A note on the best approximation by linear forms of functions

By G. ALEXITS in Budapest and S. KNAPOWSKI in Poznań

1. One of the fundamental problems of the theory of approximation seems to be that of determining the worst order of approximation effected by given means over a given class of functions. As far as we know, the first general result of this type is due to A. N. KOLMOGOROV [1]. He considered the class of functions $f(x)$, $0 \leqq x \leqq 1$, having r first derivatives $f'(x), f''(x), \ldots, f^{(r)}(x)$, where $f^{(r)}(x) \in L^2(0, 1)$, and satisfying the periodicity conditions: $f^{(\nu)}(0) = f^{(\nu)}(1)$, $0 \leqq \nu < r$. The result was: *the worst order of L^2-approximation over this class by linear forms of given n functions can be minimized by the first n functions of the trigonometric system, in fact essentially only by these functions.* KOLMOGOROV's method has been developed and applied to several related problems in recent papers by V. M. TIHOMIROV [4] and G. G. LORENTZ [5].

There are two more papers which study similar questions, using another method. The first one [2] is concerned with L^2-approximation by partial sums $s_n(f, x)$ of the development

$$f(x) \sim \sum_n c_n \varphi_n(x), \tag{1.1}$$

where $\{\varphi_n(x)\}$ is an arbitrary complete orthonormal system, and $f(x)$ belongs to the class of functions of finite variation, or to the class Lip 1. The second paper [3] gets essentially further; it supplies the required lower estimate in case of L^p-approximation ($1 \leqq p \leqq \infty$) within the class of all r times continuously differentiable functions, the approximation means being the Toeplitz means $\sum_{k=1}^{n} \lambda_k c_k \varphi_k(x)$ of (1.1) satisfying the condition $\sum_{k=1}^{n} \lambda_k^2 \leqq n$. In addition, [3] dispenses with the restriction that $\{\varphi_n(x)\}$ be a complete system.

Let $\mathfrak{K}$ be a subclass of $L^p(0, 1)$. The aim of this note is to give a simple criterion to determine a system $\{\varphi_\nu\}$ with the property that its n-th linear forms approximate in $\mathfrak{K}$ essentially no worse than the n-th linear forms of any other system. Roughly speaking, our theorem provides the best system for linear approximation within the given class.

In § 4 we give two instances to illustrate this theorem.

2. N o t a t i o n. L^p stands for $L^p(0, 1)$, $2 \leqq p \leqq \infty$; $\| \ \|_p$ denotes the L^p-norm, $\| \ \| = \| \ \|_\infty$ denotes the $C(0, 1)$-norm. Let $\{f_\nu(x)\}$ denote a given system of functions defined and L^p-integrable in [0, 1], the indexing $f_1, f_2, \dots$ is supposed to be fixed. By a linear form of order $\leqq n$, corresponding to the system, we mean an expression

$$L_n(x) = \sum_{k=1}^{n} a_{nk} f_k(x) \tag{2.1}$$

whose coefficients are real numbers. Put

$$E_n^{(p)}(f; \{f_\nu\}) = \inf \|f - L_n\|_p$$

where the infimum is to be extended over all possible linear forms (2.1) of order $\leqq n$. In case of $p = \infty$ we write simply $E_n = E_n^{(\infty)}$. If $\mathfrak{K}$ is a subclass of L^p, the "worst best approximation" in $\mathfrak{K}$ is defined by

$$E_n^{(p)}(\mathfrak{K}; \{f_\nu\}) = \sup_{f \in \mathfrak{K}} E_n^{(p)}(f; \{f_\nu\}).$$

Finally, given two systems $\{f_\nu^{(1)}\}$ and $\{f_\nu^{(2)}\}$ in L^p, we say that system $\{f_\nu^{(1)}\}$ provides, in $\mathfrak{K}$, no essentially better L^p-approximation than the system $\{f_\nu^{(2)}\}$, if

$$E_n^{(p)}(\mathfrak{K}; \{f_\nu^{(1)}\}) \geqq K_1 \cdot E_n^{(p)}(f; \{f_\nu^{(2)}\}) \qquad (n = 1, 2, \dots)$$

where $K_1 > 0$ is independent of n. In case of $p = \infty$ we will speak of uniform approximation rather than L^∞-approximation.

3. Lemma. *Let $\{\varphi_\nu\} \in L^p$ be orthonormal over* [0, 1], *further let $\mathfrak{K}$ be a subclass of L^p and n a positive integer. If there exists a positive constant K_2 such that*

$$g_k^{(n)}(x) \overset{\text{def}}{=} K_2 E_n^{(p)}(\mathfrak{K}; \{\varphi_\nu\}) \cdot \varphi_k(x) \in \mathfrak{K} \qquad (k = 1, 2, \dots, 2n),$$

then, for any orthonormal system $\{\psi_\nu\}$ in L^p,

$$E_n^{(p)}(\mathfrak{K}; \{\psi_\nu\}) \geqq K_3 E_n^{(p)}(\mathfrak{K}; \{\varphi_\nu\}) \tag{3.1}$$

where K_3 is another positive constant.[1]

Proof. Let $s_n(f; \{\psi_\nu\}) = s_n(f; \{\psi_\nu\}; x)$ stand for the n-th partial sum of the ψ_ν-Fourier series of $f \in L^p$. As is well known, among all linear forms (2.1), s_n provides the best approximation to f in the L^2-norm, namely

$$E_n^{(2)}(f; \{\psi_\nu\}) = \|f - s_n(f; \{\psi_\nu\})\|_2.$$

Putting $f = \varphi_k$, and in view of $\|f\|_2 \leqq \|f - s_n\|_2 + \|s_n\|_2$,

$$1 = \|\varphi_k\|_2 \leqq E_n^{(2)}(\varphi_k; \{\psi_\nu\}) + \left\{ \int_0^1 s_n^2(\varphi_k; \{\psi_\nu\}; x)\, dx \right\}^{\frac{1}{2}},$$

whence summing for $k = 1, 2, \dots, 2n$,

$$2n \leqq \sum_{k=1}^{2n} E_n^{(2)}(\varphi_k; \{\psi_\nu\}) + \sum_{k=1}^{2n} \left\{ \int_0^1 s_n^2(\varphi_k; \{\psi_\nu\}; x)\, dx \right\}^{\frac{1}{2}}. \tag{3.2}$$

---

[1]) If K_2 is an absolute constant then K_3 also is an absolute constant.

Applying the Schwarz inequality to the last sum, and making subsequently use of the orthonormality of ψ_ν's, we get

$$\sum_{k=1}^{2n}\left\{\int_0^1 s_n^2(\varphi_k;\{\psi_\nu\};x)\,dx\right\}^{\frac{1}{2}} \leqq \sqrt{2n}\left\{\sum_{k=1}^{2n}\int_0^1 s_n^2(\varphi_k;\{\psi_\nu\};x)\,dx\right\}^{\frac{1}{2}} =$$

$$=\sqrt{2n}\left\{\sum_{k=1}^{2n}\sum_{j=1}^{n}\left(\int_0^1 \varphi_k(x)\psi_j(x)\,dx\right)^2\right\}^{\frac{1}{2}}.$$

Finally, by Bessel's inequality

$$\sum_{k=1}^{2n}\left(\int_0^1 \varphi_k(x)\psi_j(x)\,dx\right)^2 \leqq \int_0^1 \psi_j^2(x)\,dx = 1,$$

we obtain

$$\sum_{k=1}^{2n}\left\{\int_0^1 s_n^2(\varphi_k;\{\psi_\nu\};x)\,dx\right\}^{\frac{1}{2}} \leqq n\sqrt{2}.$$

Returning to (3.2)

$$\sum_{k=1}^{2n} E_n^{(2)}(\varphi_k;\{\psi_\nu\}) \geqq (2-\sqrt{2})n,$$

and *a fortiori*

$$\max_{1\leqq k\leqq 2n} E_n^{(2)}(\varphi_k;\{\psi_\nu\}) \geqq \frac{2-\sqrt{2}}{2}. \tag{3. 3}$$

Replacing φ_k by $g_k^{(n)}=K_2E_n^{(p)}(\mathfrak{K};\{\varphi_\nu\})\cdot\varphi_k$ and nothing that

$$E_n^{(p)}(f;\{\psi_\nu\})\geqq E_n^{(2)}(f;\{\psi_\nu\}) \qquad (p\geqq 2), \tag{3. 4}$$

$$E_n^{(p)}(cf;\{\psi_\nu\})=|c|E_n^{(p)}(f;\{\psi_\nu\}), \tag{3. 5}$$

we conclude

$$\max_{1\leqq k\leqq 2n} E_n^{(p)}(g_k^{(n)};\{\psi_\nu\}) \geqq \frac{2-\sqrt{2}}{2}K_1\cdot E_n^{(p)}(\mathfrak{K};\{\varphi_\nu\}),$$

and (3.1) follows.

Theorem. *Let $\mathfrak{K}$ be a subclass of L^p and $\{f_\nu\}$ an arbitrary system of functions belonging to L^p. Suppose that there exists a system $\{\varphi_\nu\}\subset L^p$, orthonormal in* [0, 1], *and a positive constant K_4 such that for every $n=1, 2, \dots$ functions*

$$g_k^{(n)}(x)=K_4E_n^{(p)}(\mathfrak{K};\{\varphi_\nu\})\cdot\varphi_k(x) \qquad (k=1, 2, \dots, 2n)$$

belong to $\mathfrak{K}$. Then

$$E_n^{(p)}(\mathfrak{K};\{f_\nu\})\geqq K_5E_n^{(p)}(\mathfrak{K};\{\varphi_\nu\}) \tag{3. 6}$$

where K_5 is another positive constant.[2])

[2]) If K_4 is an absolute constant then K_5 also is an absolute constant.

Corollary. *Let* $\{\varphi_\nu\}$ *and* $\{\psi_\nu\}$ *be orthonormal over* $[0, 1]$. *If for* $k=1, 2, \ldots, 2n$; $n=1, 2, \ldots$

$$K_6 E_n^{(p)}(\mathfrak{K}; \{\varphi_\nu\})\cdot\varphi_k \in \mathfrak{K}$$

and

$$K_7 E_n^{(p)}(\mathfrak{K}; \{\psi_\nu\})\cdot\psi_k \in \mathfrak{K}$$

where K_6, K_7 *are positive constants, then neither of the systems* $\{\varphi_\nu\}$, $\{\psi_\nu\}$ *provides, in* $\mathfrak{K}$, *an essentially better approximation than the other.*

Proof of the theorem. First of all, we observe that without any loss of generality f_ν's can be supposed linearly independent. For if not, we would reject those expressible as linear forms of the preceding ones and consider the new system, say $\{f_\nu^*\}$, whose elements enjoy the required property. The set of linear forms (2. 1) corresponding to the new system contains obviously all linear forms derived from the original one, and consequently

$$E_n^{(p)}(f; \{f_\nu\}) \geqq E_n^{(p)}(f; \{f_\nu^*\}).$$

Hence it is enough to prove (3. 6) for $E_n^{(p)}(f; \{f_\nu^*\})$.

Next we note that f_ν^*'s can be supposed orthonormal. In fact, the familiar Schmidt orthogonalization-process of $\{f_\nu^*\}$ leaves the set (2. 1) of linear forms of order $\leqq n$ — whence also the numbers $E_n^{(p)}(f; \{f_\nu^*\})$ — unchanged.

Thus we shall suppose $\{f_\nu^*\}$ an orthonormal system. Our lemma can be applied and (3. 6) follows.

4. To illustrate the theorem, we consider two special cases:

a) $p=\infty$, $\varphi_1(x)=1$, $\varphi_{2\nu}(x)=\sqrt{2}\cos\pi\nu x$, $\varphi_{2\nu+1}(x)=\sqrt{2}\sin\pi\nu x$, $\mathfrak{K}=\mathfrak{K}_\alpha=$ the class of 1-periodic functions belonging to $\mathrm{Lip}_1\,\alpha$ $(0<\alpha\leqq 1)$ on the whole real axis. As is well known,

$$E_n(f; \{\varphi_\nu\}) \leqq K_8 n^{-\alpha} \quad \text{for} \quad f\in\mathfrak{K}$$

where $K_8>0$ is an absolute constant. Hence

$$E_n(\mathfrak{K}; \{\varphi_\nu\}) \leqq K_8 n^{-\alpha}.$$

It is easy to see that for any x', x'' and $k=1, 2, \ldots, 2n$

$$\frac{|\varphi_k(x'')-\varphi_k(x')|}{n^\alpha\sqrt{2}} \leqq \begin{cases} \pi k|x''-x'|n^{-\alpha} \\ 2n^{-\alpha} \end{cases} \leqq 2\pi|x''-x'|^\alpha,$$

so that

$$g_k^{(n)}(x) = 2^{-\frac{3}{2}}\pi^{-1}K_8^{-1}E_n(\mathfrak{K}_\alpha; \{\varphi_\nu\})\cdot\varphi_k(x)\in\mathfrak{K} \qquad (k=1, 2, \ldots, 2n).$$

This means: *No system* $\{f_\nu\}$ *provides, in* $\mathfrak{K}_\alpha$, *an essentially better uniform approximation than the trigonometric system.*

b) $p=\infty$, $\{\varphi_\nu\}$ the same as in a), $\mathfrak{K}=\mathfrak{K}^{(r)}=$ the class of 1-periodic functions f whose r-th derivative is continuous in $(-\infty, +\infty)$ and $\max\limits_{0\leqq l\leqq r}\|f^{(l)}\|\leqq 1$. By a well-known theorem

$$E_n(\mathfrak{K}^{(r)}; \{\varphi_\nu\}) \leqq K_9(r) n^{-r}.$$

Since for $0 \leqq l \leqq r$, $k = 1, 2, \ldots, 2n$

$$\left|2^{-\frac{1}{2}} n^{-r} \varphi_k^{(l)}(x)\right| \leqq \left(\frac{k\pi}{n}\right)^r \leqq (2\pi)^r,$$

it follows that

$$g_k^{(n)}(x) = 2^{-r-\frac{1}{2}} \pi^{-r} (K_9(r))^{-1} E_n(\mathfrak{K}^{(r)}; \{\varphi_\nu\}) \cdot \varphi_k(x) \in \mathfrak{K}^{(r)} \qquad (k = 1, 2, \ldots, 2n).$$

Hence: *No system $\{f_\nu\}$ provides, in $\mathfrak{K}^{(r)}$, an essentially better uniform approximation than the trigonometric system.*

Remark. The statements about the examples a) and b) are not new. But, if we take into account that the classes discussed, and many others, contain, besides the trigonometric system, also functions $g_k^{(n)}(x)$ of some other orthonormal systems, e. g. those of certain Sturm—Liouville systems, then, by our corollary we can show that all these systems provide, in the corresponding classes, essentially the same best uniform approximation; a result the direct proof of which would be rather lengthy.

5. It is obvious that our theorem and its consequences remain true, except for the constants, if we replace [0, 1] by an arbitrary finite interval $[a, b]$, and understand by "orthonormal system" a system of functions orthonormal, in $[a, b]$, relatively to a weight function $w(x) \geqq 0$.

Denote by $\{p_n(x)\}$ the system of orthonormal polynomials determined by the weight function $w(x)$. By our theorem it follows that, if the class $\mathfrak{K}$ contains the functions

$$g_k^{(n)}(x) = K_{10} E_n(\mathfrak{K}; \{p_\nu\}) \cdot p_k(x) \qquad (k = 1, 2, \ldots, 2n; n = 1, 2, \ldots),$$

we get for no system $\{f_\nu\} \subset C(a, b)$ an essentially better uniform approximation in $\mathfrak{K}$ than the best one provided by polynomials. Therefore, denoting by $\mathfrak{A}$ the class of all analytical funtions $f(x)$ possessing derivatives $\|f^{(r)}(x)\| \leqq 1, r = 0, 1, \ldots$ the essentially best uniform approximation in $\mathfrak{A}$ is provided by the system of all polynomials. This result is reversible in a certain sense:

If there exists a constant $K_{11} > 0$ such that for any $f \in \mathfrak{A}$ we have

$$E(f; \{f_\nu\}) \leqq K_{11} \|f^{(r)}\| \qquad (r = 0, 1, \ldots),$$

then the set $\{L_n(x)\}$ of all linear forms corresponding to the system $\{f_\nu(x)\}$ contains all polynomials.

Proof. Put $P_k(x)$ a polynomial of degree k having the norm $\|P_k\| \leqq 1$ in $[a, b]$, and consider the functions

$$\psi_k(x) = \frac{P_k(x)}{K_{12} \cdot k^{2k}}$$

where K_{12} is a suitably chosen positive constant. The functions $\psi_k(x)$ belong to the class $\mathfrak{A}$. Indeed, taking $r \leqq k$, by a well-known inequality of Markov—Bernstein, we obtain $\|P_k^{(r)}\| \leqq K_{12} k^{2k} \|P_k\|$ where K_{12} depends only on r and the lenght of $[a, b]$. Hence

$$\|\psi_k^{(r)}\| \leqq \frac{\|P_k^{(r)}\|}{K_{12} k^{2k}} \leqq \|P_k\| \leqq 1.$$

If $r>k$, we have evidently $\|\psi_k^{(r)}\|=0$. Fix, now, $r>k$. Since there exists at least one linear form $L_n^*(x)$ corresponding to $\{f_\nu(x)\}$ such that $\|\psi_k - L_n^*\| = E_n(\psi_k;\ \{f_\nu\})$, it follows

$$\|\psi_k - L_n^*\| = E_n(\psi_k;\ \{f_\nu\}) \leqq K_{11} \cdot \|\psi_k^{(r)}\| = 0,$$

and therefore $L_n^*(x) \equiv \psi_k(x)$, which proves our statement.

References

[1] A. N. KOLMOGOROFF, Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse, *Annals of Math.*, **37** (1936), 107—110.
[2] W. RUDIN, L^2-approximation by partial sums of orthogonal developments, *Duke Math. J.*, **19** (1952), 1—4.
[3] B. SZ.-NAGY, Approximation properties of orthogonal expansions, *Acta Sci. Math.*, **15** (1953/54), 31—37.
[4] V. M. TIHOMIROV, Diameters of sets in function spaces and the theory of best approximation, *Uspehi Mat. Nauk*, **15** (1960), 81—120. (Russian)
[5] G. G. LORENTZ, Lower bounds for the degree of approximation, *Trans. Amer. Math. Soc.*, **97** (1960), 25—34.

(Received August 5, 1965)