WEAK DISCONTINUITIES AND DISCRETE APPROXIMATIONS

OCTAV CORNEA

All the functions that can be exactly represented on a fixed computer take their values from a given discrete set. In our paper we will study, from the analysis point of view, some discrete phenomena closely related to the problem of approximating a continuous function on computer.

In the first section of the paper we will establish some necessary technical results. For a continuous function f we will point out the existence of an approximation f_A, which has some special properties. Some of them are related to the concept of weak discontinuity of the first species, concept necessary in our study. (A is the given discrete set containing the range of f_A). In the second part of the paper we prove that the natural approximations of f are converging (but not in the usual metric) to f_A. Examinating closely this process we arrive at a „uselessness" principle concerning the computational results.

Definitions and notations

Let A be a subset of a topological space X. We will denote by $a(A)$ its closure, by $Int(A)$ its interior and by $B(A)$ its boundary. Let $B \subset X$. The symmetric difference between A and B will be designated by $A \Delta B$. For $A \subset \mathbf{R}^n$ let $m'(A)$ be its exterior measure. Let E, F be two metric spaces. The point $x \in E$ is a *weak discontinuity of the first species* for a function $f : E \to F$ if : 1. x is a discontinuity point for f; 2. There exists a neighbourhood V of x and $n_x \in \mathbf{N}^*$ such that it is possible to decompose the set $V - \{x\}$ in n_x subsets P_i, $1 \leqslant i \leqslant n_x$, with the properties :

a. $$\cup\,(P_i;\ 1 \leqslant i \leqslant n_x) = V - \{x\};\ P_i \cap P_j = \emptyset\ (i \neq j)$$

b. For any $i \in \mathbf{N}$, $1 \leqslant i \leqslant n_x$ we have $\mathrm{Int}(a(P_i)) - \{x\} \subset P_i$ and there exists $y_i \in F$ such that for every sequence $\{x_n \in P : n \leqslant 1\}$ with

$$\lim(x_n;\ n \to \infty) = x \text{ we have } \lim\,(f(x_n);\ n \to \infty) = y_i.$$

c. For $i \neq j$ we have $y_i \neq y_j$. The set of the weak discontinuities of the first species of f will be denoted by D_f. A function $f : E \to F$ is said to be *closed* if the image of each closed set is a closed set. Let $M \subset \mathbf{R}$ be a finite set. For the function $f : E \to \mathbf{R}$ let's define the monotone downwardly directed rounding Df, and the monotone upwardly directed rounding Uf, as follows (see [1] page 31) :

$$Df : E \to M,\ Df(x) = \sup\{y \in M : y \leqslant f(x)\}$$

$$Uf : E \to M,\ Uf(x) = \inf\{y \in M : y \geqslant f(x)\}$$

The function $g : E \to M$ has the property S_f at the point $x \in E$ if $|f(x) - g(x)| = \min(|f(x) - Uf(x)|,\ |f(x) - Df(x)|)$. A function which

has the property S_f at each point of E will be named a *M-strong discrete approximation* (*M-s.d.a.*) *of* f. Let $B \subset \mathbf{R}^n$ be a compact set. A *division* of B is a finite class of subsets of B, pairwise disjoint, whose union covers B. Let C be a division of B. We will denote by $\|C\|$ the maximal diameter of the elements of C. Now let $\{C_n : n \leqslant 1\}$ be a sequence of division of B. We will say that this sequence *converges* to 0 if the sequence $\|C_n\|$ converges to 0 and for every n, each element of C_{n+1} is contained in an element of C_n. Let $C = \{B_i : i \leqslant n\}$ be a division of B and let $f : B \to R$. A function $g : B \to M$ such that there exists $a_i \in B_i$, $i \leqslant n$ where g has the property S_f and g is constant on B_i, will be named an *M-discrete approximation* (*M-d.a*) of associated to C. The points a_i will be named *test points* of g. For $f : E \to \mathbf{R}$ and $x \in E$ we will denote by $w(f, x)$ the oscillation of f at x.

Remarks. 1. The concept of d.a. of f models the natural representation of f on computer. Among all the functions which take values in M a M-s.d a. of f is one of the best possible approximations from the viewpoint of the natural distance.

2. For function of real variable the concept of weak discontinuity of the first species is more general than that of discontinuity of the first species. This designation was suggested to us by Professor Solomon Marcus.

3. It is possible to define the weak discontinuities of the first species and to prove some of the following results using spaces more general than the metric ones but this does not interest us here.

4. Some results concerning closed functions and related (but not very closely) to the following discussion may be found in [2], [4].

Examples. Let $I = \{(x, y) \in \mathbf{R}^2 : |x| < 2, \ |y| < 2\{$, $f : I \to [-2,2]$ $f(x, y) = (y + 2)/2$ if $x > 0$ and $f(x, y) = (y - 2)/2$ if $x \leqslant 0$ and let $M = \{-1, 0, 1\}$. A M-s.d.a. of f is $g : I \to M$, $g(x, y) = 0$ if $x > 0$ and $y < -1$; $g(x, y) = 0$ if $x \leqslant 0$ and $y > 1/2$; $g(x, y) = 1$ if $x > 0$ and $y \geqslant -1$; $g(x, y) = -1$ if $x \leqslant 0$ and $y \leqslant 1/2$. This function is closed and its discontinuities are weak of the first species. A division of I is $C = \{A_1, A_2, A_3\}$, $A_1 = \{(x, y) \in I : x > 0, \ y > 0\}$, $A_2 = \{(x, y) \in I : y \leqslant 0\}$ and $A_3 = \{(x, y) \in I : x \leqslant 0, \ y > 0\{$. Two M-d a. of f associated to C are : $f_1 : I \to M$, $f_1(x) = 1$ if $x \in A_1$; $f_1(x) = 0$ if $x \in A_2$; $f_1(x) = -1$ if $x \in A_3$ and $f_2 : I \to M$, $f_2(x) = 1$ if $x \in A_1$; $f_2(x) = -1$ if $x \in A_2$; $f_2(x) = 0$ if $x \in A_3$. Their test points are respectively (1,1/2), (1, −4/3) (−1/2, 1/4) and (1/2, 1), (−1, −1/2), (0,1).

—I—

Let E be a metric space, $J \subset R$ a closed interval, $M \subset J$ a finite set containing more than two points and let $C(E, J)$ be the set of continuous functions $f : E \to J$.

Theorem 1. *For each* $f \in C(E, J)$ *there exists a M-s.d.a. of* f, f_M, *whose discontinuities are weak on the first species.*

Proof. Let $M = \{b_1, b_2 \ldots b_N\}$ be such that $b_{i+1} > b_i$ for each $i \leqslant N - 1$. Let $B_i = \{x \in E : |f(x) - b_i| = \min(|f(x) - Uf(x)|,$ $|f(x) - Df(x)|)\}$. For $x \in E$ let $V_x = \{i \in \mathbf{N} : x \in \mathrm{Int}(B_i)\}$. and $U_x = \{i \in \mathbf{N} : x \in B_i\}$. We define $f_M : E \to M$ as follows : if $V_x \neq \emptyset$, then $f_M(x) = b_k$, $k = \min(V_x)$ and if $V_x = \emptyset$, then $f_M(x) = b_q$, $q = \min(U_x)$. Our function is well defined because $\cup (B_i : i \leqslant N) = E$. For proving that

each discontinuity of f_M is weak of the first species it is enough to prove that each set $H_i = f_M^{-1}(b_i)$ has the property that $\mathrm{Int}(a(H_i)) \subset H_i$. Let $x \in E$ such that H_i is dense on the neighbourhood V of x. We must prove that $x \in H_i$. Let $y \in H_i \cap V$. We have $|f(y) - b_i| \leqslant |f(y) - a|$ for any $a \in M$. The continuity of f implies that each point in V has the same property. Therefore $V \subset B_i$. We obtain $i \in V_x$. If there exists $k \in V_x$ such that $k < i$, then there exists a neighbourhood G of x such that for y in G we have $f_M(y) \leqslant b_k < b_i$. Hence $i = \min(V_x)$ and consequently $f_M(x) = b_i$. It is easy to notice that f_M is a M-s.d a. of f and also that f_M is a closed function (in fact each function with the range contained in M is closed).

For $f \in C(E, J)$ we will say that the function f_M defined as above is the *M-perfect discrete approximation* (*M-p.d.a*) of f. Some reasons for this designation are pointed out by the next two propositions. The sets B_i will be the same as before.

Proposition 1. *Among all the M-s.d.a. of f the M-p.d.a. has at each point the least oscillation.*

Proof. Let $x \in E$ such that $w(f_M, x) \neq 0$. There exists a neighbourhood V of x such that we can find exactly two sets B_i, B_j such that $B_i \cap V \neq \emptyset$ and $B_j \cap V \neq \emptyset$. Let us suppose that there exists a *M-s.d.a.* of f, g, such that $w(g, x) < w(f_M, x)$. Because both functions have the property S_f at x it follows $w(g, x) = 0$. ($w(g, x) < w(f_M, x) = |b_i - b_j|$). We have $g(x) = b_i$ or $g(x) = b_j$. Let s take $g(x) = b_i$. We obtain the existence of an open neighbourhood $K \subset V$ of x such that for each x in K we have $g(x) = b_i$. Consequently $K \subset V \subset B_i$. If there exists a neighbourhood G of x contained in B_j, then for $y \in K \cap G$ we have $f_M(y) = b_k$, $k = \min(i, j)$; if there does not exist such a neighbourhood, then for $y \in K$ we have $f_M(y) = b_i$. In both cases $w(f_A, x) = 0$. This contradiction ends the proof.

Proposition 2. Let E, F be two metric spaces. If $f : E \to F$ is closed then D_f is nowhere dense.

Proof. Let $x \in D_f$. There exists a partition of a neighbourhood of x : $P_1, P_2, \ldots P_m$, such that one of these sets has the property that $y_i \neq f(x)$ (y_i is the point associated to P_i by the definition of the weak discontinuity of the first species). Let us suppose that for each neighbour hood U of x there exists $h \in U \cap P_i$ such that $f(h) \neq y_i$. It follows the existence of a sequence $\{a_n : n \geqslant 1\} \subset P_i$ converging to x with the property $f(a_k) \neq y_i$, $k \in \mathbf{N}^*$. Simultaneously the sequence $\{f(a_k) : k \geqslant 1\}$ converges to $y_i \neq f(x)$. Therefore the closed set $\{a_k : k = 1\} \cup \{x\}$ has a range which is not closed. We obtain that there exists a neighbourhood W of x such that $f(W \cap P_i) = y_i$. We have $\mathrm{Int}(a(P_k)) \subset P_k$ for each $k \leqslant m$, hence $\mathrm{Int}(W \cap P_i) \neq \emptyset$. The function beeing constant on an open set near each point from D_f, it results that D_f is nowhere dense.

Corrolary. An M-p.d.a. has a nowhere dense set of discontinuities.

Remarks 1. The above two propositions are proving, in fact, that among all the s.d.a. of f there is none with a gentlier behaviour than the p.d.a.. Anyhow we must notice that the M-p.d.a. of f is not the single approximation with this feature. If we replace in its definition $\min(V_x)$ and

min (U_x) respectively by max (V_x) and max (U_x) we obtain a function with the same properties.

2. Another interesting result concerning the weak discontinuities of the first species asserts that for any function $f : E \to F$ the set D_f is of the first Baire category. (Compare with the Froda theorem [3] stating that the set of the discontinuities of the first species of a function $f : \mathbf{R} \to \mathbf{R}$ is countable.) This result is not necessary in the following hence we will just give some hints for the proof. Let $x \in D_f$ and $P_1, P_2, \ldots P_m$ a partition of a neighbourhood of x as it results from the definition of weak discontinuities. Ler $r(f, x) = \max\{d(f(x), y_i) : 1 \leqslant i \leqslant m\}$, $d(a, b)$ is the distance between a and b $(a, b \in F)$. We prove that for an $n \in \mathrm{M}^*$ the set $M_n = \{z \in D : r(f, z) \geqslant 1/n\}$ is nowhere dense by observing that its density on a neighbourhood of a point x' implies $x' \notin D_f$.

3. The weak discontinuities of the first species are useful even for approximating functions of a real variable. If I is a closed real interval there exists a continuous function $f : I \to J = [b_1, b_2]$ such that any M-s.d.a. of f has a point of discontinuity which is not of the first species. Here is an example :

$$f : I \to J,\ f(x) = (b_2 - b_1)\ (x - a)\ \sin(1/(x-a))/2(m'(\mathrm{I})) + (b_1 + b_2)/2,\ a \in I$$

— II —

The utility of the precedent long and technical discussion is pointed out by the theorem below.

Let $B \subset \mathbf{R}^n$ be a compact set, let J and M be as before $(M \subset J$, $M = \{b_1, b_2 \ldots b_N\})$ and let $L = \{(b_i + b_{i+1})/2 : 1 \leqslant i \leqslant N - 1\}$. For two functions $f, g : B \to J$ let $B(f, g) = \{x \in B : f(x) \neq g(x)\}$ and $s(f, g) = m'(B(f, g))$. Factoring the space of all functions $f : B \to J$ by the equivalence relation given by the equality almost everywhere, s becomes a metric. We will speak about convergence in s referring to functions $f : B \to J$ even if, in fact, the convergence is valid in the factor space. A measurable connected division of B is a division of B whose elements are measurable connected sets.

Theorem 2. *Let $f : B \to J$ be a continuous function such that $m'(f^{-1}(L)) = 0$. Let $\{f_n : n \geqslant 1\}$ be a sequence of M-discrete approximations of f respectively associated to the sequence of measurable connected divisions $\{C_n : n \leqslant 1\}$ of B. If $\{C_n : n \geqslant 1\}$ converges to 0, then $\{f_n\}$ converges in s to the prefect discrete approximation of f (f_M).*

Proof. For $n \in \mathbf{N}^*$ let $V_n = \{E \in C_n : E \cap f^{-1}(L) \neq \emptyset\}$ and $U_n = C_n - V_n$. Let $V_n' = \cup(E : E \in V_n)$, $U_n' = \cup(F : F \in U_n)$. The sequence $\{C_n\}$ converging to zero we obtain that $\cap(V_m' : m \geqslant 1) = f^{-1}(L)$. Because C_n is a measurable division it follows that V_n' is a measurable set and because $m'(f^{-1}(L)) = 0$ it follows $\lim(m'(V_n') : n \to \infty) = 0$. Let us notice that $U_n' \subset \{x \in B : f_n(x) = f_M(x)\}$. Indeed for $E \in U_n$ the function f_n is constant on E and there exists a point a in E such that :

$$f_n(a) = \min(H(a)) \text{ where } H(a) = \{y \in M : |f(a) - y| = \min\{|f(a) - z| : z \in M\}\}$$

Because f is continuous, E is connected and $E \cap f^{-1}(L) = \emptyset$ it follows $E \subset f^{-1}((l_1, l_2))$ with $(l_1, l_2) \cup L = \emptyset$ and $l_1, l_2 \in L$. Consequently $f_n(a) =$

b with $b \in (l_1, l_2) \cap M$. We obtain also that f_M is constant on E and $f_M(a) = b$. Hence $U'_n \subset \{x \in B : f_n(x) = f_M(x)\}$. It results $\{x \in B : f_n(x) \neq \neq f_M(x)\} \subset V'_n$ and consequently $s(f_n, f_M) = m'(V'_n)$.
It follows the convergence in s of $\{f_n\}$ to f_M.

Remark. The condition $m'(f^{-1}(L)) = 0$ is not too embarrassing. We will quickly prove that for each function $f : B \to J$ there exists $a \in \mathbf{R}$ such that for $h : B \to \mathbf{R}$, $h(x) = f(x) - a$, we have $m'(h^{-1}(L)) = 0$. Let us first notice that the set $K = \{y \in J : m'(f^{-1}(y)) > 0\}$ is countable. The reader will easely observe that for any natural number n the set $K_n = \{y \in J : m'(f^{-1}(y)) > 1/n\}$ is finite (B is compact) and $K = \cup (K_n : n \geqslant 1)$. The set L is finite. The number $r = \inf\{(b_{i+1} - b_i)/2 : i \leqslant N - 1\}$ is positive. Let $P_b = \{x + b : x \in L\}$. For $b \in (0, r)$ the sets P_b are pairwise disjoint. The set K being countable there exists $a \in (0, r)$ such that $K \cap P_a = \emptyset$. Therefore a is the wanted number.

We will study now a significant particular case. Let $I = [0, 1] \subset \mathbf{R}$, $J = [a, a + 2mt]$, $m \in \mathbf{N}^*$ and $t \in \mathbf{R}^+$, $M = \{a + t, \ a + 3t \ \ldots \ a + + (2m - 1)t\}$. Let $C = \{I_i : 0 \leqslant i \leqslant n - 1\}$ be a division of I such that $I_i = [i/n, \ (i + 1)/n)$ for $i < n - 1$ and $I_{n-1} = [(n - 1)/n, 1]$. Let $f : I \to J$ be a C^1 injective function such that for any $x \in I$ we have $f'(x) \leqslant H$. Let $g : I \to M$ be the M-d.a. of f associated to C whose test points are respectively $1/2n, \ \ldots (2i + 1)/2n, \ \ldots (2n - 1)/2n$. We interpret s as follows : $r = s(g, f_M)$ implies that for an arbitrary point $x \in I$ the probability that $|g(x) - f(x)| \leqslant t$ is valid equals $1 - r$. We will denote this probability by P and let $k = (2nt/H)$.

Proposition 3. a. If $k < 1/2$, then $P \geqslant k$; if $1/2 \leqslant k < 1$, then $P \geqslant 1/2$
b. If $k \geqslant 1$, then $P \geqslant (2[k] - 1)/(2[k])$, $[k]$ being the entire part of k.

Proof. Clearly f is monotonous. We will consider it increasing. The function f_M is also increasing. This function is constant on intervals like $(f^{-1}(a + 2vt), f^{-1}(a + (2v + 2)t)$. The minimal lenght of such intervals is $l = 2t/H$ (for each $x, y \in I$ we have $|f(x) - f(y)| \leqslant H|x - y|$). In fact each point $x \in I$ is contained in an interval I_x where f_M is constant. Now let b be a test point of g. This function is constant on the interval $J_b = [b - 1/2n, \ b + 1/2n)$. Clearly $I_b \cap J_b \subset \{x \in I : f_M(x) = g(x)\}$. We notice that $k = l/m'(J_b)$. If $k \leqslant 1/2$ then $m'(I_b \cap J_b) \geqslant 2t/H = k/n$. We have n test points, therefore $P \leqslant nk/n = k$. (Clearly $P = m'(\{x \in \in I : f_M(x) = g(x)\})$. If $k \in (1/2, 1)$ we have $m'(I_b \cap J_b) \geqslant 1/2\, n$. Hence $P \geqslant 1/2$. Let us suppose now $k \geqslant 1$. For $b \in I$, the interval I_b contains at least $[k]$ test points $a_1, a_2 \ \ldots a_k$ (they are written in increasing order). For $1 < i \leqslant [k]$ (or $1 \leqslant i < [k]$) we have $J_{a_i} \subset I_b$ and for $i = 1$ (or, respectively, $i = [k]$), $m'(J_{a_i} \cap I_b) \geqslant 1/2n$. We obtain that $m'(I_b \cap \cap (\cup (J_{a_i} : 1 \leqslant i \leqslant [k]))) \geqslant ([k] - 1)/n + 1/2n$. Consequently $P \geqslant (n/[k])$ $(2[k] - 1)/2n$. $(n/[k]$ equals the maximal number of disjoint intervals of the type of I_b).

Remarks. It seems possible to improve the statement of the second point of the precedent proposition but it is not the case for the first one. If we work on a real machine it will be desirable, ofcourse, to use the minimal t we can find. This is the least positive number accepted by the computer. Let us denote it by o. Obviously $\|C\| \leqslant o$. Consequently $k < < 1/H$ and we obtain that for the most „peaceful" values of H the pro-

bability P may be near 1/2. In the most frequent cases this result is not useful at all. It follows that it is useless to compute blindly the values of the function f at the test points whith the gratest possible precision. A better strategy is to fix before starting computation the minimal reasonable probability (P) and then to evaluate n such that t will be minimal. A rough expression of our conclusions is the following principle of uselessness : „The most detailed local computational results have little global efficiency".

REFERENCES

1. U. W. KULISCH, W. L. MIRANKER, *Computer Arithmetic*. Academic Press, New-York 1981.
2. HELENA PAWLAK, *On some conditions equivalent to the continuity of closed functions*. Demonstratio Mathematica, vol XVII, 3 (1984), 725—732.
3. MIRON NICOLESCU, *Analiza Matematica vol. II*. Edit. Tehnică, București, 1958, 133—134.
4. OCTAV CORNEA, *Proprietăți comune funcțiilor închise și celor cu proprietatea lui Darboux* Stud. Cerc. Mat. 1 (1987) 18—21.

January 15, 1987

University of Bucharest

Dept. of Mathematics
Str. Academiei 14
Bucharest, Romanil