JOURNAL OF APPROXIMATION THEORY 76, 166–176 (1994)

On the ε-Entropy of Nearly Critical Values

ANDREA ROHDE

RWTH Aachen, Lehrstuhl C für Mathematik, Templergraben 55, D-52056 Aachen, Germany

Communicated by Paul L. Butzer

Received February 18, 1992; accepted in revised form August 17, 1992

Let f be a C^k-mapping into $\mathbb{R}^m$, restricted to a compact subset of $\mathbb{R}^n$. We compute an upper bound on the ε-entropy of the nearly critical values. This refines some results of Yomdin. The bound depends only on dimensions, the differentiability degree k, and estimates of the norms of the first and kth derivative of f.

1. INTRODUCTION, NOTATIONS, AND MAIN RESULT

Critical points and values play an important role in applied mathematics. In terms of ε-entropy we get more information about gaps between critical values than by the well-known Morse–Sard theorem. In case that the image space is only one-dimensional, the situation becomes quite easy (cf. [1, Theorem 2.5.11]). A numerical application of the one-dimensional version can be found in [8]. Let R be a positive real number, $n \geqslant m$, and $k \geqslant 2$. Throughout this paper f: $[0, R]^n \to \mathbb{R}^m$ will be a mapping which extends to a C^k-mapping in some open neighbourhood of the cube $[0, R]^n$. The only information we assume to have about f are the following two bounds

$$C_1 := \max\{\|Df(x)\|_{\max} \mid x \in [0, R]^n\}$$

and

$$C_2 := \max\{\|f^{(k)}(x)\|_{\max} \mid x \in [0, R]^n\},$$

with $\|\cdot\|_{\max}$ the operator norm induced by the maximum norm. In contrast, we denote by $\|\cdot\|$ the Euclidian norm and the induced operator norm. By $B_{\max}(z, r)$ and $B(z, r)$ we mean closed balls with center z and radius r in the maximum and Euclidian norm respectively.

Let $A \subset \mathbb{R}^n$ be a *bounded* set and let $N(\varepsilon, A)$ be the *minimal* number of balls of radius ε (in the maximum norm) necessary to cover A. The base 2 logarithm of $N(\varepsilon, A)$ is called the *ε-entropy* of A (cf. [3]).

0021-9045/94 $6.00

For the definition of nearly critical points and values, we need the well-known concept of singular values (cf. [2]). In particular, for ρ_m the smallest singular value of a matrix $M \in \mathbb{R}^{m \times n}$, $n \geqslant m$, we have

$$\rho_m = \min\{ \|v^T M\| \mid v \in \mathbb{R}^m \text{ and } \|v\| = 1 \}.$$

For $\lambda \geqslant 0$ and f as above we can now define (cf. Yomdin [6]):

- x is a λ-critical point of f, iff the smallest singular value of $Df(x)$ is less or equal λ and
- y is a λ-critical value of f, iff there exists a λ-critical point x in the domain of f such that $y = f(x)$.

For λ equal to zero this gives obviously the usual definition of critical points and values. For λ *near* zero λ-critical points are *nearly* critical. Nearly critical points may appear to be critical under only finite accuracy of computation. Under small perturbations of f nearly critical points remain nearly critical with a possibly larger λ.

MAIN THEOREM. *For f: $[0, R]^n \to \mathbb{R}^m$ a C^k-mapping, with $k \geqslant 2$, $n \geqslant m$, C_1, C_2 as above and $C_2 \neq 0$ we need at most*

$$n^2 m^4 6^{n+m} k^{n+m} (1 + R/(2r))^n \, (1 + 2Cr/\varepsilon)^{m-1} \, (1 + k + 2r\lambda(\varepsilon)/\varepsilon)$$

ε-balls in the maximum norm to cover all $\lambda(\varepsilon)$-critical values of f restricted to $[0, R]^n$ with

$$r = r(\varepsilon) := \sqrt[k]{\varepsilon}\, \sqrt[k]{k!/(2C_2)} \qquad \textit{and} \qquad C := C_1 + C_2 r^{k-1}/(k-1)!.$$

Interpretation. For $\lambda(\varepsilon) \leqslant \varepsilon$, ε tending to zero, the bound above roughly grows like $(1/\varepsilon)^{m-1+(n-m+1)/k}$. If we have $k > n - m + 1$, then the total volume of the covering tends to zero and there appear *gaps* between the cubes of the covering, where only "numerically regular" values can be. So one can easily find in any fixed area of the image space a finite lattice and some $\lambda > 0$ such that at least one of the nodes is not a λ-critical value (or at least 50% etc.). For this purpose we need the *discrete* bound, not only its rate of increase (in the form of the fractal or entropy *dimension*) for ε passing to zero. A special remark on the case $k = n - m + 1$ is made at the end of Section 2. If C_2 is equal to zero, then f is a polynomial. In this case a better and more simple bound holds (cf. estimate (11)).

Remark. By the use of auxiliary variables for the approximation of an eigenvector in (3) (in Section 3; see below) we differ considerably from the approach of Yomdin [6, 7], where subdeterminants are used. Via determinants one obtains a bound where essentially the factor k^{n+m} in the final result is replaced by $(mk)^n$, which is larger, if k is small.

In the rest of the paper we prove the main theorem.

Outline of the proof. In Section 2 we replace f locally on $B_{\max}(z_0, r)$ by its Taylor polynomial $p = (p_1, ..., p_m)^T$ and see that the nearly critical points and values of p correspond to those of f. So we can compute the bound of the main theorem via a bound on the ε-entropy of the nearly critical values of p restricted to $B_{\max}(z_0, r)$.

To this aim we give in Section 3 a semi-algebraic description of the nearly critical points of p and partition them in sets $M_1, ..., M_m$ with respect to the direction of an eigenvector to the smallest eigenvalue of $Dp(x)\, Dp(x)^T$. The rest of the proof essentially gives a quantitive description of the fact that $p_i(M_i)$ cannot spread out independently of the other components.

In Section 4 we are concerned with the number of connected components of a semi-algebraic sets, with the construction of auxiliary curves and estimates of their length. These considerations are used in Section 5 to complete the computation of Section 3. Last but not least, the polynomial result is transferred back to the original function f.

2. Nearly Critical Points of the Taylor Polynomial

As f is a mapping into $\mathbb{R}^m$, we apply Taylor's formula *componentwise*. If we choose for $p = (p_1, ..., p_m)^T$ the Taylor polynomial of f *of degree* $k-1$ in z_0, we have for $\|x - z_0\|_{\max} \leqslant r$

$$\|p(x) - f(x)\|_{\max} \leqslant C_2 r^k/k!. \tag{1}$$

Then the comparision of the Taylor expansion of the directional derivative $Df(x)\, z$, x variable, z fixed, around z_0, and the directional derivative of the Taylor polynomial with respect to z gives an estimate of $\|Dp(x) - Df(x)\|_{\max}$ similar to (1), with k replaced by $k-1$. Now let us denote by $\rho_m^p(x)$, $\rho_m^f(x)$ the smallest singular values of $Dp(x)$ and $Df(x)$, respectively. Application of both mappings to such vectors of the sphere that the minimal norm of the result is attained gives for the smallest singular values

$$|\rho_m^p(x) - \rho_m^f(x)| \leqslant \|Dp(x) - Df(x)\| \leqslant C_2 \sqrt{n}\, r^{k-1}/(k-1)!. \tag{2}$$

If C_2 is not equal to zero, we define $r := \sqrt[k]{\varepsilon}\, \sqrt[k]{k!/(2C_2)}$. Then the approximational error of (1) is at most $\varepsilon/2$ on $B_{\max}(z_0, r)$. We cover

$[0, R]^n$ by at most $(1 + R/(2r))^n$ cubes of edge length $2r$ and study locally the nearly critical points and values of the Taylor polynomial.

Remark. The Taylor polynomial of degree k gives a better estimate of f. In this case one can replace C_2 in the inequalities of this section by the module on continuity of $f^{(k)}$. Then one can choose r a little bit larger and obtain, quite analogously to the quantitative result, the hard case of Sard's theorem with minimal differentiability. However, as we did not assume any quantitative knowledge about this module of continuity, we choose the Taylor polynomial of lower degree, which is more simple from the semi-algebraic point of view.

3. Semi-Algebraic Sorting

For each vector of the standard basis of the image space we sort out those critical points, for which the corresponding ellipsoid $Dp(x)(S^{n-1})$ is thin in this direction. This sorting must be done by *polynomial* inequalities to apply results from real algebraic geometry later on.

For this purpose we consider the eigenvalues of $A(x) := Dp(x)\, Dp(x)^T$ instead of the singular values of $Dp(x)$. The smallest eigenvalue of $A(x)$ is less or equal $(\lambda(\varepsilon))^2$ and there is a corresponding eigenvector pointing in a direction "close" to the jth coordinate axis, if and only if

$$\exists\, y \in \mathbb{R}^m \setminus \{0\} : y^T A(x)\, y \leqslant (\lambda(\varepsilon))^2\, \|y\|^2 \qquad \text{and} \qquad y_j^2 \geqslant \|y\|^2/m. \tag{3}$$

The second condition is true for any y for at least one $j \in \{1, ..., m\}$. Due to the homogenity of (3) we may assume that y_j is equal to one and define

$$\begin{aligned} M_i := \{(x, u) \in \mathbb{R}^{n+m-1} \mid\ & v^i(u)^T A(x)\, v^i(u) \leqslant \lambda(\varepsilon)^2\, (1 + \|u\|^2) \\ & \text{and } \|u\|^2 \leqslant m - 1\} \end{aligned}$$

with

$$v^i((u_1, ..., u_{m-1})^T) := (u_1, ..., u_{i-1}, 1, u_i, ..., u_{m-1})^T \in \mathbb{R}^m.$$

Let $\pi_x : \mathbb{R}^{n+m-1} \to \mathbb{R}^n$ be the projection to the first n components; then $\pi_x(\bigcup_{i=1}^m M_i)$ is just the set of all $\lambda(\varepsilon)$-critical values of p. The conditions (3) are polynomial conditions, so M_i is a semi-algebraic set.

We see below that $p_i(M_i)$ cannot spread out *independently* of the other components by the special construction of M_i. To this aim we consider directional derivatives and obtain eventually the inequality (4) below. Via integration over semi-algebraic curves, volume estimates are obtained from this inequality. Only by (4) we make use of the definition of nearly critical points and values.

Let π_i^e be the projection eliminating the ith component.

LEMMA 3.1. *For M_i as above, $x \in \pi_x(M_i)$ and $z \in \mathbb{R}^n$ arbitrary we have*

$$|Dp_i(x)\, z| \leqslant m^{3/2}\, \|\pi_i^e(Dp(x)\, z)\| + (1 + m^{3/2})\, m^{3/2} \lambda(\varepsilon)\, \|z\|. \tag{4}$$

Proof. The proof splits into three steps:

• First we need orthonormal bases, which we obtain by the singular value decomposition of $Dp(x)$ (cf. [2]). There are $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ orthogonal, $\rho_1 \geqslant \cdots \geqslant \rho_m \geqslant 0$ such that

$$Dp(x) = U \begin{bmatrix} \rho_1 & 0 & \cdots & 0 \\ & \ddots & & 0 \\ 0 & & \rho_m & 0 \end{bmatrix} V \quad \text{(SVD)},$$

We can transfer a weakened version of the properties of y in the definition of M_i to one of the columns of U: For $U = (u_{ij})$ let $u_j = (u_{1j}, ..., u_{mj})^T$ be the jth column of U. If $x \in M_i$, then there exists $y = (y_1, ..., y_m)^T \in \mathbb{R}^m$, with $\|y\| = 1$, $y^T A(x)\, y \leqslant (\lambda(\varepsilon))^2$ and $y_i^2 \geqslant 1/m$. In the orthonormal basis $(u_1, ..., u_m)$ we have the representation $y = \sum_{j=1}^m b_j u_j$. Due to $|y_i| \geqslant m^{-1/2}$ we can find ν in $\{1, ..., m\}$ such that both $|b_\nu| \geqslant m^{-3/2}$ and $|u_{i\nu}| \geqslant m^{-3/2}$. As a consequence

$$b_\nu^2 \cdot \rho_\nu^2 \leqslant y^T A(x)\, y \leqslant \lambda(\varepsilon)^2 \qquad \text{implies} \qquad \rho_\nu \leqslant m^{3/2} \cdot \lambda(\varepsilon).$$

So u_ν is the column of U we consider instead of y with the important properties

$$\|u_{i\nu}\| \geqslant m^{-3/2}, \qquad \rho_\nu \leqslant m^{3/2} \lambda(\varepsilon), \qquad (u_1, ..., u_n) \quad \text{an orthonormal basis.} \tag{5}$$

• Second, using Schwarz's inequality one easily shows that for $b \in \mathbb{R}^m$, $\mu \in (0, 1)$, $j \in \{1, ..., m\}$ with $\|b\| = 1$, $u_i^2 \geqslant \mu^2$ holds

$$\forall v \in \mathbb{R}^m\colon (v \perp b \Rightarrow \|\pi_i^e(v)\| \geqslant \mu\, \|v\|).$$

For the special case of (5) this yields

$$\forall v \in \langle u_1, ..., u_{\nu-1}, u_{\nu+1}, ..., u_n \rangle_{\mathbb{R}}\colon \|\pi_i^e(v)\| \geqslant m^{-3/2}\, \|v\|. \tag{6}$$

• Summing up the previous considerations we have by (6) and (5) for $w := Dp(x)\, z = (w_1, ..., w_n)^T = \sum_{j=1}^m \hat{w}_j u_j$

$$\left\| \sum_{j=1, j \neq j_0}^{m} \hat{w}_j u_j \right\| \leqslant m^{3/2} (\|\pi_i^e(Dp(x) \cdot z)\| + \|\pi_i^e(w_\nu u_\nu)\|)$$

and

$$\|w_\nu u_\nu\| + m^{3/2} \|\pi_i^e(w_\nu u_\nu)\| \leqslant (1 + m^{3/2}) m^{3/2} \lambda(\varepsilon) \|z\|.$$

By triangle inequality and elementary transformations we obtain the desired result. ∎

For regularity conditions in Section 4 we need a version of (4) under little perturbation of x. So we define

$$\omega(\Delta, z_0, r) := \sup\{ \|Dp(x) - Dp(y)\| \mid \|x - y\| \leqslant \Delta, x \in B(z_0, r), y \in \mathbb{R}^n \}.$$

Then we have for x in a Euclidian Δ-neighbourhood of $\pi_x(M_i) \cap B(z_0, r)$

$$|Dp_i(x) z| \leqslant m^{3/2} \|\pi_i^e(Dp(x) z)\| + (1 + m^{3/2}) m^{3/2} (\lambda(\varepsilon) + \omega(\Delta, z_0, r)) \|z\|.$$

with

$$\omega(\Delta, z_0, r) \to 0 \qquad \text{for} \quad \Delta \to 0. \tag{7}$$

This is due to the fact that for $\|Dp(x) - Dp(y)\| \leqslant \omega$ we have for the projections that both $\|Dp_i(x) - Dp_i(y)\| \leqslant \omega$ and $\|\pi_i^e(Dp(x)) - \pi_i^e(Dp(y))\| \leqslant \omega$.

Now we begin with the construction of the covering. Without loss of generality we assume that $i = 1$.

First we take care of those directions of the image space for which we have no special restrictions, only the general bound to the first derivative of p. We can cover

$$(p_2, ..., p_m)^T ([0, R]^n \cap B_{\max}(z_0, r))$$

with $(m-1)$-dimensional (2ε)-cubes, only using the bound

$$\forall x \in [0, R]^n \cap B_{\max}(z_0, r): \|Dp(x)\|_{\max} \leqslant C = C_1 + C_2 r^{k-1} (k-1)!. \tag{8}$$

We need at most $(Cr\varepsilon^{-1} + 1)^{m-1}$ such cubes.

The difficult part is the estimate for the first component. Let $\square$ be one of the cubes fixed. We consider the image under p_1 of those nearly critical points of p in $B_{\max}(z_0, r)$ which are mapped into $\square$ by $(p_2, ..., p_n)$. The number of ε-intervals we need for this covering gives the last factor for the desired estimate.

We can write $\square$ in the form $\square = B_{\max}(y_0, \varepsilon)$, $y_0 = (y_{2,0}, ..., y_{m,0})^T$. To reduce the number of polynomial conditions, we replace the cubes by

Euclidian balls just including them. So the subset of nearly critical points we are interested in has the form

$$M_1(\square, \varepsilon) := \Big\{(x, u) \in \mathbb{R}^{n+m-1} \mid (1, u^T) A(x)(1, u^T)^T \leqslant (\lambda(\varepsilon))^2 (1 + \|u\|^2),$$

$$\sum_{i=1}^{n} (x_i - z_{i,0})^2 \leqslant r^2 n, \sum_{j=2}^{m} (p_j(x) - y_{j,0})^2 \leqslant \varepsilon^2(m-1)$$

$$\text{and } \|u\|^2 \leqslant m-1\Big\} \tag{9}$$

with $z_0 = (z_{1,0}, ..., z_{n,0})^T$. Then we have for $M := \pi_x(M_1(\square, \varepsilon))$

$$M \supseteq \pi_x(M_1) \cap [0, R]^n \cap B_{\max}(z_0, r) \cap (p_2, ..., p_m)^{-1}(\square),$$

and it suffices to cover $p_1(M)$ with (2ε)-intervals. To this aim we need

- bounds on the number of connected components of $p_1(M)$ and
- bounds on the total length of $p_1(M)$.

The number of connected components of $p_1(M)$ is less than or equal to the number of connected components of the bounded semi-algebraic set $M_1(\square, \varepsilon)$. This number is bounded independently of ε, $\square$ (cf. Section 4). To show that the total length of $p_1(M)$ tends to zero for ε, $\lambda(\varepsilon) \to 0$ we need the special choice of M_1 in (9) via (3). For any set $S \subset \mathbb{R}^n$ let $[S]_\Delta$ be the Δ-neighbourhood of S. We replace M by a family of semi-algebraic C^∞-curves, $(s^\Delta)_{\Delta>0}$, with uniformly bounded number of connected components, $s^\Delta \subset [M]_\Delta$ and $p_1(M) \subset [p_1(s^\Delta)]_\Delta$. We obtain

$$\text{total length } (p_1(M)) \leqslant \limsup_{\Delta \to 0} \text{length } (p_1(s^\Delta)).$$

Now application of (7) to $Dp_1(s^\Delta(t))\, \dot{s}^\Delta(t)$ yields

$$\text{length}(p_1(s^\Delta)) \leqslant \int_0^1 m^{3/2} \|(Dp_2(s^\Delta(t)), ..., Dp_m(s^\Delta(t)))^T \dot{s}^\Delta(t)\| \, dt$$

$$+ \int_0^1 (1 + m^{3/2})\, m^{3/2}(\lambda(\varepsilon) + \omega(\Delta, z_0, r))\, \|\dot{s}^\Delta(t)\| \, dt. \tag{10}$$

So the estimate of the total length of $p_1(M)$ is reduced to the construction of the $(s^\Delta)_{\Delta>0}$ and estimates of the lengths of the curves (s^Δ) and $((p_2, ..., p_m)^T (s^\Delta))$ for $\Delta \to 0$.

For length (s^Δ) a constant bound independent of Δ is sufficient, as $\lambda(\varepsilon)$ is small. The length of $((p_2, ..., p_m)^T (s^\Delta))$ tends to zero with ε, as the curve moves inside a ball of diameter $2\varepsilon\sqrt{m-1}$ and the number of changes of direction of a semi-algebraic curve is bounded in polynomial degrees only, i.e., independent of Δ (cf. Lemma 4.3).

4. Semi-algebraic Estimates

Let A be a semi-algebraic set of the form

$$A=\bigcap_{j=1}^{a}\{x\in\mathbb{R}^q \mid p_j(x)\geqslant 0\},\qquad p_j\in\mathbb{R}[x_1,...,x_q],\qquad d:=\sum_{j=1}^{a}\operatorname{degree}(p_j).$$

We call $x\in\mathbb{R}^q$ a *non-degenerate* solution of the system $f_1(x)=\cdots=f_m(x)=0$, $m\leqslant q$, if and only if $Df(x)$ is surjective with $f:=(f_1,...,f_m)$.

In the next two lemmata the following idea (cf. [4]) is used: For a (polynomial) manifold $\{x\in\mathbb{R}^q \mid p(x)=0\}$ one may choose a generic linear form l, so that all critical points of l restricted to the manifold are non-degenerate. These critical points are non-degenerate solutions of the polynomial system obtained by the Lagrange multiplier rule. The number of such solutions is bounded by the product of the degrees in virtue of the real version of Bezout's theorem (cf. [1]).

Lemma 4.2 (cf. [7], Theorem 4.5). *For A as above the number of bounded connected components is at most $(\frac{1}{2})\,d(d-1)^q$.*

Lemma 4.3. *For A as above, A compact, $\tilde{p}\in\mathbb{R}[x_0,...,x_q]$ of degree $\tilde{d}$ we obtain a family of semi-algebraic C^∞-curves, $(s^\Delta)_{\Delta>0}$, $s^\Delta=(s_1^\Delta,...,s_q^\Delta)^T$ with*

- *$s^\Delta\subset[A]_\Delta$, $\tilde{p}(A)\subset[\tilde{p}(s^\Delta)]_\Delta$ and the number and degrees of the polynomials defining the curves independent of Δ*
- *s_i^Δ attains a generic value at most $(d+2\tilde{d})^q$ times for all $1\leqslant i\leqslant q$.*

For any polynomial $p_0\in\mathbb{R}[x_1,...,x_q]$ of degree d_0 we further have

- *$p_0\circ s^\Delta$ attains a generic value at most $(d+2\tilde{d})^q\,d_0$ times.*

Proof. We make use of some ideas to be found in [7]. However, the bounds there are very much too large for our special case and do not include the bound for the curve of $p_0\circ s^\Delta$. So we give a sketch of the proof here.

Construction of s^Δ. In fact Lemma 4.3 is based on the same ideas as Lemma 4.2, only an additional parametrisation (in the image of $\tilde{p}$) takes place. The graph of $\tilde{p}\,|_A$ is approximated by *some components* of a polynomial *manifold* $G:=\{(x,y)\in\mathbb{R}^{q+1} \mid p(x,y)=\eta\}$, where p is the product of the slightly perturbed polynomials p_j of the definition of A and $\tilde{p} \pm \varepsilon$. Milnors idea is applied to all but finitely many cross sections $G_y = \{x \in \mathbb{R}^q \mid p(x, y) = \eta\}$ restricted to the relevant components of G.

The choice of η and l uniformly for all but finitely many y is possible essentially with the following ideas. First we have that η and l may be chosen generically for fixed y. Second, by the compactness of A, the requirements for Milnor's approach and Bezout's theorem are stable under little perturbations in all variables (in y, too). Third, by representation as a semi-algebraic set, we get a bound to the number of connected components of the set of y values without the desired properties uniformly in η, l (cf. Lemma 4.2).

With the parameter y the critical points of l on each cross section form a curve c^{Δ} and the projection $\pi_x(c^{\Delta})$ gives s^{Δ}. From the special choice of l in Milnor's approach we have that the implicit function theorem can be applied—so we obtain C^{∞}-functions.

Intersection multiplicities. Now the condition that s_i^{Δ} or $\tilde{p}(s^{\Delta})$ attains a certain value simply gives the last polynomial equation to the system of equations describing the implicit function c^{Δ} and one can apply Bezout's theorem *directly*. By this simple idea we reduce the estimate of [7] by a factor *exponential in* q.

It is easy to show that the 2-tuples of regular points and corresponding values of the curves s_i^{Δ} and $p_0(s^{\Delta})$ respectively are non-degenerate solutions of the system of equations above. The regular values of the C^{∞}-functions are dense by Sard's theorem (cf. [5]). As A is assumed to be compact and bounded and the semi-algebraic curves are smoothly extendable [1], so the set of regular values is open, too. ∎

From the intersection multiplicities and the diameter of A, $p_0(A)$ one easily obtains estimates for the lengths of the curves s^{Δ} and $p_0 \circ s^{\Delta}$: Let $s\colon [0, 1] \to \mathbb{R}^n$, $s = (s_1, ..., s_n)$, be a C^1-function. Further assume that s_i attains a generic value at most h_i times and $s_i(t) \in [a, b]$ for all $t \in [0, 1]$; then

$$\text{length}(s) = \int_0^1 \|\dot{s}(t)\| \, dt \leqslant \sum_{i=1}^{n} \int_0^1 |\dot{s}_i(t)| \, dt$$

and

$$\int_0^1 |\dot{s}_i(t)| \, dt \leqslant h_i(b-a).$$

This remains true, of course, if s is only piecewise differentiable and not even continuous. In particular the observation above allows the computation of the length of any projection curve of s.

5. Final Evaluation

Now we choose for A the set $M_1(\square, \varepsilon)$, for $\tilde{p}$ the polynomial p_1 and for p_0 the polynomials $p_2, ..., p_m$ (cf. Sections 2 and 3). Then we have $d=4k$, $q=n+m-1$. Construction of s^Δ according to Lemma 4.3 gives a curve with the properties necessary to continue the estimate (10). We get

$$\limsup_{\Delta \to 0} \operatorname{length}(s^\Delta) \leqslant n \cdot 2r\sqrt{n} \cdot (6k)^{n+m-1}$$

and

$$\limsup_{\Delta \to 0} \operatorname{length}((p_2, ..., p_m)^T \circ s^\Delta) \leqslant (m-1)\, 2\varepsilon \sqrt{m-1}\, 6^{n+m-1} k^{n+m}.$$

So we have by (10) and elementary transformations

$$\text{total length}(p_1(M)) \leqslant m^3 6^{n+m} k^{n+m} (\varepsilon + n^{3/2} r \lambda(\varepsilon)).$$

Lemma 4.2 gives $(\frac{1}{2})(4k)^{n+m-1}$ as an upper bound for the number of connected components of $M_1(\square, \varepsilon)$. Projection and application of p_1 only decrease the number of connected components, so the bound holds for $p_1(M)$, too. We divide the bound on the total length by 2ε and add the bound for the number of connected components. This gives a bound for the number of 2ε-intervals necessary to cover $p_1(M)$. Multiplication with $(Cr\varepsilon^{-1}+1)^{m-1}$ (for all possible choices of $\square$) yields a bound to the number of 2ε-cubes necessary to cover $p(\pi_x(M_1) \cap B_{\max}(z_0, r))$ (cf. (8) and (9)). Due to symmetry the bound for M_1 holds for M_j, $2 \leqslant j \leqslant m$, too. So we need at most

$$m^4 6^{n+m} k^{n+m} (Cr\varepsilon^{-1}+1)^{m-1} (1 + n^{3/2} r \lambda(\varepsilon)/\varepsilon) \tag{11}$$

ε-balls in the maximum norm to cover all $\lambda(\varepsilon)$-critical values of p on $B_{\max}(z_0, r)$.

Now we return to our original function f. By the choice of $r = r(\varepsilon)$ in Section 2 we get a covering of all $\lambda(\varepsilon)$-critical values of f restricted to $B_{\max}(z_0, r)$ by covering all $(\lambda(\varepsilon) + \varepsilon k\sqrt{n}/(2r))$-critical values of p with $\varepsilon/2$-balls and blow each ball up to an ε-ball. Replacing ε, $\lambda(\varepsilon)$ by $\varepsilon/2$ and $\lambda(\varepsilon) + \varepsilon k\sqrt{n}/(2r)$, respectively, we can therefore use the polynomial result (11) with an additional factor for the number of r-balls in the maximum norm necessary for a covering of $[0, R]^n$, the original domain of f. This yields the result of the main theorem.

ACKNOWLEDGMENT

I thank H. Th. Jongen for suggesting this research and for very useful discussions.

REFERENCES

1. R. BENEDETTI AND J.-J. RISLER, "Real Algebraic and Semi-algebraic Sets," Hermann, Paris, 1990.
2. G. H. GOLUB AND C. F. VAN LOAN, "Matrix Computations," Johns Hopkins Univ. Press, Baltimore, 1983.
3. A. N. KOLMOGOROV AND V. M. TIHOMIROV, ε-entropy and ε-capacity of sets in functional spaces, *Amer. Math. Soc. Transl. Ser. 2* **17** (1961), 277–364.
4. J. W. MILNOR, On the Betti numbers of real varieties, *Proc. Amer. Math.* **15**, No. 2 (1964), 275–280.
5. J. W. MILNOR, "Topology from the Differentiable Viewpoint," Univ. Press of Virginia, Charlottesville, VA, 1965.
6. Y. YOMDIN, The geometry of critical and near-critical values of differentiable mappings, *Math. Ann.* **264** (1983), 495–515.
7. Y. YOMDIN, Metric semialgebraic geometry with applications in smooth analysis, preprint, 1988.
8. Y. YOMDIN, Sard's theorem and its improved versions in numerical analysis, *in* "Computational Solution of Nonlinear Systems of Equations" (E. L. Allgower and K. Georg, Eds.), pp. 701–706, Lectures in Applied Mathematics (AMS), Vol. 26, Amer. Math. Soc., Providence, RI, 1990.