shapiro, Y. Yomdin, On functions, representable as a difference of two convex functions, and necessary conditions in a constrained optimigation, preprint, 1981.
8. C. Tricot, Douze définitions de la densité logarithmique, C. R. Acad. Sc., 193 (1981), 549-552.
9. M. Whitney, A function non constant on a connected set of critical points, Duke Math. I. 1 (1935), 514-517.

10. Y. Yomdin, The geometry of critical and near-critical values of differentiable mappings, Math. Ann., 264 (1983), 495-515.

11. Y. Yomdin, On functions representable as a supremum of a family of smooth functions, SIAM J. Math. An.al. 14, 2 (1983), 239-246.

12. Y. Yomdin, On functions, representable as a supremum of a family of smooth functions, II, to appear.