Journal of Approximation Theory **106**, 58–65 (2000)
doi:10.1006/jath.2000.3477, available online at http://www.idealibrary.com on IDEAL®

COMMENTARY

On the Article by V. L. Goncharov, ''The Theory of the Best Approximation of Functions''

Vladimir M. Tikhomirov

Moscow State University, Moscow 119899, Russia

Received February 21, 2000; accepted February 21, 2000;
published online August 30, 2000

Research in approximation theory in Russia dates back to P. L. Chebyshev's memoir "Théorie des mécanismes connus sous le nom de parallélogrammes" (*Mém. Prés. Acad. Imp. Sci. Pétersb. Divers Savants*, 1854, **VII**, 539–568). This memoir posed the problem of the best approximation of functions by polynomials and presented the first results concerning exact expressions for such approximations.

An interesting article by V. L. Goncharov, "The theory of best approximation of functions," included in the collection "Scientific Heritage of Chebyshev, Mathematics," Moscow, 1945, reviews Chebyshev's work and work of his collaborators in that early period when approximation theory was being established in Russia.

The present paper provides a brief commentary on the part of Goncharov's article devoted to the development of Chebyshev's ideas.

Goncharov justly points out that Chebyshev's memoir contains "a series of mathematical facts and ... statements of utmost importance, undoubtedly forming the basis of his theory." Indeed, Chebyshev gives exact expressions for the best approximations to many different functions. The same subject is treated in a second memoir of Chebyshev, "Sur les questions de minima qui se rattachent à la représentation approximative des fonctions" (*Mém. Acad. Imp. Sci. Pétersb. (6) Sci. Math. Phys.* **VII** (1859), 199–291).

Let us consider both topics (foundation of the general theory and exact solutions) through the eyes of contemporary mathematicians.

1. STATEMENT OF PROBLEMS

In the first-mentioned memoir, Chebyshev wrote: "Soit $f(x)$ une fonction donnée, U un polynome du degré n avec des coefficients arbitraires. Si l'on choisit ces coefficients de manière à ce que la différence $f(x)-U$, depuis

0021-9045/00 $35.00

$x = a - h$, jusqu'à $x = a + h$, reste dans les limites le plus rapprochées de 0, la différence $f(x) - u$ jouira, comme on le sait, de cette propriété: « Parmi les valeurs les plus grandes et les plus petites de la différence $f(x) - U$ entre les limits $x = a - h$, $x = a + h$, on trouve au moins $n + 2$ fois la même valeur numérique »."

This problem has the following general geometric interpretation. Let X be a normed linear space, let A be some subset of X (the *approximation set*), and let $x \in X \backslash A$ be an arbitrary element. The problem of approximating the fixed element x from the fixed set A requires us to solve the following extremal problem:

$$\|x - \xi\| \to \min, \qquad \xi \in A. \tag{P}$$

The *value* of this problem is the distance between x and A in the metric of the space X. A *solution* of this problem, i.e., an element $\hat{\xi} \in A$ such that $d(x, A, X) := \inf_{\xi \in A} \|x - \xi\| = \|x - \hat{\xi}\|$, is called an *element of best approximation*.

Some natural questions could be asked about the problem (P):

(1) Does there exist a solution to the problem or not?

(2) What are the conditions satisfied by a solution (necessary, sufficient, necessary and sufficient)?

(3) Is a solution unique or not?

(4) Is it possible to express a solution in explicit form?

And so on. Chebyshev was mainly interested in the last question, i.e., in "exact solutions" (other questions were considered later).

2. CHEBYSHEV'S ALTERNATION THEOREM AND ITS GENERALIZATIONS

In the quotatron above, Chebyshev discusses necessary conditions on the solution of the problem of best approximation of a continuous function $x(\cdot)$ by algebraic polynomials. Here is how *Chebyshev's alternation theorem* usually is formulated these days.

THEOREM 1. *Let $x(\cdot)$ be a continuous function on a finite interval $[t_0, t_1]$. Then*:

(1) *a polynomial $\hat{p}(\cdot)$ of best approximation exists*;

(2) *the polynomial $\hat{p}(t) = \sum_{k=0}^{n} \hat{y}_{k+1} t^k$ is a polynomial of best approximation iff the difference $x(\cdot) - \hat{p}(\cdot)$ has an $(n+2)$-alternation (this means*

that there is a monotone sequence of $n+2$ *points at which the difference* $x(\cdot)-\hat{p}(\cdot)$ *takes on its absolutely largest value, with alternating sign*);

(3) *the polynomial of best approximation is unique.*

Goncharov's article is devoted to the history of this theorem. The present article comments on the modern view of the subject.

(A) *Existence of an Element of Best Approximation.*

As a rule, existence theorems in approximation theory are corollaries of the following general principles of functional analysis:

Weierstrass–Lebesgue Compactness Principle. A lower semicontinuous function on a compact topological space attains its (global) infimum.

Banach–Alaoglu–Bourbaki Theorem. The polar of a neighborhood is a weakly compact set (in every topology generated by duality).

As a corollary of these principles, the existence of elements of best approximation in any finite-dimensional subspace of a Banach space (in particular, in Chebyshev's case), in any closed subset of a reflexive Banach space, and in many other different cases, can be obtained. Existence of best ε-nets, of extremal subspaces in conjugate spaces, and so on, are also corollaries of the compactness principles quoted above.

(B) *Duality and Criteria*

Let us suppose that the approximation set A is convex. Then the problem (P) is a problem of convex programming. One of the main principles of convexity is the following: *Every convex phenomenon* (*a convex set*, *a convex function*, *or a convex problem*) *has a dual description in the dual space*. An application of this principle to the problem (P) leads to the following result:

Theorem 2. *Let X be a normed space and A a convex set. Then*

$$d(x, A, X) = \sup_{x^*} (\langle x^*, x\rangle - sA(x^*) : \|x^*\| \leqslant 1),$$

where $\langle x^*, x\rangle$ is the value of the linear functional x^* at the element x, and

$$sA(x^*) := \sup_{x \in A} \langle x^*, x\rangle$$

is the value at x^* of the support function of the set A.

Many duality theorems in approximation theory were obtained by S. M. Nikolskii, M. G. Krein, S. Ya. Khavinson, A. L. Garkavi and other mathematicians.

We now formulate a general criterion for an element to be a best approximation from a convex set.

THEOREM 3. *If A is a convex set, then $\hat{\xi}$ is a solution of problem (P) iff there exists a linear functional x^* such that*

$$\|x^*\|_{X^*}=1, \qquad \langle x^*, x-\hat{\xi}\rangle = \|x-\hat{\xi}\|, \qquad \langle x^*, \hat{\xi}\rangle = sA(\hat{\xi}).$$

Both results are proved by standard methods of convex analysis.

In preparation for the short survey of the approximation theory in C-spaces offered at the end of this section, we now recall that Chebyshev's problem is a particular case of the more general problem of convex programming:

Let T be a compact set and $f\colon T\times\mathbb{R}^n\to\mathbb{R}^n$ a function of variables $t\in T$ and $y\in\mathbb{R}^n$. Suppose that the function $f(t,\cdot)$ is convex for all $t\in T$. Consider the problem

$$\sup_{t\in T} f(t, y)\to \min. \tag{P'}$$

Now we formulate an important theorem of convex analysis which gives a universal approach to criteria for elements of best approximation in C-spaces.

THEOREM 4 (Decomposition (or Refinement) Theorem). *Let T be a compact topological space, $f\colon T\times\mathbb{R}^n\to\mathbb{R}$ (or, equivalently, $(f_t(\cdot)\colon t\in T)$ a family of real functions) such that:*

(a) *$f(t,\cdot)$ is convex for all $t\in T$;*

(b) *$f(\cdot, y)$ is upper semicontinuous for all $y\in\mathbb{R}^n$;*

(c) $m:=\inf_y \max_t f(t,y) > -\infty$.

Then there exist $r\leqslant n+1$ points $\{\tau_i\}_{i=1}^r$ such that

$$m=\inf_y \max_{1\leqslant i\leqslant r} f(\tau_i, y).$$

This theorem allows one to state a criterion for the solution of the problem

$$\max_t f(t, y)\to \min, \tag{P''}$$

with f as described in the theorem. For its formulation and derivation, we need the following important definition and two theorems from convex analysis.

DEFINITION. Let $g: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function. The set

$$\partial g(\hat{y}) := \{z \in \mathbb{R}^n : g(y) - g(\hat{y}) \geqslant \langle z, y - \hat{y} \rangle\} \tag{1}$$

(with $\langle a, b \rangle := \sum_{i=1}^n a_i b_i$) is called the *subdifferential of g at the point $\hat{y}$.*

Dubovitskii–Miliutin Formula. If f is the pointwise maximum of the two continuous convex functions f_1 and f_2, i.e., $f(y) := \max(f_1(y), f_2(y))$, and $f_1(\hat{y}) = f_2(\hat{y})$, then

$$\partial f(\hat{y}) = \operatorname{co}(\partial f_1(\hat{y}) \cup \partial f_2(\hat{y})) \tag{2}$$

(co A means the convex hull of A).

Fermat's Theorem for Convex Functions. Let f be a convex function. Then $\hat{y}$ is a minimum of f iff

$$0 \in \partial f(\hat{y}). \tag{3}$$

From the decomposition theorem and these results from convex analysis, one concludes the following.

THEOREM 5. *If $\hat{y}$ is a solution of the problem (P''), then there exist $r \leqslant n+1$ points, $\{\tau_i\}_{i=1}^r$, in T and positive scalars, $(\alpha_i)_{i=1}^r$, with $\sum_{i=1}^r \alpha_i = 1$ such that*

(a) $f(\tau_i, \hat{y}) = m$ $(:= \inf_y \max_t f(t, y))$;

(b) $0 \in \sum_{i=1}^r \alpha_i \, \partial f_{\tau_i}(\hat{y})$.

Chebyshev's criterion, i.e., Theorem 1, is an immediate corollary. In Chebyshev's case, $f(t, y) := |x(t) - \sum_{k=0}^n y_{k+1} t^k|$, $T = [t_0, t_1]$. It follows from Theorem 5 that there exist $r \leqslant n+2$ points $\{\tau_i\}_{i=1}^r$ in T and positive scalars $\{\alpha_i\}_{i=1}^r$ with $\sum_{i=1}^r \alpha_i = 1$ such that

$$\left| x(\tau_i) - \sum_{k=0}^n \hat{y}_{k+1} t^k \right| = m := \|x(\cdot) - \hat{p}(\cdot)\|_{C([t_0, t_1])} \tag{4}$$

and

$$\sum_{i=1}^r \alpha_i \operatorname{sgn}(x(\tau_i) - \hat{p}(\tau_i)) \, \tau_i^k = 0, \qquad k = 0, 1, ..., n. \tag{5}$$

It follows that the homogeneous linear system in (5) of $n+1$ equations in the $r \leqslant n+2$ unknowns $z_i := \alpha_i \operatorname{sgn}(x(\tau_i) - \hat{p}(\tau_i)))$ has a nontrivial solution. Hence (as follows from properties of Vandermonde's matrix), $r = n+2$, and $z_i z_{i+1} < 0$, $0 \leqslant i \leqslant n+1$. Chebyshev's criterion is proved.

In a similar way, one can prove the criteria of Bernstein, Kolmogorov, Zukhovitsky and Krein, Zukhovitsky and Stechkin, Singer, and many others.

The majority of such results are immediate corollaries of the following reformulation of the decomposition theorem (it includes also some results devoted to the problem of approximation of single elements in the case of a metric which is "nonsymmetric").

Let T be a topological space, Y a real or complex locally convex linear topological space, L_n an n-dimensional subspace of the space $C(T, Y)$ of continuous mappings from T into Y, $x(\cdot) \in C(T, Y) \backslash L_n$, $\{p_t(\cdot)\}_{t \in T}$ a family of sublinear continuous functionals on Y (which may be nonsymmetric).

In this setting, consider the following extremal problem:

$$\sup_{t \in T} p_t(x(\cdot) - \xi(\cdot)) \to \min, \qquad \xi(\cdot) \in L_n. \tag{P'''}$$

THEOREM 6. *If, in* (P'), *the mapping* $t \mapsto p_t(y)$ *is upper semicontinuous for each* $y \in Y$, *then* $\hat{\xi}(\cdot)$ *is a solution of* (P') *iff there exist r points* $\{t_j\}_{j=1}^r$ *in T (with* $r \leqslant n+1$ *in the real and* $r \leqslant 2n+1$ *in the complex case), and r functionals* $\{y_j^*\}_{i=1}^r$ *in* Y^*, *and r positive numbers* $\{\alpha_j\}_{j=1}^r$ *with* $\sum_{j=1}^r \alpha_j = 1$, *such that*

(a) $\operatorname{Re} \langle y_j^*, y \rangle \leqslant p_{t_j}(y)$, *for all* $y \in Y$;

(b) $p_{t_j}(x(t_j) - \hat{\xi}(t_j)) = \sup_{t \in T} p_t(x(\cdot) - \hat{\xi}(\cdot))$, *for all j*;

(c) $\sum_{j=1}^r \alpha_j \operatorname{Re} \langle y_j^*, y(t_j) \rangle = 0$, *for all* $y(\cdot) \in L_n$

(*with* Re *of course ignorable in the real case*).

3. EXACT EXPRESSION FOR THE ELEMENTS OF BEST APPROXIMATION OF INDIVIDUAL FUNCTIONS

Chebyshev and Zolotarev found explicit expressions for polynomials and rational functions of best approximation to some important functions $x(\cdot)$. Here are three examples.

THEOREM 7. (a) *For* $x(\cdot) = t^n$, $d(x(\cdot), \mathscr{P}_{n-1}, C([-1, 1])) = 2^{-(n-1)}$, *and*

$$x(t) - \hat{p}(t) = T_n(t) := 2^{-(n-1)} \cos n \arccos t$$

(*Chebyshev; the polynomials* $T_n(\cdot)$ *are called Chebyshev polynomials*).

(b) *For* $x(t)=1/(t-a)$ *with* $|a|>1$, $d(x(\cdot), \mathscr{P}_n, C([-1,1]))=M:=(a-\sqrt{a^2-1})/(a^2-1)$, *and*

$$\frac{1}{t-a}-\hat{p}(t)=\frac{M}{2}\left(\frac{\alpha-w}{1-\alpha w}\,w^n+\frac{1-\alpha w}{\alpha-w}\,w^{-n}\right),$$

$$t=\frac{1}{2}(w+w^{-1}).$$

(c) *For* $x(t)=t^n+\sigma t^{n-1}$, $d(x(\cdot), \mathscr{P}_{n-2}, C([-1,1]))=\|Z_{n\sigma}(\cdot)\|$, *where* $Z_{n\sigma}$ *is a Zolotarev polynomial and has the parametric representation*

$$Z_{n\sigma}(t)=C\left(\left(\frac{H(K/n-w)}{H(K/n+w)}\right)^n+\left(\frac{H(K/n+w)}{H(K/n-w)}\right)^n\right),$$

$$t=\frac{\operatorname{sn}^2 w+\operatorname{sn}^2 K/n}{\operatorname{sn}^2 w-\operatorname{sn}^2 K/n}$$

with $H(\cdot)$, $\operatorname{sn}(\cdot)$ *elliptic functions, and* C *a known constant.*

In view of these results, one is tempted to recall the words of Jacques Hadamard: "The shortest path between two truths in real analysis lies in the complex domain."

In all the cases discussed (and indeed in almost all cases considered at that first stage of the development of approximation theory), the difference $e(\cdot):=x(\cdot)-\hat{p}(\cdot)$ between the approximated function $x(\cdot)$ and the polynomial $\hat{p}(\cdot)$ of best approximation is represented in the parametric form $e=f(w)$, $t=g(w)$, such that $e(\cdot)$ has the necessary alternation on $[-1,1]$.

This idea is nicely illustrated by the first two examples of the theorem. In case (a) the representation in question is

$$e=\tfrac{1}{2}(w^n+w^{-n}), \qquad t=\tfrac{1}{2}(w+w^{-1}).$$

It is easy to see that the function $t\mapsto e(t)$ on the complex domain has no singularities except at $t=\infty$, where it has a pole of degree n with the residue equal to 1. Hence $e(\cdot)\in\mathscr{P}_n$ and $e(t)=t^n+\cdots$. If $w=\exp(i\theta)$, $0\leqslant\theta\leqslant\pi$, then t runs from 1 to -1 and $e(\cdot)$ has an $(n+1)$-alternation. From the Chebyshev criterion (Theorem 1), we obtain (a).

In (b), $e(\cdot)$ has two singularities, at $t=a$ and $t=\infty$. In both cases, the singularities are poles (of degree one with the residue equal to 1 at the point $t=a$, and of degree n at $t=\infty$). Therefore, $e(t)=1/(t-a)+\hat{p}(t)$. And again (from the theorem of the argument), $e(\cdot)$ has an $(n+2)$-alternation, therefore, by Chebyshev's criterion, it follows that $\hat{p}(\cdot)$ is the polynomial of best approximation to the function $t \mapsto 1/(t-a)$.

The representation in (c) is also completely natural. The details can be found in Akhiezer's book.

The remainder of the article by Goncharov is devoted to the development of approximation theory in the 20th century.

On the problems considered above (and the problems of unicity) see also [1, 2].

REFERENCES

1. N. I. Akhiezer, "Theory of Approximation," Ungar, New York, 1956.
2. V. M. Tikhomirov, Theory of Approximation, *in* "Encyclopaedia of Mathematical Sciences," Vol. 14, pp. 93–255, Springer-Verlag, Berlin/New York, 1990.