On the characterization of classes of functions by their best linear approximation

By G. ALEXITS in Budapest

1. In the theory of approximation it is an attractive problem to characterize different classes of functions by means of their best approximation. Denote by $\{E_n\}$ a positive non-increasing number sequence. We say that the class of functions $\mathscr{C}$ is characterized by $\{E_n\}$, if the nth best approximation of every $f \in \mathscr{C}$ by polynomials, or another by appropriate system of functions, remains $\leqq E_n$ for every n (so-called direct theorem) and there exists a constant $K > 0$ such that every $f(x)$ for which the nth best approximation is $\leqq KE_n$ for every n belongs to $\mathscr{C}$ (so-called inverse theorem).

It is well known that many important classes of functions are characterized by the sequence of their best trigonometric polynomial approximation. Such are e.g. the classes $\mathrm{Lip}_M\, \alpha$ for $0 < \alpha < 1$ and a fixed $M > 0$ (Lipschitz-constant), the characteristic sequence $\{E_n\}$ being $\{CM \cdot n^{-\alpha}\}$ where C is an absolute constant. But, for $\alpha = 1$, the class $\mathrm{Lip}_M\, 1$ is no more characterized by $\{CM \cdot n^{-1}\}$, this sequence being characteristic for the Zygmund class Z_M containing all 2π-periodic functions for which

$$\max_{0 \leqq x \leqq 2\pi} |f(x+h) + f(x-h) - 2f(x)| \leqq M|h|.$$

In the following, we intend to investigate the nature of the classes $\mathscr{C}$ characterizable by the sequence of their best approximation with linear combinations of an arbitrary system of functions $\{f_\nu(x)\}$. Our main result can be expressed, roughly speaking, about so: if a class $\mathscr{C}$ is characterizable by a sufficiently regular sequence $\{E_n\}$ of any best approximation, then E_n grants the order of magnitude of the absolutely best possible nth linear approximation of $\mathscr{C}$.

One may be tempted to think that the order of the absolutely best approximation could be essentially improved if we turned to non-linear methods of approximation. But, it is not easy to find out what kind of a non-linear method could accomplish this task. If we confront, for instance, the linear method of approximation by polynomials with the non-linear method of approximation by rational functions, we shall show that, for the classes characterizable by polynomial approximation, the best approximation by rational functions gives no essentially better result. Hence, only the polynomially non-characterizable classes are worth to be investigated concerning their best approximation by rational functions.

. Approximations in a Banach space

2. Let B be a Banach space, $\|x\|_B$ the norm of $x \in B$ and $\{y_\nu\}$ a sequence of elements of B set in a fixed order. We form linear combinations $\sum_{k=1}^{n} a_k y_k$ of the first n elements where the a_k's are real numbers. The non-negative number

$$E_n^{(B)}(x, \{y_\nu\}) = \inf_{a_k} \|x - \sum_{k=1}^{n} a_k y_k\|_B$$

is the *nth best* $\{y_\nu\}$*-approximation* of $x \in B$. We call $\{y_\nu\}$ the *basis* of this approximation. If $\mathscr{C}$ is any subset of B, then

$$E_n^{(B)}(\mathscr{C}, \{y_\nu\}) = \sup_{x \in \mathscr{C}} E_n^{(B)}(x, \{y_\nu\})$$

is the nth best $\{y_\nu\}$-approximation *of the set* $\mathscr{C}$.

Denote by $\mathscr{C}(\{E_n\}, \{y_\nu\})$ the set of all those elements $x \in B$ for which

$$E_n^{(B)}(x, \{y_\nu\}) \leqq E_n \qquad (n = 1, 2, \cdots).$$

We call $\mathscr{C}(\{E_n\}, \{y_\nu\})$ the $\{E_n\}$*-saturation set* of the $\{y_\nu\}$-approximation. This is the set of all elements x of B for which a "direct theorem" with the best approximation sequence $\{E_n\}$ exists (referring to $\{y_\nu\}$-approximation). Saturation sets are closely connected with characterizable classes: a set $\mathscr{C} \subset B$ is called $\{E_n\}$*-characterizable,* if there exists a sequence $\{y_\nu\}$ and a positive absolute constant K_1 such that

$$\mathscr{C}(\{K_1 E_n\}, \{y_\nu\}) \subset \mathscr{C} \subset \mathscr{C}(\{E_n\}, \{y_\nu\}).$$

($K_1, K_2, \cdots$ will denote always positive absolute constants.) In the following, we shall suppose that $\{E_n\}$ is a positive non-decreasing number sequence tending to zero.

3. Theorem 1. *A saturation set* $\mathscr{C}(\{E_n\}, \{y_\nu\})$ *is a closed and convex subset of* B*, provided that the elements* $\{y_\nu\}$ *are independent.*

First of all, $\mathscr{C}(\{E_n\}, \{y_\nu\})$ contains infinitely many elements because of a theorem of Bernstein (cf. [3], p. 332) according to which, for every $p \geqq 1$, there exists an element x_p such that

$$E_n^{(B)}(x_p, \{y_\nu\}) = \frac{E_n}{p} \qquad (n=1, 2, \cdots),$$

hence every x_p belongs to $\mathscr{C}(\{E_n\}, \{y_\nu\})$. After this, let x, y be two different elements of $\mathscr{C}(\{E_n\}, \{y_\nu\})$ and consider the element $z=(1-\lambda)x+\lambda y$ where $0<\lambda<1$. Since, for every n,

$$E_n^{(B)}(z, \{y_\nu\}) \leqq (1-\lambda)E_n^{(B)}(x, \{y_\nu\})+\lambda E_n^{(B)}(y, \{y_\nu\}) \leqq ((1-\lambda)+\lambda)\cdot E_n = E_n,$$

the element z belongs to $\mathscr{C}(\{E_n\}, \{y_\nu\})$, this set is therefore convex. — As for the closure, let x be an element of accumulation and $\{x_j\}$ a sequence of different elements of $\mathscr{C}(\{E_n\}, \{y_\nu\})$ converging to x. By assumption, to every x_j belongs at least one linear combination $\sum_{k=1}^{n} a_k^{(j)} y_k$ such that

$$\| x_j - \sum_{k=1}^{n} a_k^{(j)} y_k \|_B = E_n^{(B)}(x_j, \{y_\nu\}) \leqq E_n \qquad (n=1, 2, \cdots).$$

Then we have

$$E_n^{(B)}(x, \{y_\nu\}) \leqq \| x - \sum_{k=1}^{n} a_k^{(j)} y_k \|_B \leqq \|x-x_j\|_B + E_n,$$

hence, going over to the limit $j\to\infty$,

$$E_n^{(B)}(x, \{y_\nu\}) \leqq E_n \qquad (n=1, 2, \cdots),$$

i.e. $x\in\mathscr{C}$, what we had to prove.

Corollary 1. *A set $\mathscr{C}$ is characterizable only if it contains a convex continuum, namely $\mathscr{C}(\{K_1 E_n\}, \{y_\nu\})$.*

4. The class $\mathscr{C}$ of convex functions having bounded nth derivatives was, in the last time, subject of investigations ([8], I and II) concerning its best approximation by rational functions. This class $\mathscr{C}$ can serve as an example for a class which is absolutely not characterizable by linear approximations in any Banach space. Because if, on the contrary, there were a characteristic sequence $\{E_n\}$ to a linear $\{y_\nu\}$-approximation, then $\mathscr{C}$ would contain $\mathscr{C}(\{K_1 E_n\}, \{y_\nu\})$. So choose two convex functions $f(x)$ and $g(x)$ contained in $\mathscr{C}$ and a number $0<\lambda<1$ such that $h(x)=(1-\lambda)f(x)+\lambda g(x)$ should not be convex. Then $K_1 f(x)$ and $K_1 g(x)$ would be contained in $\mathscr{C}(\{K_1 E_n\}, \{y_\nu\})$ but $K_1 h(x)$ not, contrary to Corollary 1.

5. We call the number sequence $\{E_n\}$ *slowly decreasing*, if it is positive, non-increasing and tends to zero such that $E_{2n}\geqq K_2 E_n$ for every n.

Theorem 2. *If $\{E_n\}$ is slowly decreasing and the basis of approximation bounded* ($\|y_\nu\|_B \leqq K_3$), *then there exists a* $K_4 > 0$ *such that*

$$K_4 E_n \cdot y_k \in \mathscr{C}(\{E_n\}, \{y_\nu\}) \qquad (k=1, 2, \cdots, 2n;\ n=1, 2, \cdots).$$

The boundedness of the basis $\{y_\nu\}$ implies

$$\left\| \frac{E_{2n}}{2K_3} \cdot y_k - \frac{E_{2n}}{2K_3} \cdot y_m \right\|_B \leqq \frac{E_{2n}}{2K_3} (\|y_k\|_B + \|y_m\|_B) \leqq E_{2n}.$$

Therefore, if $k \leqq 2n$ and $m \leqq k$,

$$E_m^{(B)}\left(\frac{E_{2n}}{2K_3} \cdot y_k, \{y_\nu\}\right) \leqq E_{2n} \leqq E_m.$$

This estimation is the more satisfied when $m > k$, because then we have

$$E_m^{(B)}\left(\frac{E_{2n}}{2K_3} \cdot y_k, \{y_\nu\}\right) = 0.$$

So it follows

$$\frac{E_{2n}}{2K_3} \cdot y_k \in \mathscr{C}(\{E_n\}, \{y_\nu\}) \qquad (k = 1, 2 \cdots, 2n;\ n = 1, 2, \cdots).$$

Put $K_4 = K_2/2K_3$, then, by the slow decrease of $\{E_n\}$, we obtain

$$K_4 E_n \leqq \frac{E_{2n}}{2K_3},$$

hence the more we have $K_4 E_n \cdot y_k \in \mathscr{C}(\{E_n\}, \{y_\nu\})$ for $k = 1, 2, \cdots, 2n$ and $n = 1, 2, \cdots$.

Corollary 2. *If $\{E_n\}$ is slowly decreasing and $\{y_\nu\}$ bounded, then a set $\mathscr{C}$ can be $\{E_n\}$-characterisable by $\{y_\nu\}$-approximation only if there exists a K_4 such that $K_4 E_n \cdot y_k \in \mathscr{C}$ for $k = 1, 2, \cdots, 2n$ and $n = 1, 2, \cdots$.*

6. It is known [4] that the best linear approximation in the space C of continuous functions with the basis $\{w_\nu(x)\}$ of Walsh functions provides, for the classes $\mathrm{Lip}\,\alpha$ with $0 < \alpha < 1$ and Lipschitz constant 1, about the same order of magnitude as the best polynomial approximation, namely $\{n^{-\alpha}\}$. Although, for the $\{w_\nu(x)\}$-approximation of $\mathrm{Lip}\,\alpha$ only direct theorems can be obtained. Because if $\mathscr{C} = \mathrm{Lip}\,\alpha$ were $\{E_n\}$-characterizable by $\{w_\nu(x)\}$-approximation, then by Corollary 2 the functions $K_4 n^{-\alpha} w_k(x)$ $(k = 2n)$ would belong to $\mathscr{C}$ and this is impossible, since $w_k(x)$ is not continuous.

Approximation in Banach spaces contained in a Hilbert space

7. Let H be a Hilbert space, (x, y) the inner product of the elements x, y of H and $\|x\|_H = \sqrt{(x, x)}$ the norm of $x \in H$. By B we denote now a Banach space $B \subset H$ for which the relation $\|x\|_B \geqq K_5 \|x\|_H$ is valid for all $x \in B$.

We prove now the counterpart of Theorem 2, a statement proved previously, in a somewhat less general form, by KNAPOWSKI and myself [2].

Theorem 3. *Assume that for a set $\mathscr{C} \subset B$ and for an orthonormal system $\{\xi_\nu\} \subset B$ we have*

(1) $$E_n \cdot \xi_k \in \mathscr{C} \qquad (k = 1, 2, \cdots, 2n;\ n = 1, 2, \cdots),$$

then, for any system $\{y_\nu\} \subset B$, $E_n(\mathscr{C}, \{y_\nu\}) \geqq K_6' E_n \qquad (n = 1, 2, \cdots)$.

First, let $\{\eta_\nu\} \subset B$ be an arbitrary orthonormal system and

$$s_n(\xi_k, \{\eta_\nu\}) = \sum_{\nu=1}^{n} (\xi_k, \eta_\nu) \cdot \eta_\nu .$$

By the orthonormality, we have

$$1 = \|\xi_k\|_H \leqq \|\xi_k - s_n(\xi_k, \{\eta_\nu\})\|_H + \|s_n(\xi_k, \{\eta_\nu\})\|_H .$$

Hence, owing attention to $E_n^{(H)}(\xi_k, \{\eta_\nu\}) = \|\xi_k - s_n(\xi_k, \{\eta_\nu\})\|_H$, and summing for $k = 1, 2, \cdots, 2n$, we get

(2) $$2n \leqq \sum_{k=1}^{2n} E_n^{(H)}(\xi_k, \{\eta_\nu\}) + \sum_{k=1}^{2n} \|s_n(\xi_k, \{\eta_\nu\})\|_H .$$

Applying CAUCHY's inequality, we see that

$$\sum_{k=1}^{2n} \|s_n(\xi_k, \{\eta_\nu\})\|_H \leqq \sqrt{2n} \left\{ \sum_{k=1}^{2n} \|s_n(\xi_k, \{\eta_\nu\})\|_H^2 \right\}^{\frac{1}{2}} = \sqrt{2n} \left\{ \sum_{k=1}^{2n} \sum_{j=1}^{n} (\xi_k, \eta_j)^2 \right\}^{\frac{1}{2}} .$$

Both systems $\{\xi_\nu\}$ and $\{\eta_\nu\}$ being orthonormed, by BESSEL's inequality,

$$\sum_{k=1}^{2n} (\xi_k, \eta_j)^2 \leqq \|\eta_j\|_H^2 = 1,$$

hence we obtain the estimation $\sum_{k=1}^{2n} \|s_n(\xi_k, \{\eta_\nu\})\|_H \leqq \sqrt{2}\, n$.

Therefore (2) leads to $\sum_{k=1}^{2n} E_n^{(H)}(\xi_k, \{\eta_\nu\}) \geqq (2 - \sqrt{2}) n$, and consequently to

(3) $$\max_{1 \leqq k \leqq 2n} E_n^{(H)}(\xi_k, \{\eta_\nu\}) \geqq \frac{2 - \sqrt{2}}{2} .$$

By assumption, $\|x\|_B \geqq K_5 \|x\|_H$ for all $x \in B$, hence also $E_n^{(B)}(x, \{\eta_\nu\}) \geqq K_5 E_n^{(H)}(x, \{\eta_\nu\})$. Thus, if we apply (3) to the elements $E_n \cdot \xi_k$ instead of ξ_k, we obtain

$$\max_{1 \leqq k \leqq 2n} E_n^{(B)}(E_n \cdot \xi_k, \{\eta_\nu\}) \geqq K_5 \frac{2-\sqrt{2}}{2} E_n.$$

Since we assumed that, for $k \leqq 2n$, $E_n \cdot \xi_k$ is an element of $\mathscr{C}$, hence

$$E_n^{(B)}(\mathscr{C}, \{\eta_\nu\}) \geqq K_6 E_n. \tag{4}$$

The estimation (4) shows that our theorem holds good for the best approximations with an orthonormal basis $\{\eta_\nu\} \subset B$. Consider, now, an arbitrary approximation basis $\{y_\nu\} \subset B$. Delete from $\{y_\nu\}$ the linearly dependent elements and denote by $\{y_\nu^*\}$ the remainder system. Since there are less linear combinations of $y_1, y_2, \cdots, y_\nu$ than those of $y_1^*, y_2^*, \cdots, y_n^*$, we have $E_n^{(B)}(x, \{y_\nu\}) \geqq E_n^{(B)}(x, \{y_\nu^*\})$. But $\{y_\nu^*\}$ can be orthonormalized such that the νth element η_ν of the orthonormalized system should be a linear combination of the elements $y_1^*, y_2^*, \cdots, y_\nu^*$. Thus the linear combinations of $y_1^*, y_2^*, \cdots, y_n^*$ are the same as those of $\eta_1, \eta_2, \cdots, \eta_n$, therefore

$$E_n^{(B)}(x, \{y_\nu\}) \geqq E_n^{(B)}(x, \{y_\nu^*\}) = E_n^{(B)}(x, \{\eta_\nu\})$$

for all $x \in B$ and $n = 1, 2, \cdots$. Hence, by (4), $E_n^{(B)}(\mathscr{C}, \{y_\nu\}) \geqq K_6 E_n$, as we asserted.

8. For two positive number sequences $\{a_n\}$ and $\{b_n\}$, we write $\{a_n\} \approx \{b_n\}$, if $a_n \leqq K_7 b_n$ and $b_n \leqq K_8 a_n$ for $n = 1, 2, \cdots$, i.e. if $\{a_n\}$ and $\{b_n\}$ have the same order of magnitude. If there exists an orthonormal system satisfying (1) for a set $\mathscr{C} \subset B$ then, by Theorem 3, it follows that the numbers

$$E_n^{(B)}(\mathscr{C}) = \inf E_n^{(B)}(\mathscr{C}, \{y_\nu\}) \qquad (n = 1, 2, \cdots)$$

are positive where the inf has to be taken for all possible basises $\{y_\nu\} \subset B$. We call $E_n^{(B)}(\mathscr{C})$ the nth *absolutely best* linear approximation of $\mathscr{C}$. (We are not concerned with the question whether $E_n^{(B)}(\mathscr{C})$ is attained or not by a system $\{y_\nu\} \subset B$.)

If $\{\xi_\nu\}$ is an orthonormal system, set $E_n^* = \sup e_n$, where the 'sup' has to be taken for all $e_n > 0$ for which

$$e_n \cdot \xi_\kappa \in \mathscr{C} \qquad (k = 1, 2, \cdots, 2n;\ n = 1, 2, \cdots).$$

We shall show that, in many cases, the sequence $\{E_n^*\}$ is equivalent to $\{E_n\}$ and $\{E_n^{(B)}(\mathscr{C})\}$.

Theorem 4. *Let $\{\xi_\nu\}$ be a bounded orthonormal system and $\mathscr{C}$ a closed set $\{E_n\}$-characterizable by best $\{\xi_\nu\}$-approximation. If $\{E_n\}$ is slowly decreasing, then*

$$\{E_n\} \approx \{E_n^{(B)}(\mathscr{C})\} \approx \{E_n^*\}.$$

Since $\mathscr{C}$ is closed, we have

$$E_n^* \cdot \xi_k \in \mathscr{C} \qquad (k = 1, 2, \cdots, 2n;\ n = 1, 2, \cdots).$$

By Theorem 3 it follows then $E_n \geqq K_6 E_n^*$. By Theorem 2, there is a K_4 such that $K_4 E_n \cdot \xi_k \in \mathscr{C}(\{E_n\}, \{\xi_\nu\})$ for $k=1, 2, \cdots, 2n$, hence

$$(5) \qquad K_1 K_4 E_n \cdot \xi_k \in \mathscr{C}(\{K_1 E_n\}, \{\xi_\nu\}) \subset \mathscr{C} \qquad (k=1, 2, \cdots, 2n;\ n=1, 2, \cdots)$$

and therefore $K_1 K_4 E_n \leqq E_n^*$, i.e. $E_n \approx E_n^*$. We have still to prove $E_n^* \approx E_n^{(B)}(\mathscr{C})$. First, we have $E_n^{(B)}(\mathscr{C}) \geqq K_6 E_n^*$ by Theorem 3. But $E_n^{(B)}(\mathscr{C}) \leqq E_n$ and (5) implies

$$K_1 K_4 E_n^{(B)}(\mathscr{C}) \cdot \xi_k \in \mathscr{C} \qquad (k=1, 2, \cdots, 2n;\ n=1, 2, \cdots),$$

hence $K_1 K_4 E_n^{(B)}(\mathscr{C}) \leqq E_n^*$, thus also $E_n^* \approx E_n^{(B)}(\mathscr{C})$ is proved.

Comparison of the best approximations by polynomials and by rational functions

9. The mostly used linear method is the approximation by polynomials; its simplest extension to a non-linear approximation method consists in the substitution of the polynomials by corresponding rational functions. We intend to compare efficacy of these two methods and shall see that, for classes characterizable by best polynomial approximation, the non-linear method provides no better results than the linear method does.

Denote by $r_n(x)$ a rational function of degree n, i.e. $r_n(x) = P_n(x)/Q_n(x)$ where $P_n(x)$ and $Q_n(x)$ are polynomials of degree $\leqq n$. Then, $\mathscr{C}$ being a given class of continuous functions, we call

$$\varrho_n(\mathscr{C}) = \sup_{f \in \mathscr{C}} \inf_{r_n} \|f - r_n\|_C$$

the nth best rational approximation of $\mathscr{C}$ in the space C.

Szüsz—Turán [8], Freud [5], and Szabados [7] have proved that, for some classes, the best rational approximation in the space C may be essentially better than the best polynomial approximation. The classes considered by these autors are not characterizable by polynomial approximation. But, for the classical polynomially characterizable classes, the best rational approximation is equivalent to the best polynomial approximation. (Newman [6], Szabados [7].) We shall see that this phenomenon occurs for all polynomially characterizable classes.

Theorem 5. *Let $\mathscr{C}$ be a class of continuous functions $\{E_n\}$-characterizable by polynomial approximation in the space C where $\{E_n\}$ is slowly decreasing. Then, for the best rational approximation, we have $\{\varrho_n(\mathscr{C})\} \approx \{E_n\}$.*

Consider the function

$$f(x) = K_{11} \cdot \sum_{k=0}^{\infty} (E_{3^k} - E_{3^{k+1}}) T_{3^k}(x)$$

where $T_\nu(x)$ denotes the ν-th normed Chebysheff polynomial and K_{11} an appropriate constant. Let $s_n(x)$ be the nth partial sum of this series. Since $|T_\nu(x)| \leqq \sqrt{2/\pi}$, we have

$$\max_{-1\leqq x\leqq 1} |f(x)-s_{3^m}(x)| \leqq K_{11} E_{3^m}.$$

A result of BERNSTEIN (cf. ACHYESER [1], p. 79) states that $s_{3^m}(x)$ represents, for $3^m \leqq n < 3^{m+1}$, the best approximating polynomial of degree n and even the best approximating rational function of degree n. Thus, denoting by $\varrho_n(f)$ the best approximation of the function $f(x)$ by rational functions of degree n, we obtain

$$\varrho_n(f) = E_n(f, \{T_n\}) = K_{11} E_n \qquad (3^m \leqq n < 3^{m+1};\ m=0, 1, \cdots). \tag{7}$$

Because of the characterisability of $\mathscr{C}$ by polynomial approximation, there is a K_1 such that

$$\mathscr{C}(\{K_1 E_n\}, \{T_n\}) \subset \mathscr{C},$$

while $\{E_n\}$ is slowly decreasing; therefore, by appropriate choice of K_{11}, it follows that

$$K_{11} E_{3^m} \leqq K_1 E_{3^{m+1}} \leqq K_1 E_n \qquad (3^m \leqq n < 3^{m+1};\ m=0, 1, \cdots).$$

Thus, in accordance with (7), we see that $f \in \mathscr{C}$ and so $\varrho_n(\mathscr{C}) \geqq \varrho_n(f) = K_{11} E_n$. This is just our assertion.

References

[1] Н. И. Ахиезер, *Лекции по теории аппроксимации* (Москва—Ленинград, 1947).

[2] G. ALEXITS—S. KNAPOWSKI, A note on the best approximation by linear forms of functions, *Acta Sci. Math.*, **27** (1966), 49—54.

[3] PH. J. DAVIS, *Interpolation and approximation* (New York—Toronto—London, 1963).

[4] N. J. FINE, On the Walsh functions, *Trans. Amer. Math. Soc.*, **63** (1949), 372—414.

[5] G. FREUD, Über die Approximation reeller Funktionen durch gebrochene rationale Funktionen, *Acta Math. Acad. Sci. Hung.*, **17** (1966), 313—324.

[6] D. J. NEWMAN, Rational approximation to $|x|$, *Michigan Math. Journ.*, **11** (1964), 11—14.

[7] J. SZABADOS, Generalization of two theorems of G. Freud concerning the rational approximation, *Studia Sci. Math. Hung.*, **2** (1967), 73—80.

[8] P. SZÜSZ—P. TURÁN, On the constructive theory of functions. I, *Publ. Math. Inst. Hung. Acad. Sci.*, **9** (1964), 495—502; II, *Studia Sci. Math. Hung.*, **1** (1966), 65—69; III, *Studia Sci. Math. Hung.*, **1** (1966), 315—322.

(Received May 11, 1967)