V.L. Goncharov

The Theory of Best Approximation of Functions

1. Approximating formulæ by Poncelet

The history of mathematics testifies to the fact that the main idea of a theory often is more or less clearly formulated, though sometimes only in passing, by the predecessors of the mathematician whose name later becomes inseparably associated with it.

The authenticity of the approximation method due to Chebyshev can be thought of as well established. This makes it all the more interesting to investigate the source of the impetus that led our genius fellow countryman to his brilliant and deep constructions.

There is nothing mysterious about it. Chebyshev himself tells us, in a perfectly clear way, about the purpose of his research in approximation theory, and names the person whose results were his starting point. He states: "From among the many research subjects that I encountered in studying and comparing different mechanisms of motion transfer, especially in a steam engine, where efficiency and reliability depend much on the way the power of steam is transferred, I was especially occupied by the theory of mechanisms known as parallelograms... While trying to derive the rules for constructing specific parallelograms directly from their properties, I encountered problems in analysis that were not well known then. Everything done in this area is due to Mr. Poncelet, a member of the Paris Academy who is well known in applied mechanics. His formulæ are widely used in computing the friction † in mechanisms..." [6]. And in a different place [1]: "With respect to the approximation method just mentioned, we are only equipped with the findings of Poncelet, who gave linear formulæ often used for approximating the following three expressions:

$$\sqrt{x^2+y^2}, \quad \sqrt{x^2-y^2}, \quad \sqrt{x^2+y^2+z^2}\text{...''}$$

It is not the applications of Chebyshev's theory that are of interest to us at the moment, but rather its genesis. The results of Poncelet just mentioned are stated (other than in lithographic lecture notes not available to us) in his work on approximate radical values published in Crelle's Journal in 1835 [64]. The results are not likely to be known to our reader, and it therefore seems appropriate to give a feeling for them here by presenting the following simplest example.

In trying to simplify the computation of the efficiency of certain mechanisms (as reported by Résal), Poncelet (who, at the time, was serving as a captain and teaching in the School of Engineering and Artillery in Metz) poses the problem of finding an approximating formula for $\sqrt{a^2+b^2}$ of the form

$$\sqrt{a^2+b^2} \sim \alpha a + \beta b$$

in such a way that the absolute value of the relative error

$$\frac{\alpha a + \beta b}{\sqrt{a^2+b^2}} - 1,$$

which obviously depends only on the ratio $\frac{a}{b}$, have the smallest possible maximum over all values of a and b satisfying $\frac{a}{b} \geq k$ (where k is a positive given number). In other words, using modern notation, one must choose α and β so that the expression

$$\max_{x \geq k} |r(x)|,$$

where

$$r(x) = \frac{\alpha x + \beta}{\sqrt{x^2+1}} - 1,$$

† lit.: useless resistance

attains its minimum. The function $r(x)$ is increasing for $x < \frac{\alpha}{\beta}$, and is decreasing for $x > \frac{\alpha}{\beta}$. Therefore the maximum, which we will denote by $F(\alpha,\beta)$, is equal to the largest of the three numbers $|r(k)|$, $|r(\frac{\alpha}{\beta})|$, and $|r(\infty)|$. Let

$$k = \operatorname{ctg}\omega,$$
$$\frac{\alpha k+\beta}{\sqrt{k^2+1}} = \alpha\cos\omega + \beta\sin\omega = \alpha',$$
$$\frac{-\alpha+\beta k}{\sqrt{k^2+1}} = -\alpha\sin\omega + \beta\cos\omega = \beta'.$$

It then turns out that the values of the continuous function $F(\alpha,\beta)$ on the domains

$$\begin{aligned}
&(\mathrm{A})\ \beta^2 > 4(1-\alpha), \quad {\beta'}^2 > 4(1-\alpha'),\\
&(\mathrm{B})\ \beta^2 < 4(1-\alpha), \quad \alpha' > \alpha,\\
&(\mathrm{C})\ {\beta'}^2 < 4(1-\alpha'), \quad \alpha' < \alpha,
\end{aligned}$$

are equal, respectively, to

$$r\left(\frac{\alpha}{\beta}\right) = \sqrt{\alpha^2+\beta^2-1}, \quad -r(\infty) = 1-\alpha, \quad -r(k) = 1-\alpha'.$$

The latter quantities, as is obvious from geometric considerations,* are minimized at the point common to the boundaries of the three domains, namely at

$$\begin{aligned}
\alpha &= \frac{\cos\frac{\omega}{2}}{\cos^2\frac{\omega}{4}} = \frac{2}{1+\sqrt{2(k^2+1)-2k\sqrt{k^2+1}}},\\
\beta &= \frac{\sin\frac{\omega}{2}}{\cos^2\frac{\omega}{4}} = \frac{2(\sqrt{k^2+1}-1)}{1+\sqrt{2(k^2+1)-2k\sqrt{k^2+1}}}.
\end{aligned} \tag{1}$$

The relative error ε is the value of $F(\alpha,\beta)$ at this point,

$$\varepsilon = 1-\alpha = 1-\alpha' = \sqrt{\alpha^2+\beta^2}-1,$$

given by the formula

$$\varepsilon = \operatorname{tg}^2(\frac{\omega}{4}) = \frac{\sqrt{2(k^2+1)-2k\sqrt{k^2+1}}-1}{\sqrt{2(k^2+1)-2k\sqrt{k^2+1}}+1}. \tag{2}$$

---

* Let us view the parameters α and β as the Cartesian coordinates in a plane. Then the whole plane $O\alpha\beta$ is divided into three domains, by the line $\alpha = \alpha'$, by the parabola $\beta^2 = 4(1-\alpha)$, and by the congruent parabola obtained by rotating the first one about the focus O by the angle ω. In the domain (A), which is outside of both parabolas, the function $F(\alpha,\beta)$ is equal to $\sqrt{\alpha^2+\beta^2}-1$, and it attains its smallest value at the boundary point P closest to the origin. In the domain (B), which is inside of the first parabola and to the same side from the line $\alpha = \alpha'$ as the semi-axis α', the function $F(\alpha,\beta)$ is equal to $1-\alpha$, and, therefore, attains its maximum, again, at the point P. The same is true for the domain (C). It is worth noting that the analytic expressions defining the function $F(\alpha,\beta)$ in the different domains are such that in each case the minimum is attained on the boundary, at the point common to the three domains.

Poncelet presents the following table:

a and b	k	α	β	ε
arbitrary	0	0.82840	0.82240	0.17160 or $\frac{1}{6}$
$a > b$	1	0.96046	0.39783	0.03954 or $\frac{1}{25}$
$a > 2b$	2	0.98592	0.23270	0.10408 or $\frac{1}{71}$
$a > 3b$	3	0.99350	0.16123	0.00650 or $\frac{1}{154}$
$a > 4b$	4	0.99625	0.12260	0.00375 or $\frac{1}{266}$
$a > 5b$	5	0.99757	0.09878	0.00243 or $\frac{1}{417}$
$a > 6b$	6	0.99826	0.08261	0.00174 or $\frac{1}{589}$
$a > 7b$	7	0.99875	0.07098	0.00125 or $\frac{1}{800}$
$a > 8b$	8	0.99905	0.06220	0.00095 or $\frac{1}{1049}$
$a > 9b$	9	0.99930	0.05535	0.00070 or $\frac{1}{1428}$
$a > 10b$	10	0.99935	0.04984	0.00065 or $\frac{1}{1538}$

Poncelet solves the same problem for $\sqrt{a^2 - b^2}$ by a similar method, where he introduces the restriction $k \leq \frac{a}{b} \leq k'$. As concerns the square root of the sum of three squares $\sqrt{a^2 + b^2 + c^2}$, he reduces the problem to the first one as follows

$$\sqrt{a^2 + b^2 + c^2} = \sqrt{a^2 + (\sqrt{b^2 + c^2})^2} \sim \alpha a + \beta\sqrt{b^2 + c^2} \sim$$
$$\sim \alpha a + \beta(\alpha' b + \beta' c) = \alpha a + \alpha'\beta b + \beta\beta' c.$$

Note that, apart from the work of several French authors ([50], [65]), a more detailed solution of the last problem under the restriction of the form $k \leq \frac{a}{b} \leq k'$ was obtained much later, in the spirit of Poncelet, by A.A. Markov [59].

The following quote from Poncelet's work is interesting in two ways. First, it tells us that Poncelet is following even earlier authors in posing the problem. Second, in somewhat vague and cautious terms, it suggests possible generalizations, and outlines the path followed later by Chebyshev. In connection with the first of the problems considered, Poncelet says the following:

"The method that allowed us to obtain the general expressions for α, β, and ϵ in terms of k and ω, can, obviously, be applied to any function of two variables a and b, more or less complex, given that the goal is to approximate it by a linear expression of the form $\alpha a + \beta b + \gamma$, provided that the formula for the relative error that results from this substitution takes on its maximum or minimum in the range of values of a and b that are being considered. In some cases, the method might also be applied to functions of any number of variables a, b, c, d, etc., if one starts with arguments analogous to those by Laplace and Fourier that allowed them to determine the values of the unknown variables from a system of equations so that the absolute value of the maximum error resulting from substituting experimental data for the unknowns is minimized.* Indeed, the whole difficulty is to find, in each particular case, analytic expressions for the extrema of the possible error; to equate their absolute values; and it then becomes possible (provided that the number of equations thus obtained is equal to the number of unknowns) to calculate the values of the variables that satisfy the required conditions."

And further:

"These remarks show quite clearly that the method can be applied under very general circumstances. It allows one to replace (provided that the replacement is possible at all) any complex function of any number of variables by another function that is simpler and more suitable for computations or analytical transformations. The example we have considered gives one a feeling for the machinery that should be applied in each particular case, as well as for the advantages of our procedure compared to the traditional methods – decomposing in a series or in a continued fraction".

Even though the problems considered by Poncelet are modest, it can be seen from his words that this distinguished researcher had the foresight to appreciate the scientific and practical importance of his principle. Choose the parameter values in such a way that the maximum of the error is minimized.

* Poncelet cites "Mécanique celeste" by Laplace, Ed. 2, t. 2, page 126, and "Analyse des équations" by Fourier, part 1, page 81.

2. The memoir "Théorie des mécanismes connus sous le nom de parallélogrammes"

The first of the two major memoirs that contain Chebyshev's research on the best approximation of functions was presented to the Academy of Sciences in January 1853. This is soon after Chebyshev had returned from a long trip abroad, where he visited the most important European scientific and industrial centers, and where he paid equal attention to studying factories, plants, and different kinds of "interesting subjects in applied mechanics", as well as making personal contacts and "conversing" with famous "geometers", mainly French ones.

The name of Poncelet is not among those mentioned by Chebyshev in his report but, undoubtedly, during his stay in Paris or in Metz, Chebyshev was in touch with the circle of his ideas. It is difficult to decide whether he had considered problems of that kind before his trip abroad. In any case, it is significant that all his manuscripts written in the preceding period were devoted to pure mathematics (number theory, integration, probability), while, after this trip, Chebyshev showed a lively interest in practical applications of mathematical problems. In particular, if the actual words of the author are to be taken literally, Chebyshev had undertaken his research to advance Watt's parallelogram theory. Though it is mentioned that the applications of the general formula that were obtained "are not limited to investigating these mechanisms", and that "applied mechanics and other applied sciences have a whole range of questions for which these formulæ are necessary". Hence appeared a treatise entitled "The theory of mechanisms known by the name of parallelograms", which was published in the "Notes of the Academy of Sciences" in French.

Chebyshev writes: "Because of the lack of time and the breadth of the subject, I was only able to finish the first part of my note." The structure of this part is roughly the following. After an extensive introduction of an exclusively technical nature, which illustrated the drawbacks (inaccuracy of motion) of Watt's mechanisms, Chebyshev '*ex abrupto*' turns to solving a purely mathematical problem, which we are going to consider now, and only his closing words bring us back to the problem that he initially claimed to be concerned with: "In the following paragraphs, we shall illustrate the application of the derived formulæ for finding the parameters of the parallelograms that satisfy the conditions that make the precision of motion best possible". Where are those following paragraphs? There is no sign of them. Apparently, the initial plan was not implemented, and the incompatible pieces broke apart. There further appeared, on one hand, the excellent self-contained memoir "Sur les questions des minima", which contains the basics of the mathematical theory of approximation and makes "Theory of mechanisms" look like a mere draft, and, on the other hand, a number of later articles and notes on hinged mechanisms, which were of so much interest to Chebyshev in the second half of his scientific activity.*

When analyzing the content of the memoir [1], one notices that, even in its purely mathematical part, this treatise is somewhat unbalanced. *In passing*, Chebyshev establishes here a series of mathematical facts, and states results of major importance, which are undoubtedly fundamental to his theory and contain the source of its further development. *Explicitly*, however, the subject of the paper is a rather technical question of limited importance, though difficult to solve, requiring cumbersome computations and a great deal of mathematical insight. The author himself says that this question is motivated by applications, but does not go into details. Common sense and caution make one act in this way when the authenticity of the results obtained is in doubt, or when the results themselves are at risk of not being appreciated by biased arbiters.

By the way, the memoir [1] is apparently the first to contain a formulation of the general problem. Given a continuous function f, find a polynomial of a given degree such that "the maximum of its deviation from $f(x)$ in a given interval is smaller than that of all other polynomials of the same degree" [1]. In other words, given an interval $[a, b]$, one has to find the coefficients p_i of the n-th degree polynomial

$$P(x) = p_0 x^n + p_1 x^{n-1} + \cdots + p_n$$

so that the expression

$$\max_{a \le x \le b} |f(x) - P(x)|,$$

which depends on the coefficients p_i, is minimized.

* Of interest is an introductory paragraph to the note "On one mechanism" [7] that testifies that theory and practice were inseparable in Chebyshev's work in his last years.

How to find the approximating polynomial $P(x)$? The difference

$$R(x) = f(x) - P(x),$$

"as is known" (according to Chebyshev), necessarily has the following property: "the set of its numerical* maxima and minima in the given interval contains the same number at least $n+2$ times". In other words, if $|R(x)| \leq L$ for $a \leq x \leq b$, and there exist points x in which $|R(x)| = L$ (the "deviation points"), then the number of such points is at least $n+2$ (the latter number is one more than the number of parameters p_i). Neither here nor in any other place does Chebyshev discuss the existence or the uniqueness of the polynomial $P(x)$, because in all particular cases he is able to find the unique solution. Similarly, there is no statement made about the number of points of positive deviation $R(x) = +L$ versus the number of points of negative deviation $R(x) = -L$, nor about their relative positioning.

Chebyshev does not give a proof of his statement, so there remains the question of whether the property mentioned was formulated by someone else (in France, in Poncelet' school, perhaps, during a conversation) or was established by Chebyshev himself, and, due to its apparent simplicity, was considered as not deserving a proof. The words "is known" can be understood either way.

In addition, while tacitly assuming that the function $f(x)$ is differentiable, Chebyshev takes up the problem of finding the polynomial $P(x)$. At the points of deviation x_i, besides the conditions

$$R^2(x_i) = L^2,$$

the equality

$$(x_i - a)(x_i - b)R'(x_i) = 0$$

holds, and therefore there are at least $2n+4$ equations, from which, theoretically speaking, the points of deviation x_i, the coefficients p_i, and, finally, the deviation L itself can be determined. If the system leads to several polynomials rather than to a single one, then Chebyshev, according to his way of thinking, would select the correct polynomial by direct comparison.

Actually, the particular cases considered are such that either solving the system of algebraic equations can be avoided, or, at least, its order can be reduced. For simplicity, set $a = -1$ and $b = +1$. In this particular memoir, we are concerned, in fact, with only the one particular case, where

$$f(x) = x^{n+p} \quad (p \text{ is an integer} \geq 1).$$

It follows on algebraic grounds that, in this case, the fraction

$$\frac{R^2(x) - L^2}{R'^2(x)}$$

can be reduced to

$$\frac{(x^2-1)A(x)}{B^2(x)},$$

where $A(x)$ and $B(x)$ are polynomials of degree $2(p-1)$ and $p-1$, respectively. In this way, Chebyshev arrives at the differential equation

$$\frac{\mathrm{d}R}{\sqrt{R^2 - L^2}} = \frac{B(x)\,\mathrm{d}x}{\sqrt{(x^2-1)A(x)}}. \tag{3}$$

If $p = 1$, then $A(x)$ and $B(x)$ are constants, so integrating the equation, using the initial conditions, and comparing the leading coefficients gives the result

$$R(x) = \frac{1}{2^n} T_{n+1}(x),$$

* that is, absolute values of

where

$$T_n(x) = \frac{1}{2}[(x+\sqrt{x^2-1})^n + (x-\sqrt{x^2-1})^n] = \cos n \arccos x.* \tag{4}$$

So, the sought-for polynomial has the form

$$P(x) = x^{n+1} - \frac{1}{2^n}T_{n+1}(x).$$

Simultaneously, the following problem is solved. Find a polynomial of degree n with a given coefficient σ_0 of x^n that deviates least from zero on the interval $[-1,+1]$.

This polynomial is

$$\frac{\sigma_0}{2^{n-1}}T_n(x).$$

The case $p=2$ is essentially the same, because

$$R(x) = \frac{1}{2^{n+1}}T_{n+2}(x).$$

As far as the case $p \geq 3$ is concerned, without solving the problem completely, Chebyshev points to a way of doing so, while ingeniously using the results of his 1847 dissertation *pro venia legendi* "On integration by means of logarithms" [14]. While deriving from (3) that the integral

$$\int \frac{B(x)\,\mathrm{d}x}{\sqrt{(x^2-1)A(x)}}$$

can be represented in the form

$$\frac{1}{2}\ln\frac{R(x)+\sqrt{R^2(x)-L^2}}{R(x)-\sqrt{R^2(x)-L^2}},$$

he concludes that the polynomial $A(x)$ must satisfy $p-1$ conditions. The same number of conditions comes from the fact that the polynomial $R(x)$ does not contain the powers $x^{n+1},\dots,x^{n+p-1}$. All these conditions, are, theoretically speaking, sufficient for computing $R(x)$.

The above-mentioned results are undoubtedly of major importance, however they look somewhat like a by-product in the memoir [1]. Let us now turn to the problem that Chebyshev describes as its major subject.**

Let us assume, as does Chebyshev, that the given function $f(x)$ is analytic at the point $x=a$, that is, it can be represented by its Taylor series

$$f(x) = \sum_0^\infty k_m(x-a)^m$$

in some neighborhood of that point. Let $p(x,h)$ be a polynomial of degree n that approximates $f(x)$ best in the sense of Chebyshev on the interval $a-h \leq x \leq a+h$. Given the coefficients k_m, one has to find a power series expansion of the polynomial $p(x,h)$ in powers of h.

If we set

$$F(x) \equiv f(a+x) = \sum_0^\infty k_m x^m, \quad P(x,h) = p(a+hx,h),$$

then the problem can be reformulated in the following way. Given the coefficients k_m, find the expansion of the polynomial $P(x,h)$ in powers of h that provides a best approximation to the function $F(hx) \equiv \sum_0^\infty k_m h^m x^m$ on the fixed interval $-1 \leq x \leq +1$.

---

* Note that the trigonometric form of the polynomial that deviates least from zero can be encountered in Chebyshev's 1873 work. The notation $T_n(x)$ was first used by S.N. Bernstein in his dissertation.

** The exposition in the following paragraph is somewhat modernized.

We cannot find in [1] a proof of the fact that the polynomial $P(x,h)$, as well as the smallest deviation $L(x)$, are analytic functions of h (at $h=0$). As a matter of fact, Chebyshev does not need it, since explicitly, and from the viewpoint of possible applications, he is only interested in finding an approximating polynomial $P_N(x,h)$ of degree n with respect to x and of degree N with respect to h such that, as $h\to 0$,

$$P(x,h)=P_N(x,h)+O(h^{N+1}).$$

The question of convergence of $P_N(x,h)$ to $P(x,h)$ as $N\to\infty$ is ignored.

Chebyshev solves the problem posed in several steps.

1. Let $N=n$. Denote by $S(x)$ the partial Taylor sum of the function $F(x)$,

$$S(x)=\sum_0^n k_m x^m.$$

Then

$$\max_{|x|\le 1}|S(hx)-F(hx)|=O(h^{n+1}),$$

and, therefore, by the definition of the polynomial $P(x,h)$, similarly,

$$\max_{|x|\le 1}|P(x,h)-F(hx)|=O(h^{n+1}).$$

Thus,

$$\max_{|x|\le 1}|P(x,h)-S(hx)|=O(h^{n+1}),$$

which implies that $P(x,h)\equiv S(hx)$.

2. Let $k_{n+1}=k_{n+2}=\cdots=k_{n+p-1}=0$, but $k_{n+p}\ne 0$. Assume that $N=n+p$. Then the polynomial $P(x,h)$ must have the form

$$P(x,h)=S(hx)+O(h^{n+p}),$$

or

$$P(x,h)=S(hx)+h^{n+p}Q(x,h),$$

where $Q(x,h)$ is a polynomial of degree n with respect to x, and we shall assume

$$Q(x,h)=Q_0(x)+hQ_1(x)+\cdots+h^\nu Q_\nu(x)+O(h^{\nu+1}),\quad (\nu=0,1,2,\ldots).$$

By the definition of $P(x,h)$, the polynomial $Q(x,h)$ should be selected in such a way that the expression

$$\max_{|x|\le 1}|F(hx)-P(x,h)|=\max_{|x|\le 1}\Big|\sum_{m=n+p}^{\infty}k_m h^m x^m-h^{n+p}Q(x,h)\Big|=$$
$$=h^{n+p}\max_{|x|\le 1}|k_{n+p}x^{n+p}-Q_0(x)+O(h)|$$

be minimized. This is achieved by minimizing the expression

$$\max_{|x|\le 1}|k_{n+p}x^{n+p}-Q_0(x)|.$$

Therefore, the polynomial $Q_0(x)$ of degree n deviates least from the function $k_{n+p}x^{n+p}$, so it is k_{n+p} times the polynomial of degree n that deviates least from x^{n+p}, whose construction was discussed earlier. Chebyshev points out that the case $p=1$ "is the only one that makes sense in the parallelogram theory", and in this case, as we have seen, the solution can be obtained via the trigonometric polynomial

$$Q_0=k_{n+1}\{x^{n+1}-\frac{1}{2^n}T_{n+1}(x)\}$$

found by Chebyshev.

3. While assuming that $k_{n+1} \neq 0$ for the rest of the argument, Chebyshev computes the polynomial $Q_1(x)$ in the following way. By a property of $P(x,h)$, the polynomial

$$Q_1(x,h) = \frac{Q(x,h) - Q_0(x)}{h} = Q_1(x) + hQ_2(x) + \cdots$$

of degree n with respect to x minimizes the expression

$$\frac{1}{h^{n+1}} \max_{|x|\leq 1} |F(hx) - P(x,h)| = \max_{|x|\leq 1} |k_{n+2}hx^{n+2} + k_{n+1}\frac{T_{n+1}(x)}{2^n} - hQ_1(x,h) + O(h^2)|,$$

and, therefore, the equations

$$\{k_{n+1}\frac{T_{n+1}(x)}{2^n} + k_{n+2}hx^{n+2} - hQ_1(x,h) + O(h^2)\}^2 = L^2(h)$$

and

$$(x^2-1)\frac{d}{dx}\{k_{n+1}\frac{T_{n+1}(x)}{2^n} + k_{n+2}hx^{n+2} - hQ_1(x,h) + O(h^2)\} = 0$$

have at least $n+2$ common roots. The first of these equations can be written as

$$k_{n+1}^2\frac{T_{n+1}^2(x)}{2^{2n}} + 2hk_{n+1}\frac{T_{n+1}(x)}{2^n}[k_{n+2}x^{n+2} - Q_1(x)] - L^2(h) + O(h^2) = 0. \qquad (5)$$

As far as the second one is concerned, its roots are within $O(h)$ of those of $(x^2-1)T'_{n+1}(x)$, so they have the form

$$x_m(h) = x_m + O(h), \quad \text{where} \quad x_m = \cos\frac{m\pi}{n+1} \quad (m = 0, 1, \ldots, n+1).$$

Note that

$$L(h) = L(0) + hL'(0) + O(h^2) = \frac{|k_{n+1}|}{2^n} + kL'(0) + O(h^2)$$

and

$$T_{n+1}(x_m(h)) = T_{n+1}(x_m + O(h)) = T_{n+1}(x_m) + O(h)T'_{n+1}(x_m) + O(h^2) = (-1)^m + O(h^2).$$

Set $x = x_m(h)$ in (5):

$$2hk_{n+1}\frac{(-1)^m}{2^n}[k_{n+2}x_m^{n+2} - Q_1(x_m)] - h\frac{|k_{n+1}|}{2^{n-1}}L'(0) + O(h^2) = 0.$$

Since this is an identity with respect to h, the coefficient of h has to be zero, i.e.,

$$k_{n+2}x_m^{n+2} - Q_1(x_m) = (-1)^m\lambda, \quad (\lambda = L'(0)\,\mathrm{sign}\,k_{n+1}),$$

or

$$k_{n+2}x_m^{n+2} - Q_1(x_m) - \lambda T_{n+1}(x_m) = 0.$$

Therefore, the polynomial

$$k_{n+2}x^{n+2} - Q_1(x) - \lambda T_{n+1}(x)$$

of degree $n+2$ has the same roots as $(x^2-1)T'_{n+1}(x)$, which means that one polynomial is a multiple of the other:

$$k_{n+2}x^{n+2} - Q_1(x) - \lambda T_{n+1}(x) = \mu(x^2-1)T'_{n+1}(x),$$

so

$$Q_1(x) = k_{n+2}x^{n+2} - \lambda T_{n+1}(x) - \mu(x^2-1)T'_{n+1}(x).$$

But the polynomial $Q_1(x)$ has degree n, so the coefficients of x^{n+1} and x^{n+2} have to be zero. This implies that $\lambda = 0$, $\mu = \frac{k_{n+2}}{(n+1)2^n}$, therefore

$$Q_1(x) = k_{n+2}\big\{x^{n+2} - \frac{1}{(n+1)2^n}(x^2-1)T_{n+1}(x)\big\}.$$

4. Chebyshev continues to carry out the same kind of computation, turning it into a recursive argument, and finding next $Q_2(x)$, $Q_3(x)$, etc. Given the functions $Q_m(x)$, one can determine $Q(x,h)$, $P(x,h)$, and finally $p(x,h)$ and $L(h)$.

Without getting into the details, let us illustrate the result of the example considered by Chebyshev for the case $n = 4$, $k_5 \neq 0$.

$$\begin{aligned} p(x,h) = &\big\{k_0 + \frac{1}{16}k_6h^6 + \frac{7k_5^2k_8 + 2k_5k_6k_7 - k_6^3}{64k_5^2}h^8 + \cdots\big\} \\ &+ \big\{k_1 - \frac{5}{16}k_5h^4 - \frac{31k_5k_7 - 3k_6^2}{64k_5}h^6 + \cdots\big\}(x-a) \\ &+ \big\{k_2 - \frac{13}{16}k_6h^4 - \frac{87k_5^2k_8 + 10k_5k_6k_7 - 5k_6^3}{64k_5^2}h^6 + \cdots\big\}(x-a)^2 \\ &+ \big\{k_3 + \frac{5}{4}k_5h^2 + \frac{22k_5k_7 - k_6^2}{16k_5}h^4 + \cdots\big\}(x-a)^3 \\ &+ \big\{k_4 + \frac{7}{4}k_6h^2 + \frac{36k_5^2k_8 + 2k_5k_6k_7 - k_6^3}{16k_5^2}h^4 + \cdots\big\}(x-a)^4. \end{aligned}$$

In this formula, the quantities that contain positive powers of h constitute "the changes that have to be made in the approximate quantity $f(x)$, given by its expansion into ascending powers of $(x-a)$, in order to achieve the smallest possible deviation between $x = a - h$ and $x = a + h$, for h being quite small".

3. The memoir "Sur les question des minima qui se rattachent à la représentation approximative des fonctions"

For several years, Chebyshev's ideas matured and, in 1857, instead of continuing the memoir [1], he presents a new memoir to the Academy of Science, which is composed in a purely theoretic manner, and which contains a complete exposition of the method of best approximation. It is entitled "Sur les question des minima qui se rattachent à la représentation approximative des fonctions" [4]. We find here: (1) a general theory leading to "Chebyshev's necessary conditions"; (2) an application of the theory to three basic problems ('cases'), which are further solved completely.

The question is generalized in the following way. A function $F(x; p_1, p_2, \ldots, p_n)$, which depends on a variable x and parameters $p_1, p_2, \ldots, p_n$, is given. The variable x belongs to some closed interval that one can, without loss of generality, take to be $[-1, +1]$. As far as the parameter values are concerned, we shall assume that they belong to an open domain (P). We shall assume that the function F is continuously differentiable with respect to both x and p_i *. One has to find conditions on the values of the parameters p_i that are necessary to make the quantity

$$\max_{|x|\leq 1} |F(x; p_1, p_2, \ldots, p_n)| \tag{6}$$

smaller than for any other values that are sufficiently close.

* This assumption is actually present in the original manuscript: "In order to simplify the investigation, we leave aside the case of F or its derivatives with respect to x and the parameters not being finite and continuous."

Let L denote the specified maximum, and let the "deviation points" be the points x at which F is equal to $-L$ or $+L$. Suppose that the number of deviation points is finite**, and let us denote them by $x_1, x_2, \ldots, x_\mu$.

Chebyshev claims that if a system of parameter values in hand provides the minimum, then one of the following two statements hold. Either the number of deviation points is at least one more than the number of parameters

$$\mu \geq n + 1$$

or the rank of the matrix

$$\begin{bmatrix} P_{11} & P_{12} & \ldots & P_{1\mu} \\ P_{21} & P_{22} & \ldots & P_{2\mu} \\ \vdots & \vdots & \ldots & \vdots \\ P_{n1} & P_{n2} & \ldots & P_{n\mu} \end{bmatrix}, \tag{7}$$

where

$$P_{ik} = \frac{\partial F}{\partial p_i}(x_k) \equiv \frac{\partial F}{\partial p_i}(x_k; p_1, p_2, \ldots, p_n),$$

is smaller than μ.

This theorem is proved by contradiction. Assume that $\mu \leq n$, so that our matrix has at most as many columns as rows and that its rank is equal to the number of columns μ. Then the linear system in the n variables $N_1, N_2, \ldots, N_n$

$$\sum_{i=1}^{n} P_{ik} N_i = F_k \quad (k = 1, 2, \ldots, \mu) \tag{8}$$

has a non-trivial (non-zero) solution for any values of F_k, as long as they are not all zero. Let the numbers $N_1, N_2, \ldots, N_n$ be a non-trivial solution to the linear system (8) for

$$F_k = F(x_k; p_1, p_2, \ldots, p_n) \quad (= \pm L \neq 0).$$

Then there exists a number ω such that the value of (6) at $p_1 - \omega N_1, p_2 - \omega N_2, \ldots, p_n - \omega N_n$ is smaller than its value at $p_1, p_2, \ldots, p_n$. Indeed, let

$$\Phi(x, \omega) \equiv F(x; p_1 - \omega N_1, \ldots, p_n - \omega N_n).$$

Then

$$\Phi(x, \omega) = F(x; p_1, \ldots, p_n) - \omega \sum_{i=1}^{n} N_i \frac{\partial F}{\partial p_i}(x; p_1 - \theta\omega N_1, \ldots, p_n - \theta\omega N_n), \quad (0 < \theta < 1)$$

and, therefore,

$$\Phi(x_k, \omega) = F(x_k; p_1, \ldots, p_n) - \omega \sum_{i=1}^{n} N_i \left[\frac{\partial F}{\partial p_i}(x_k; p_1, \ldots, p_n) + \varepsilon_i\right] =$$
$$= F_k - \omega(F_k + \sum_{i=1}^{n} N_i \varepsilon_i) = (1 - \omega) F_k - \omega \sum_{i=1}^{n} N_i \varepsilon_i,$$

where the ε_i tend to zero as ω tends to zero. If we take ω positive and sufficiently small, we shall have

$$|\Phi(x_k, \omega)| < |F_k|,$$

that is,

$$|F(x_k; p_1 - \omega N_1, \ldots, p_n - \omega N_n)| < |F(x_k; p_1, \ldots, p_n)| \quad (k = 1, 2, \ldots, \mu).$$

** Note that the number of deviation points may not even be countable.

From this, using continuity, it is easy to conclude that, for a sufficiently small positive ω, the following inequality also holds

$$\max_{|x|\le 1}|F(x;p_1-\omega N_1\dots,p_n-\omega N_n)|<\max_{|x|\le 1}|F(x;p_1\dots,p_n)|.$$

The theorem just proved is fundamental for Chebyshev. It opens the road to computing the parameters. If the number μ of deviation points exceeds the number n of parameters by 1 or more, then the relations

$$F^2(x_i)=L^2,\quad (x_i^2-1)F'(x_i)=0$$

hold at the deviation points x_i. The number of these relations is 2μ, and, therefore, if the problem is viewed from a standpoint that is characteristic of Chebyshev, the $\mu+n+1$ variables $x_1,\dots,x_\mu$, $p_1,\dots,p_n$, and L can be computed. If the number μ of deviation points does not exceed the number n of parameters, then the missing $n-\mu+1$ equations are obtained after the rank of the matrix (7) is decreased.*

The problem considered by Chebyshev in [1], to which he returns here ("the first case"), corresponds to the assumption

$$F(x;p_1,\dots,p_n)\equiv p_1x^{n-1}+p_2x^{n-2}+\dots+p_n-f(x)$$

(along with Chebyshev, we take the degree of the approximating polynomial to be $n-1$). With this assumption,

$$P_{ik}=x_k^{n-i},$$

and, since the x_i's are distinct, the matrix (7) has the form

$$\begin{bmatrix} x_1^{n-1} & x_2^{n-1} & \dots & x_\mu^{n-1}\\ \vdots & \vdots & \dots & \vdots\\ x_1 & x_2 & \dots & x_\mu\\ 1 & 1 & \dots & 1\end{bmatrix},$$

and its rank is necessarily equal to μ. Therefore, in this case the number of deviation points exceeds the number of parameters by at least 1. This fills the gap left in the "Theory of Mechanisms" [1].

Chebyshev's "second case" corresponds to the more general assumption

$$F(x;p_1,\dots,p_n)\equiv \varrho(x)\{p_1x^{n-1}+p_2x^{n-2}+\dots+p_n-f(x)\},$$

where $\varrho(x)$ is a given function that is positive on the basic interval (a "weight", in the terminology of more recent authors). It is easy to check that the rank of the matrix (7) cannot be decreased in this case. Note that Chebyshev only considers weights of the form

$$\varrho(x)=\frac{1}{A(x)},\tag{9}$$

where $A(x)$ is a polynomial that does not vanish on the basic interval.**

The "third case" concerns the approximation of a given function by a rational function with variable coefficients

$$F(x;p_1,\dots,p_n)\equiv\frac{p_1x^{n-l-1}+p_2x^{n-l-2}+\dots+p_{n-l}}{p_{n-l+1}x^l+p_{n-l+2}x^{l-1}+\dots+p_nx+1}-f(x).$$

In this case, as is shown by the computation offered by Chebyshev, the number of deviation points x_k does not necessarily exceed the number n of parameters, but this happens only when the approximating function

* The simplest examples that illustrate how the rank of the matrix can be decreased are (1) $F(x,p)=(x-p)^2-\frac{1+x^2}{2}$ for $p=0$; (2) $F(x,p)=1-x^2-px$ for $p=0$. The geometric meaning in these examples is obvious. In both cases there is only one deviation point.

** Recall the first problem of Poncelet, where the weight is $\varrho(x)=\frac{1}{\sqrt{x^2+1}}$.

is a fraction such that the first $n+1-\mu$ leading coefficients vanish both in the numerator and in the denominator.

While mentioning that, "with the help of traditional methods of algebra", the problems of the general type he considered "require computations that are absolutely impossible", Chebyshev then takes up some specialized examples, where he is able to reduce the problem to "questions of indeterminate analysis†".

First of all, a different solution (without using differential equations) is given to the problem of finding a polynomial $P(x)$ of a given degree and with a given leading coefficient, say equal to 1, that deviates least from zero. Namely, the ratio $\frac{P^2(x)-L^2}{x^2-1}$ is the exact square of a polynomial $Q(x)$ of degree $n-1$, so that

$$P^2(x)-(x^2-1)Q^2(x)=L^2. \tag{10}$$

This identity can be rewritten as

$$\frac{P(x)}{Q(x)}=\sqrt{x^2-1}+\frac{L^2}{Q(x)[P(x)+Q(x)\sqrt{x^2-1}]},$$

so it is clear that $\frac{P(x)}{Q(x)}$ is a fraction appropriate for decomposing $\sqrt{x^2-1}$ into a continued fraction

$$\sqrt{x^2-1}=x-\frac{1|}{|2x}-\frac{1|}{|2x}-\cdots.$$

This allows one to calculate both the polynomial $P(x)$, and the polynomial $Q(x)$, up to a constant multiple, and that can be easily found from an additional condition.

An argument of the same kind is used by Chebyshev for finding the polynomial $P(x)$ that deviates least from zero (with the same additional condition) for the case of an arbitrary weight of the form (9), where

$$A(x)=\prod_{\nu=1}^{m}(x-a_\nu).$$

The identity (10) now generalizes in the following way:

$$P^2(x)-(x^2-1)Q^2(x)=L^2A^2(x).$$

In the search for the most general form of the polynomials $P(x)$, $Q(x)$, and constant L, Chebyshev uses the following trick. He starts with a "particular" solution

$$P_0(x)=\frac{1}{2}\left\{\prod_{\nu=1}^{m}\left(\sqrt{\frac{x-1}{\alpha_\nu-1}}+\sqrt{\frac{x+1}{\alpha_\nu+1}}\right)^2+\prod_{\nu=1}^{m}\left(\sqrt{\frac{x-1}{\alpha_\nu-1}}+\sqrt{\frac{x+1}{\alpha_\nu+1}}\right)^2\right\},$$

$$Q_0(x)=\frac{1}{2\sqrt{x^2-1}}\left\{\prod_{\nu=1}^{m}\left(\sqrt{\frac{x-1}{\alpha_\nu-1}}+\sqrt{\frac{x+1}{\alpha_\nu+1}}\right)^2-\prod_{\nu=1}^{m}\left(\sqrt{\frac{x-1}{\alpha_\nu-1}}+\sqrt{\frac{x+1}{\alpha_\nu+1}}\right)^2\right\},$$

$$L_0=\prod_{\nu=1}^{m}\frac{2}{\alpha_\nu^2-1},$$

and shows that any solution $P(x)$, $Q(x)$, L must have the property that the polynomials $P_0(x)P(x)-(x^2-1)Q_0(x)Q(x)$ and $P_0(x)Q(x)-P(x)Q_0(x)$ are divisible by $A^2(x)$, while the respective ratios

$$X(x)=\frac{P_0(x)P(x)-(x^2-1)Q_0(x)Q(x)}{A^2(x)},\quad Y(x)=\frac{P_0(x)Q(x)-P(x)Q_0(x)}{A^2(x)}$$

† "questions d'Analyse indéterminée"

satisfy the identity

$$X^2(x) - (x^2-1)Y^2(x) = (L_0L)^2.$$

But all solutions of this last identity are known from the "first case". From this, one can immediately compute the polynomials $P(x)$ and $Q(x)$, and the constant multiple that enters the solution can be found from the additional condition.

The third problem considered by Chebyshev in the memoir under discussion is a generalization of the first one, as is the second one, but it is noticeably more complex. Among all rational fractions $\frac{U(x)}{V(x)}$ where the degree of both numerator and denominator are given (here the coefficients of both the numerator and the denominator are now arbitrary), one has to find a fraction that deviates least from a given polynomial $u(x)$ of a degree that is one more than the degree of the numerator of the fraction.

Based on the derived necessary conditions on the deviation points, Chebyshev concludes that the polynomials $U(x)$ and $V(x)$ necessarily satisfy an identity of the form

$$[u(x)V(x) - U(x)]^2 - L^2V^2(x) = (x^2-1)W^2(x),$$

where $W(x)$ is some polynomial. Further investigation requires decomposing one of the functions

$$\sqrt{\frac{[u(x)+L](x^2-1)}{u(x)-L}} \quad \text{or} \quad \sqrt{\frac{[u(x)+L](x+1)}{[u(x)-L](x-1)}}$$

into a continued fraction of the form $q_0 + \frac{1|}{|q_1} + \frac{1|}{|q_2} + \cdots$, where the q_i denote polynomials. As a result of a very subtle analysis, a precisely formulated rule is obtained that allows one to compute first the deviation L and then the fraction $\frac{U(x)}{V(x)}$ itself.

The memoir under discussion is naturally complemented, without introducing anything fundamentally new, by Chebyshev's later work "On functions that deviate little from zero for some values of the variables" (1881) [10]. The following problems are solved here:

(1) Among all algebraic polynomials of degree n that take on a given value M at a given point $x = H$ ($H > 1$), find the one that deviates least from zero on the interval $|x| \leq 1$.

(2) Among all trigonometric polynomials* $A_0 + \sum_{m=1}^{n}(A_m \cos mx + B_m \sin mx)$ of degree n that take on a given value M at the point $x = x_1$, find the one that deviates least from zero in the interval $|x| \leq x_0$ ($0 < x_0 < x_1 < 2\pi$).

The solutions to these problems are expressed in terms of polynomials, namely:

$$(1)\ M\frac{T_n(x)}{T_n(H)}, \quad (2)\ M\frac{T_{2n}\left(\frac{\sin\frac{x}{2}}{\sin\frac{x_0}{2}}\right)}{T_{2n}\left(\frac{\sin\frac{x_1}{2}}{\sin\frac{x_0}{2}}\right)}.$$

4. Best approximation in a normed linear space

Let us return to Chebyshev's main problem but, in order to get closer to the essence of the questions it raises, we shall give it a more transparent, a more abstract form.

We say that a family of elements form a *metric space* R if, for each pair of elements a and b from R, there is a real number $\delta(a,b)$, called the *distance* between a and b, that has the following properties:

1° $\delta(a,b) \geq 0$; $\delta(a,b) = 0$ if and only if $a = b$, that is, a and b coincide;

2° $\delta(a,b) = \delta(b,a)$;

3° $\delta(a,b) \leq \delta(a,c) + \delta(c,b)$ ('triangle inequality').

* It deserves special attention that the trigonometric problem, albeit in later work, is nevertheless presented by Chebyshev.

A subset E of a space R is called *bounded* if, for each a in R, the set of numbers $\delta(a,x)$, where x ranges over E, is bounded. It follows from the triangle inequality that the boundedness of a set E is equivalent to the existence of at least one element a_0 so that the set of numbers $\delta(a_0,x)$, where $x \in E$, is bounded.

A sequence of elements $\{x_n\}$ *has limit* a if $\lim_{n\to\infty}\delta(a,x_n)=0$. Any sequence either does not have a limit or has only one limit.

The distance $\delta(x,y)$ is a continuous function of both variables x and y, as follows from the triangle inequality.

A subset E is said to be *compact*† if any bounded infinite subset of E contains a sequence that converges, that is, has a limit.* A subset E is said to be *closed* if it contains the limits of all sequences in E that converge.

Let E be a subset of the space R, and let a be an element of R. The lower bound L of distances $\delta(a,x)$, where x ranges over E, is called the distance from the element a to the set E:

$$L \equiv L(a,E) = \inf_{x\in E}\delta(a,x).$$

If an element x_0 of a set E has the property that its distance from a is the same as the distance from E to a,

$$\delta(x_0,a) = L,$$

then it is said that x *provides a best approximation to* a *from the set* E. The quantity L itself is called *the magnitude of the best approximation* or simply the *distance*.

The questions arise:

(1) Does there exist an element x_0 in E that provides a best approximation to a?

(2) Is this element x_0 unique?

Let us consider the first question first.

By the definition of the infimum, there is a sequence $\{x_n\}$ in E, such that

$$\lim_{n\to\infty}\delta(a,x_n) = L.$$

Obviously the set that consists of the elements $\{x_n\}$ is bounded. If the set E is compact, then one can choose a convergent subsequence $\{x_{p_n}\}$. Let $x_{p_n}\to x_0$. If, in addition, the set E is closed, then the element x_0 also belongs to the space E. Then, since the distance is continuous, $\delta(x_0,a)=L$, that is, x_0 provides a best approximation.

So, *if a set* E *in a metric space* R *is compact*† *and closed, then, for any* a *in* R*, the set* E *contains a best approximation to* a.

The case of special interest is that of R being a *linear* space. This means that the following operations are defined: (1) addition, (2) multiplication by a scalar (real number). These operations satisfy all the usual algebra laws. The sum of elements a and b is denoted by $a+b$. The product of an element a by a scalar λ is denoted by λa. The zero element of the space is denoted by 0.

A collection of elements $a_1,a_2,\ldots,a_n$ of a linear space is called *linearly independent* if the equation

$$\sum_1^n \lambda_i a_i = 0$$

implies that $\lambda_1=\lambda_2=\cdots=\lambda_n=0$. A space is *infinite dimensional* if there exist linear independent collections with any number of elements. In the opposite case the space is *finite dimensional* and the maximum number of linear independent elements is called the dimension of the space.

Let the dimension of a space be p, and let

$$e_1,e_2,\ldots,e_p$$

---

† *locally precompact*, in present-day parlance.

* Note that we require that a convergent sequence exists in any *bounded* infinite set, as opposed to the usual definition of a compact set.

be linearly independent elements in the space. Then, for any x in the space, the collection $x, e_1, e_2, \ldots, e_p$ is not linear independent, and, therefore, there is a dependence of the form

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_p e_p,$$

where $\xi_1, \xi_2, \ldots, \xi_p$ are some uniquely defined numbers. Therefore, an element of a p-dimensional linear space is defined by the values of p scalar parameters (coordinates), and this dependence is linear.

A linear space is said to be *normed* if it has a metric defined in terms of a *norm*. For each element a, there is a number $\|a\|$ (the norm of the element a) with the properties:

1° $\|a\| \geq 0$, $\|a\| = 0$ if and only if $a = 0$,

2° $\|\lambda a\| = |\lambda| \|a\|$,

3° $\|a + b\| \leq \|a\| + \|b\|$.

A space is called *strictly normed* if the equation

$$\|a + b\| = \|a\| + \|b\|$$

implies that $\lambda a = \mu b$ for some non-negative λ and μ $(\lambda^2 + \mu^2 > 0)$.

In a normed linear space a metric is defined as follows:

$$\delta(a, b) = \|a - b\|.$$

All the required properties of a metric are satisfied, since they follow from the properties of the norm.

In a finite-dimensional linear space, as we have seen, an element x is defined by finitely many parameters ξ_i. The norm of the element x is a continuous function of the parameters. Indeed, let $\xi_i^{(n)} \to \xi_i$ $(i = 1, 2, \ldots, p)$. Then, by setting $x^{(n)} = \sum_1^p \xi_i^{(n)} e_i$, $x = \sum_1^p \xi_i e_i$, we obtain:

$$\|x^{(n)} - x\| = \| \sum_1^p (\xi_i^{(n)} - \xi_i) e_i \| \leq \sum_1^p |\xi_i^{(n)} - \xi_i| \|e_i\| \to 0,$$

and, therefore,

$$x^{(n)} \to x, \quad \|x^{(n)}\| \to \|x\|.$$

This implies that *a finite dimensional normed linear space is necessarily compact and closed*. To prove this, look at a closed set of points in the p-dimensional space of parameters ξ_i, for which

$$\max\{ |\xi_1|, |\xi_2|, \ldots, |\xi_p| \} = 1.$$

On this set, denoted by K, the function $\|x\|$ that is continuous in the parameters ξ_i, attains its smallest value δ. This value δ is positive since $\|x\|$ becomes zero only if $x = 0$. Thus, if the largest of the absolute values of the parameters is equal to one, then the norm of the element is greater than or equal to δ. Let us now take a bounded sequence $\{ x^{(n)} \}$ of elements. Assume that $\|x^{(n)}\| < M$. Let $x^{(n)} = \sum_1^p \xi_i^{(n)} e_i$. Consider a sequence $\{ y^{(n)} \}$, where

$$y^{(n)} = \frac{x^{(n)}}{\sigma_n}, \quad \sigma_n = \max\left\{ |\xi_1^{(n)}|, |\xi_2^{(n)}|, \ldots, |\xi_p^{(n)}| \right\}.$$

Then the elements $y^{(n)}$ belong to the set K, and, as proved earlier, $\|y^{(n)}\| \geq \delta$, that is, $\frac{\|x^{(n)}\|}{\sigma_n} \geq \delta$, which implies that $\sigma_n \leq \frac{\|x^{(n)}\|}{\delta} < \frac{M}{\delta}$.

But this means that $|\xi_i^{(n)}| < \frac{M}{\delta}$ for all values of i and n, and, therefore, one can choose a subsequence $\{ m_n \}$, so that $\xi_i^{(m_n)} \to \xi_i$ $(i = 1, 2, \ldots, p)$, and then $x^{(m_n)} \to x$, where $x = \sum_1^p \xi_i e_i$. This proves compactness. The closedness is obtained even more easily.* Indeed, let the elements of the sequence $\{ x^{(m)} \}$

* The proof is necessary, of course, only in the case when the finite dimensional space in question is a proper subset of the linear space R.

that converges to x belong to a p-dimensional normed linear space E. Assume that e_i $(i = 1, 2, \dots, e_p)$ is a system of linear independent elements from E, and that $x^{(n)} = \sum_1^p \xi_i^{(n)} e_i$. Choose a subsequence $\{ m_n \}$ so that $\xi_i^{(m_n)} \to \xi_i$ $(i = 1, 2, \dots, p)$. Then $x^{(m_n)} \to \sum_1^p \xi_i e_i$, and, on the other hand, $x^{(m_n)} \to x$, therefore, $x = \sum_1^p \xi_i e_i$, that is, x belongs to E.

Collecting the results obtained, one can make the following statement. If the set E in a normed linear space R is also linear and finite dimensional, then, for each element a in R, there is an element x_0 in E that provides a best approximation to a.

Passing to the next question – about the *uniqueness* of best approximation, – let us note here the following *sufficient* condition, under which uniqueness necessarily takes place. This condition consists of the requirement that *the space R be strictly normed.*

Assume to the contrary that the space E has two different elements x_0' and x_0'' that are best approximations to a, so that†

$$\|x_0' - a\| = L, \quad \|x_0'' - a\| = L, \quad x_0' \neq x_0'', \quad L > 0.$$

Then,

$$\|\frac{x_0' + x_0''}{2} - a\| = \|\frac{x_0' - a}{2} + \frac{x_0'' - a}{2}\| \leq \|\frac{x_0' - a}{2}\| + \|\frac{x_0'' - a}{2}\| = L.$$

Here the inequality is, in fact, strict, because otherwise, since the space is strictly normed, it would follow that

$$\lambda'(x_0' - a) = \lambda''(x_0'' - a) \quad (\lambda', \lambda'' \geq 0, \quad \lambda'^2 + \lambda''^2 \neq 0)$$

The equality $\lambda' = \lambda''$ is impossible because $x_0' \neq x_0''$. If $\lambda' \neq \lambda''$, then

$$a = \frac{\lambda' x_0' - \lambda'' x_0''}{\lambda' - \lambda''}, \quad \|\frac{\lambda' x_0' - \lambda'' x_0''}{\lambda' - \lambda''} - a\| = 0,$$

which is also impossible. Therefore, a contradiction is obtained:

$$\|\frac{x_0' + x_0''}{2} - a\| < L.$$

The results presented can be given a somewhat more general form.

It is obvious that a subset E of elements of a linear space is a linear space itself if the element $\lambda a + \mu b$ belongs to E whenever the elements a and b belong to E, for all scalars λ and μ.

Let us call a set M of elements of a linear space R an *affine set* if the element $\lambda a + \mu b$ belongs to M, whenever the elements a and b belong to M, and the scalars λ and μ satisfy the relation $\lambda + \mu = 1$.

Each linear space is an affine set. The converse statement holds if and only if the affine set in question contains the zero element.

If x ranges over the elements of a linear space E, and a is an arbitrary element of R, then $y = x + a$ ranges over the elements of some affine set M. Conversely, if y ranges over the elements of an affine set M, and a is one of its elements, then $x = y - a$ ranges over the elements of some linear space.

Let us now consider the following generalized problem: *find an element y_0 of a given affine set M that provides a best approximation to the zero element*. In other words: *find an element y_0 of an affine set M that "deviates least from zero"*.

The previous remark reduces this generalized problem to the one considered earlier. The results pertaining to the existence and uniqueness of the approximation are thus generalized automatically.

Everything discussed above can be given a "geometric" interpretation, by considering the so-called *gauge body* of Minkowski.

The *gauge body* (Eichkörper) or *unit ball* K is the set of elements x of a linear space R with the norm less then or equal to one:

$$\|x\| \leq 1.$$

Let us note the following properties of the gauge body:

† The case $L = 0$ is trivial.

(1) if an element a belongs to K, then any element λa, where $|\lambda| \leq 1$ also belongs to K;

(2) K is a *convex* body. If a and b belong to K, then, given that $\lambda, \mu \geq 1$, $\lambda + \mu = 1$, the element $\lambda a + \mu b$ also belongs to K;

(3) in particular, if the space R is strictly normed, then the property (2) is strengthened in the sense that the element $\lambda a + \mu b$ not only belongs to K but it also is in the interior of K. That is, all elements that are sufficiently close to $\lambda a + \mu b$ also belong to K.

Along with the gauge body $K \equiv K_1$, let us consider the "similar" bodies K_λ $0 < \lambda < \infty)$ defined by the inequalities

$$\|x\| \leq \lambda.$$

It is clear that the bodies K_λ also have the properties (1)–(3).

Let us call an affine set M *supporting* for a body K_λ if the minimum of the norms of elements of E is equal to λ. This means, first, that there is at least one element that is common to the set E and the body K_λ. And, second, that for arbitrarily small $\varepsilon > 0$, the set E and the body $K_{\lambda-\varepsilon}$ do not have any common elements.

Property (2) implies that the set of points that are common to a affine set and a body K_λ is also a convex set. Property (3) implies that, in a strictly normed linear space, a supporting affine set has only one common element with the corresponding body K_λ. If the space is not strictly normed, it then depends on the choice of the affine set M as to whether there will be only one common element or an infinite set (necessarily convex) of them.

Let us now turn to the specific realizations of the general schemes presented earlier.

These specific realizations take one form or another, depending on the nature of the elements that make up the space R, and what is being understood by a "sum", a "multiplication by a scalar", and a "norm". In the applications of interest here, the role of elements is played by functions of one or several variables, real or complex, defined on a fixed domain (D). The "sum" of the elements is the usual sum of functions. The "multiplication by a scalar" is the multiplication of a function by a constant. As far as, finally, the "norm" is concerned, it can be defined in different ways, and the resulting "function space" depends on the definition of the norm. Of course, the set of elements of the space depends on the choice of the norm.

On the other hand, the role of elements can also be played by points, say, in n-dimensional Euclidean space (the n-tuples of numbers that are their coordinates). Then the "addition" is "geometric" or "vector" addition, in which the n-tuples are added componentwise. The "multiplication by a scalar" is multiplication of all coordinates by that scalar. The same can be said about choosing the norm as in the case of function spaces.

The name given to such a space depends on the nature of its elements and the choice of the norm. The following table† lists the spaces that will be referred to in the sequel. To be definite, we shall restrict the choice of the functional space to the case of one independent real variable, and we shall even assume that the domain (D) mentioned is the fixed interval $(-1, +1)$. Also, $\rho > 0$.

Notation	Elements	Norm
$L^{(s)}(\varrho), \quad (s \geq 1)$	$\left\{ f : \int_{-1}^{+1} \varrho(x)\|f(x)\|^s \,\mathrm{d}x < \infty \right\}$	$\left\{ \int_{-1}^{+1} \varrho(x)\|f(x)\|^s \,\mathrm{d}x \right\}^{\frac{1}{s}}$
$l_n^{(s)}(\varrho), \quad (s \geq 1)$	$\{ (x_1, x_2, \ldots, x_n) \}$	$\{ \sum_{i=1}^n \varrho_i \|x_i\|^s \}^{\frac{1}{s}}$
$L^{(s)}, \quad (s \geq 1)$	$\left\{ f : \int_{-1}^{+1} \|f(x)\|^s \,\mathrm{d}x < \infty \right\}$	$\left\{ \int_{-1}^{+1} \|f(x)\|^s \,\mathrm{d}x \right\}^{\frac{1}{s}}$
$l_n^{(s)}, \quad (s \geq 1)$	$\{ (x_1, x_2, \ldots, x_n) \}$	$\{ \sum_{i=1}^n \|x_i\|^s \}^{\frac{1}{s}}$
$C(\varrho)$	Continuous functions f.	$\max_{\|x\| \leq 1} \varrho(x)\|f(x)\|$
$c(\varrho)$	$\{ (x_1, x_2, \ldots, x_n) \}$	$\max_{i \in \{1,n\}} \varrho_i \|x_i\|$
C	Continuous functions f.	$\max_{\|x\| \leq 1} \|f(x)\|$
c	$\{ (x_1, x_2, \ldots, x_n) \}$	$\max_{i \in \{1,n\}} \|x_i\|$

† rearranged to fit the printed page here

For brevity, the spaces that are denoted by letters C or c will be called the spaces with Chebyshev's norm, or *Chebyshev spaces*. The spaces that are denoted by the letters L or l will be called the spaces with the power-norm, or *power-spaces*.

In the case of the power-spaces $L^{(s)}(\varrho)$, $L^{(s)}$, the functions that differ only on a set of measure zero are identified. Thus, the elements in this case are classes of functions.

The triangle inequality is obvious for the Chebyshev spaces. For the $L^{(s)}$ spaces with $s \geq 1$, it follows from the so-called Minkowski inequality. The case $s = 2$ is especially important for two reasons. First, the metrics in the spaces $L^{(2)}$ and $l^{(2)}$ are completely analogous to that in Euclidean spaces. And, second, the conditions of the extrema in this case are linear. If $s < 1$, then the power-spaces become defective in the sense that the triangle inequality fails. Despite this fact, in the case of finite dimensional affine sets, a best approximation nevertheless exists.*

The $L^{(s)}$ spaces with $s > 1$ are strictly normed**, and this, as we have seen, implies uniqueness of a best approximation. This kind of statement, however, would be incorrect, for both the $L^{(s)}$ spaces with $s = 1$ and the Chebyshev spaces. Under these circumstances, it is easy to visualize the situation, given that in the case of the point spaces $l_n^{(s)}(\varrho)$ with $s > 1$, the gauge body is smooth, and the supporting planes are the tangent planes. For $s = 1$ and in the case of the Chebyshev space $c_n(\varrho)$, the gauge body is a convex polyhedron, so that the supporting plane can coincide with the "faces" or the "edges". Finally, for the $L^{(s)}$ spaces with $s < 1$, even the convexity is lost.

The case of a Chebyshev space is especially interesting to us. It is, in some sense, the limit case, since the norm in the space $L^{(s)}$ turns into that in the space C as $s \to \infty$. Whether the problem of approximating an element a from a linear space E has a unique solution depends, generally speaking, on both the element a and the space E. However, Haar [47] found a property that characterizes the finite dimensional linear spaces E in which uniqueness holds for an arbitrary element a. This property consists of the requirement that every element of the p-dimensional space L be a function with no more than $p - 1$ zeros in the basic interval. The systems of functions that generate this kind of spaces were called *Chebyshev systems* by S.N. Bernstein [27]. The simplest examples are exactly the spaces of polynomials of a given degree with arbitrary weight considered by Chebyshev.

In the case of functions of several variables, as in the case of functions of a complex variable, the statement about uniqueness of a best approximation by a polynomial of a given degree does not hold. The first counterexamples were found by Tonelli [72].

5. Necessary and sufficient conditions of approximation in Chebyshev problems

In his major memoirs [1] and [4], Chebyshev only deals with the spaces that we denoted C or $C(\varrho)$ and he is only interested in approximating a given element – a continuous function – from a finite dimensional space of polynomials of a given degree n. While starting with the "known" fact about the deviation taking on its maximal value at least $n+2$ times, and developing a series of consequences that follow from it, he comes with necessity to the formula for the sought-for polynomial. The subtle argument by Chebyshev, based on the theory of continued fractions, was reproduced by Bertrand in his "Calcul différentiel" in 1864, and this is how Chebyshev polynomials became widely known. However, it took almost half a century before it was noticed that the original argument by Chebyshev admits significant reduction. Indeed, one can formulate rather simple necessary and sufficient conditions for a polynomial that provides a best approximation, and then it only remains to check that the polynomials found by Chebyshev satisfy these conditions. The fact that the "lengthy considerations dealing with the theory of continued fractions" can be avoided was apparently first noticed in 1901 by Blichfeldt, who pointed out [40] that the graph of the deviation on a given interval is characterized by the existence "of at least $n+2$ alternations of two kinds of maxima", while mentioning that

* The proof only uses the consequences of the triangle inequality: (1) boundedness of the set; and (2) continuity of the norm. Both follow from the "generalized triangle inequality": there should be a function $\phi(u)$ defined on the interval $0 \leq u < \infty$, positive, increasing, and continuous, with the properties $\phi(0) = 0$, $\lim_{u\to\infty} \phi(u) = \infty$ and $\phi(||a + b||) \leq \phi(||a||) + \phi(||b||)$. In the case of the $L^{(s)}$ spaces with $0 < s < 1$, one can take $\phi(u) = u^s$.

** See, for example, [48], page 148, Theorem 200.

he does not know of any original works by Chebyshev where the above property could be found. Indeed, the Collected Works of Chebyshev do not contain anything pertaining to the sign changes of the deviation, which does not imply, of course, that the fact of alternation itself was not known to him.

In 1902 in Göttingen, there appeared a dissertation by Kirchberger [53] where the problem of signs was given full consideration, even in the case of functions of many variables. In the case of one variable, the modern formulation of the approximation conditions was given by E. Borel in his monograph "Leçons sur les fonctions de variables réelles et les développements en séries de polynômes" (1905) [41]. A polynomial $P(x)$ of degree n is a best approximation to the function $f(x)$ on some interval if and only if the given interval contains at least $n+2$ points x_i such that $x_1 < x_2 < \cdots < x_{n+2}$ and $R(x_i) = \varepsilon(-1)^i L$ ($\varepsilon = +1$ or $\varepsilon = -1$), where $R(x) = f(x) - P(x)$. The sufficiency of this condition follows from the following argument. If the deviation $R(x) = f(x) - P(x)$ satisfies the condition, and $Q(x)$ is another polynomial of the same degree, such that $|f(x) - Q(x)| < L$ on the given interval, then the polynomial $P(x) - Q(x)$ of degree n has the same signs at the points x_i as does $R(x)$ and is therefore equal to zero at least $n+1$ times, which is impossible. The minimal property of the polynomial $T_n = \cos n \arccos x$ $(-1 \le x \le 1)$ follows immediately from this, because

$$T_n\big(\cos\frac{n-i+1}{n}\pi\big) = \varepsilon(-1)^i L, \quad (i = 1, 2, \ldots, n+1),$$

where

$$\varepsilon = (-1)^{n+1}, \quad L = 1.$$

These conditions can be generalized to the case of many variables, however they become cumbersome and difficult to use.

Let a function f be approximated in some domain (D) by generalized polynomials of the form $\sum_{i=1}^n c_i\phi_i$, where ϕ_i are functions that are continuous in (D), and c_i are arbitrary parameters (the variable arguments are omitted for brevity). Then a polynomial P of the specified form deviates least from f in (D) if and only if there does not exist a polynomial Q of the same form that takes on positive values at the points Δ_+ where $f - P = +L$ and negative values at points Δ_- where $f - P = -L$. Indeed, if the polynomial P deviates least from f and the polynomial Q had the specified property, then, for sufficiently small positive values of λ, the polynomial $P - \lambda Q$ would deviate less from f than does P. On the other hand, if there were some polynomial Q deviating from f in (D) less than P, then the polynomial $P - Q$ would be positive at the points Δ_+ and negative at the points Δ_-. In the particular case of only one independent variable and the "polynomials" $\sum_{i=1}^n c_i\phi_i$ obtained from a Chebyshev system of functions ϕ_i, the condition formulated above is equivalent to the existence of at least $n+1$ points of deviation with the alternating signs (Chebyshev-Borel condition). In another particular case of an arbitrary number of independent variables and linear approximating polynomials, the condition reduces to non-existence of an $(n-1)$-dimensional plane separating the set Δ_+ from the set Δ_- (Kirchberger condition).

6. Least-squares approximation and the memoir "On functions that deviate least from zero"

Chebyshev's works contain the explicit and generally formulated problem of best approximation to a given function $f(x)$ by a polynomial $P_n(x)$ of degree n in the space C, that is, with the norm equal to the maximum of the modulus. However, a similar problem in the space $L^{(2)}$, that is, with the norm equal to the square root of the integral of the square of the difference, was already solved long before then. Indeed, the polynomial $P_n(x)$ that minimizes the integral

$$\int_{-1}^{+1} [f(x) - P_n(x)]^2\,\mathrm{d}x$$

is nothing but the sum of the first n terms of the Legendre expansion of the function $f(x)$. Chebyshev came very close to the analogous problem in the space $L^{(2)}(\varrho)$, that is, with an arbitrary weight. However, instead of an integral, he introduces a sum, distributed over the sequences of points in the given interval. In this way, he finds in his 1855 memoir "On continued fractions" [2] the polynomial $P_n(x)$ that minimizes the sum

$$\sum_{i=1}^{m} \varrho(x_i)[f(x_i) - P_n(x_i)]^2 \quad (-1 \le x_1 < x_2 < \cdots < x_m \le 1).$$

It is interesting that the choice $s = 2$ is motivated by considerations from probability theory, namely, the effect of the error in the interpolation data on the sought-for quantity is being minimized. It already follows from Gauss' research that, under the assumption of the normal distribution of the error, the least-squares method should be used for data fitting.

Chebyshev's results can be extended, without significant changes, to the case of the integral

$$\int_{-1}^{+1} \varrho(x)[f(x) - P_n(x)]^2 \, \mathrm{d}x.*$$

The polynomial $P_n(x)$ is the sum of n terms in the expansion of $f(x)$ into a series of orthogonal polynomials of increasing degrees $\Phi_n(x)$, $(n = 0, 1, 2, \ldots)$

$$f(x) \sim \sum_{\nu=0}^{\infty} c_\nu \Phi_\nu(x), \quad c_\nu = \int_{-1}^{+1} \varrho(x) f(x) \Phi_\nu(x) \, \mathrm{d}x,$$

where the polynomial system $\{ \Phi_\nu(x) \}$ is defined by the weight $\varrho(x)$ according to the conditions

$$\int_{-1}^{+1} \varrho(x) \Phi_i(x) \Phi_k(x) \, \mathrm{d}x = \begin{cases} 0, & i \neq k, \\ 1, & i = k. \end{cases}$$

In particular, notice the following fact. Among all polynomials of degree n with the leading coefficient equal to one, $P_n(x) = x^n + \cdots$, the integral

$$\int_{-1}^{+1} \varrho(x) P_n^2(x) \, \mathrm{d}x$$

is minimized by a scalar multiple of $\Phi_n(x)$. To verify that, it is enough to write $P_n(x)$ as a linear combination of the orthogonal polynomials $\Phi_\nu(x)$.

If $\varrho(x) = \text{const}$, then the polynomials $\Phi_\nu(x)$ turn, up to a scalar multiple, into Legendre polynomials, defined by the known expansion

$$\frac{1}{\sqrt{1 - 2sx + s^2}} = \sum_{\nu=0}^{\infty} X_\nu(x) s^\nu.$$

A more general case $\varrho(x) = \frac{\text{const}}{(1+x)^\lambda (1-x)^\mu}$ was considered by Jacobi in 1859. The respective polynomials $\Phi_\nu(x)$ are scalar multiples of the Jacobi polynomials $J^{(\lambda,\mu)}(x)$ defined by the expansion

$$\frac{(1 + s + \sqrt{1 - 2sx + s^2})^\lambda (1 - s + \sqrt{1 - 2sx + s^2})^\mu}{2^{\lambda+\mu} \sqrt{1 - 2sx + x^2}} = \sum_{\nu=0}^{\infty} J_\nu^{(\lambda,\mu)}(x) s^\nu.$$

The orthogonality property of Jacobi polynomials was, among other things, derived by Chebyshev directly from this expansion in his 1869 note "On functions that are similar to the Legendre functions" [8].

Both theories, the one dealing with uniform or Chebyshev approximation, and the other dealing with least-squares approximation, are shown to be related in Chebyshev's 1872 memoir "On functions that deviate least from zero" [9]. Here, as previously, one is looking for a polynomial with the leading coefficient equal to one, $P(x) = x^n + \cdots$, that minimizes the maximum of the modulus in the interval $(-1, +1)$, with the additional restriction that the polynomial $P(x)$ has to be monotone. That is, $P(x)$ is either non-increasing or non-decreasing. Its largest and smallest values are therefore equal in absolute value, but opposite in sign, so that the maximum of its absolute value is equal to the magnitude of the integral

$$L = \frac{1}{2} \int_{-1}^{+1} P'(x) \, \mathrm{d}x.$$

* See Comp., t. II, 1907, page 200.

While mentioning that all zeros of the derivative $P'(x)$ that lie within the interval have to be of even order, and that there should be no zeros outside of the interval, Chebyshev comes to the conclusion that the polynomial $P'(x)$ of degree $n-1$ has the form

$$P'(x) = n(x-1)^{\varrho}(x+1)^{\varrho_0}U^2(x),$$

where the numbers ϱ and ϱ_0 can have values 0 or 1. Therefore, one obtains

$$L = \frac{n}{2}\int_{-1}^{+1}(x-1)^{\varrho}(x+1)^{\varrho_0}U^2(x)\,\mathrm{d}x.$$

It is now easy to minimize this integral by considering the multiple $(x-1)^{\varrho}(x+1)^{\varrho_0}$ as a weight. One only needs to distinguish between the four possible cases of values of ϱ and ϱ_0, depending on the parity of n, and on whether the polynomial $P(x)$ is increasing or decreasing.

The polynomial $U(x)$ is a scalar multiple of the Jacobi polynomial $J^{(-\varrho_0,\varrho)}_{\frac{n-1-\varrho-\varrho_0}{2}}(x)$, which Chebyshev finds from the generating function obtained earlier. It is not necessary to present the details and the exact result of the computations. Let us only note that the additional requirement of monotonicity does not affect the deviation much. Without the restriction it is exactly equal to $\frac{1}{2^{n-1}}$, while with the restriction it is asymptotically equal, as $n \to \infty$, to $\frac{\pi n}{2^n}$, that is, it only increases by a factor of $\frac{\pi}{2}n$. This comparison, which is natural in present research, did not escape Chebyshev. In his memoir, he gives, by the way, not an asymptotic estimate of the result, but an exact estimate with an inequality. In any case, these circumstances testify to the fact that, despite the exact content of his work and his reputation, Chebyshev was not indifferent to asymptotic questions.

7. Applications of the theory of best approximation by Chebyshev. Chebyshev's latest works

Chebyshev's statements about the relationships between a mathematical theory and its applications illustrate quite clearly not just the source of his creativity, but also his scientific and philosophical positions. He says "The convergence of theory and practice gives the most fruitful results, and it is not just the practice that gains from it. The sciences themselves are being developed under its influence. It opens new research subjects, or new aspects in subjects that are already known for a long time... If a theory gains much from a new application of an old method, or from its new developments, then it receives even more by discovering new methods, and in this case science finds a reliable advisor in practice." Without any doubt, when Chebyshev wrote these words, he was thinking mainly about the theory of best approximation created by him. In addition, even though Chebyshev's works of the early period belonged, judging by their topics and within the tradition of his great predecessors, to the abstract areas of science, and even though he later made a sharp turn, as we have mentioned, towards practical applications, in the latest period, both tendencies went hand in hand, and they are united in some harmonic balance. Furthermore, it is necessary to note that applications are understood by Chebyshev in a broad and original sense. They are not limited to the area of technical sciences, but are rather related to very different forms of human activity, or serve the internal needs of mathematics itself (designing tables, interpolation, quadratures, solving equations). They are being evaluated critically from the viewpoint of the relationship between the "means" used and the "goals" achieved.

It is not quite true that Chebyshev's works reflect completely the family of questions to which he had a chance to apply the methods of approximation that he developed and which he used with incomparable proficiency. Rather, his works contain just a portion of such applications. Having made this remark, let us now consider the applications that Chebyshev mentions *explicitly*.

i. Kinematics of mechanisms. As was already mentioned, this is the point of origin for the theory of approximation of functions by polynomials, or, generally speaking, by functions of various kinds that depend on several parameters. This is his favorite area and it attracted his thought for several decades. It is not the purpose of this article, however, to give consideration to the numerous notes that are related to this subject.

ii. Solving algebraic equations (separating roots). In the memoir [4], there are about ten theorems (6-11, 15-19) that are derived from the basic propositions on best approximation. These theorems state that, under certain conditions, the polynomial of interest has at least one zero in some interval. The length of the interval depends, on the one hand, on the value of the polynomial at the center of the interval, on the other hand, on specific assumptions on the coefficients or on the zeros of the polynomial. For example, Theorem 10 states that if the polynomial $f(x) = x^{2n+1} + \cdots + K$ does not contain any even powers of x, then it has at least one zero in the interval $|x| < 2\left(\frac{|K|}{2}\right)^{\frac{1}{2n+1}}$. Let us give a proof, assuming, of course, that $K \neq 0$. If the polynomial $f(x)$ did not have any zeros in the specified interval, then the same would be true for the polynomials $f(x) - 2K$ and $[f(x) - K]^2 - K^2 \equiv f(x)[f(x) - 2K]$. Since the latter polynomial is negative at $x = 0$, it would also be negative in the whole interval $|x| < 2\left(\frac{|K|}{2}\right)^{\frac{1}{2n+1}}$, so that $f(x) - K$ would deviate from zero by less than K. But this is impossible since it means that the polynomial $\frac{1}{2^{2n}|K|} f\left(2x\left(\frac{|K|}{2}\right)^{\frac{1}{2n+1}}\right) \equiv x^{2n+1} + \cdots$ would deviate from zero by less than $\frac{1}{2^{2n}}$ on the interval $|x| < 1$. Let us also formulate Theorem 9, which uses a different assumption: if a polynomial of degree n with leading coefficient equal to one, $f(x) \equiv x^n + \cdots$, has only real roots, then, for any t, there is a real root in the interval $|x - t| < 4\sqrt[n]{\frac{|f(t)|}{4}}$. Later, in his 1872 memoir, using a result on monotone polynomials, Chebyshev narrows the interval, replacing it with the following: $|x - t| < 4\sqrt[n]{\frac{|f(t)|}{2(n-1)\pi}}$.*

iii. Interpolation (remainder estimate). To minimize the error in the Lagrange interpolation formula, Chebyshev suggests to take the nodes of the interpolation [say, in the interval $(-1, +1)$] to be the zeros of the polynomial $T_n(x) \equiv \cos n \arccos(x)$, since, for a given function $f(x)$, the remainder has the form $R(x) = f^{(n+1)}(\xi) P_n(x)$, where $P_n(x) = \prod_{i=1}^{n}(x - x_i)$, and x_i are the nodes. Therefore, with $R_n(x) \leq M_{n+1} \max |P_n(x)|$, where $M_{n+1} = \max |f^{(n+1)}(x)|$, the choice $P_n(x) = T_n(x)$ is the most profitable. Here Chebyshev partly envisions the later result of Runge, that says that, as $n \to \infty$, Chebyshev interpolation converges for any function that is regular in the basic interval (while this is not true for Newton interpolation with equally spaced nodes).

iv. A rule for finding approximately distances on the surface of the Earth.** Let us quote it completely: "(1) take the differences between the two latitudes and the two longitudes and express them in minutes; (2) double the difference of the latitudes; (3) out of these two numbers, the difference of longitudes, and the doubled difference of latitudes, multiply the smaller by three, multiply the larger one by 7, and then add the two products; (4) the result divided by three will give the desired distance in versts."

It is not difficult to guess that here one is talking about applying the approximating formula of Poncelet $\sqrt{a^2 + b^2} \sim \alpha a + \beta b$ to the infinitesimal formula

$$\Delta s \sim R\sqrt{\Delta u^2 + \cos^2 u \Delta v^2},$$

where Δs is the length of the main arc connecting the two points, R is the radius of the Earth, which is equal to 5971 versts, u is latitude, and Δu and Δv are the differences of latitudes and longitudes, in radians. The Poncelet parameter k is obviously taken to be 1. The rule was probably obtained in the following way:

$$\begin{aligned} R\sqrt{\Delta u^2 + \cos^2 u \Delta v^2} &= R\cos u \sqrt{\Delta v^2 + \left(\frac{\Delta u}{\cos u}\right)^2} \sim R\cos u\sqrt{\Delta v^2 + (2\Delta u)^2} \sim \\ &\sim R\cos u(\alpha \min\{\,\Delta v, 2\Delta u\,\} + \beta \max\{\,\Delta v, 2\Delta u\,\}) = \\ &= \frac{R\pi}{180 \cdot 60}\cos u(\alpha \min\{\,\Delta V, 2\Delta U\,\} + \beta \max\{\,\Delta V, 2\Delta U\,\}) = \\ &= \cos u(1.67 \min\{\,\Delta V, 2\Delta U\,\} + 0.68 \max\{\,\Delta V, 2\Delta U\,\}), \end{aligned}$$

where ΔU and ΔV are the differences of latitudes and longitudes in minutes. The coefficients $\frac{7}{8}$ and $\frac{3}{8}$ in the formulation of the rule indicate that $\cos u$ was approximately taken to be 0.53, which corresponds to the latitude of 58°.

v. Approximate quadratures. The applications of best approximation to this particular question were the subject of Chebyshev's last two works: "On approximate expressions for the square root of a variable

* A further improvement of this result is due to A.A. Markov [58].

** Printed in [13].

in terms of simple fractions" (1889) [11], and "On polynomials that represent the values of the simplest rational functions best when the argument is bounded between two given limits" (1892) [12]. Chebyshev writes: "While computing quadratures, it is often necessary to replace the functions that present difficulty for integration by approximate expressions". In the first of the two articles mentioned, the problem of the best relative approximation to the expression $\frac{1}{\sqrt{x}}$ by a rational expression of the form $A+\sum_1^n \frac{B_i}{C_i+x}$ is being solved, and a special application to computing integrals of the form $\int \frac{U}{\sqrt{V}}\,\mathrm{d}x$ is given, in particular, for elliptic integrals

$$\int \frac{\mathrm{tg}^{p-1}\,x}{\sqrt{1-\lambda^2\sin^2 x}}\,\mathrm{d}x \quad (0<p\le 1).$$

In the second article, an approximate equality is established for the interval $|x|\le h$

$$\frac{1}{H-x} \sim \frac{1}{T_n\left(\frac{H}{h}\right)} \cdot \frac{T_n\left(\frac{x}{h}\right)-T_n\left(\frac{H}{h}\right)}{x-H} \quad (h<H).$$

This means that, among all polynomials of degree $n-1$, the one on the right gives the smallest relative error on the interval $|x|\le h$. This allowed Chebyshev to approximate the integrals of the form $\int_{-h}^{+h}\frac{f(x)}{H-x}\,\mathrm{d}x$ by linear combinations of integrals of the form $\int_{-h}^{+h} x^k f(x)\,\mathrm{d}x$.

vi. Constructing geographic maps. If one needs to draw on a map some piece of the Earth' surface with a given boundary, then there is a choice among infinitely many projections that provide the infinitesimal similarity and preservation of the scale in each point and in all directions. These are the so-called conformal projections. However, as follows from Gauss' *Theorema egregium*, among all conformal projections of a ball to a plane, it is impossible to find a projection that would preserve the scale for all points of the surface.

In a talk given on January 30 (18) of 1856 published under the title "Sur la construction des cartes géographiques" in the "Notes of the Academy of Sciences" [3], Chebyshev posed the problem of finding a conformal projection for which the logarithm of the scale would vary within tightest possible bounds, that is, it should have the smallest possible deviation from some average value. Without a proof, he claimed that, if the specified condition is satisfied, then the scale should be constant *on the boundary* of the map (note that this last requirement can be satisfied by the Dirichlet principle). Chebyshev's statement was proved much later by Academician D.A. Grave.*

The problem is reduced to finding a function U harmonic in the given domain (D) that deviates least from some given function ϕ. In the simplest case, $\Delta\phi$ does not change sign in the domain (D), where Δ is the Laplace operator. That is, as one now says, the function ϕ is subharmonic or superharmonic. In the cartographic problem of interest here, one has to deal with a superharmonic function.

In the simplest case just mentioned, it turns out that the function U only differs by an additive constant from some harmonic function U_0 that coincides with ϕ on the boundary (so that $U_0<\phi$ in the interior of D)

$$U = U_0 + C.$$

As far as the constant C is concerned, it is easy to see that it equals

$$C = \frac{1}{2}\max_{(D)}(\phi - U_0).$$

* See [45]. The proof is simplified in [46]. It is reproduced in an article by N.G. Chebotarev in "A collection devoted to the memory of Academician D.A. Grave", 1940.

8. The extremal problems solved by E.I. Zolotarev, A.A. and V.A. Markov, and N.I. Akhiezer

Already in Chebyshev's time, research in the theory of best approximation of functions was continued by other authors – his students.

For example, E.I. Zolotarev in his 1868 dissertation [78] considered the problem of finding a polynomial of the form

$$P_n(x) = x^n + \sigma_1 x^{n-1} + p_2 x^{n-2} + \cdots + p_n \quad (\sigma_1 \text{ is given}),$$

that deviates least from zero on some given interval. It turned out that the solution could be expressed, generally speaking, in terms of elliptic functions, similar to Chebyshev's simplest problem where the solution can be expressed in terms of trigonometric functions. A different problem, to minimize the polynomial of the form

$$P_n(x) = x^n + p_1 x^{n-1} + \cdots + p_n$$

on some given interval, with the additional condition

$$P_n(H) = M \quad (H > 1),$$

reduces to the first problem.

In 1884, A.A. Markov generalized Chebyshev's "second case" [56] to the case of a weight of the form

$$\varrho(x) = \frac{1}{\sqrt{f(x)}},$$

where $f(x)$ is a given polynomial of even degree.

In 1889, the same author [57] answered the question posed by D.I. Mendeleev in his work "An investigation of liquid substances by their specific weight" (§ 86): Assuming that the maximum of the absolute value of some polynomial $P_n(x)$ of degree n on the interval $|x| \leq 1$ is equal to 1, find the upper bound for the maximum of the modulus of its derivative in the same interval. Markov proved that this upper bound is equal to n^2. As is easy to check, it is attained by the Chebyshev polynomial at the points ± 1. In particular, this implies that the polynomial $T_n(x)$ is a solution to the following problem. Among all polynomials of the form $P_n(x) = p_0 x^n + \cdots + p_n$ with absolute value not exceeding 1 in the interval $|x| \leq 1$, find the one that maximizes the quantity $P_n'(1) = np_0 + (n-1)p_1 + \cdots + p_{n-1}$. This problem is equivalent to the following one. Among all polynomials $P_n(x) = p_0 x^n + \cdots + p_n$ with coefficients satisfying the linear condition $np_0 + (n-1)p_1 + \cdots + p_{n-1} = 1$, find the one that least deviates from zero.

In 1892, V.A. Markov, the younger brother of A.A. Markov, formulated [60] a general problem of finding a polynomial $P_n(x)$ of degree n that deviates least from zero on the interval $|x| \leq 1$ under a given linear condition on the coefficients

$$\alpha_0 p_0 + \alpha_1 p_1 + \cdots + \alpha_n p_n = 1,$$

which is equivalent to computing the maximum of $|\alpha_0 p_0 + \cdots + \alpha_n p_n|$ under the condition $\max_{|x|\leq 1} |P_n(x)| \leq 1$. In particular, he found the upper bounds $\left|\frac{P_n^{(n-k)}(0)}{(n-k)!}\right| = |p_k|$ under this condition. In other words, he found the polynomials of degree n that deviate least from zero when the coefficient of an *arbitrary* power of x is given. Besides, in connection with the problem of least deviation under the condition $P_n^{(k)}(\xi) = 1$ (where ξ is a given value of x), V.A. Markov found the exact upper bounds on the expression $\max_{|x|\leq 1} |P_n^{(k)}(x)|$ under the assumption $\max_{|x|\leq 1} |P_n(x)| = 1$.

The problem of finding a polynomial of degree n that deviates least from zero under two linear conditions on the coefficients (a generalization of the Zolotarev problem: $p_0 = 1$, $p_1 = \sigma$) was studied later by A.P. Psheborsky [62] and other authors.

More recently, the research in best approximation in the classical direction started by Chebyshev and Zolotarev was successfully continued by N.I. Akhiezer, who obtained a series of new exact results in his major 1928 work [15]. Of them, we shall just mention a few:

1. A solution to the Zolotarev problem of minimizing a polynomial with two given coefficients

$$P_n(x) = x^n + \sigma_1 x^{n-1} + p_2 x^{n-2} + \cdots + p_n$$

with an arbitrary Chebyshev weight $\varrho(x) = \frac{1}{A(x)}$, where $A(x)$ is a polynomial.

2. A solution to the problem of minimizing a polynomial

$$P_n(x) = x^n + \sigma_1 x^{n-1} + \sigma_2 x^{n-2} + p_3 x^{n-3} + \cdots + p_n$$

with *three* given coefficients under a constant weight.

3. A solution of the fundamental problem of minimizing a polynomial with one given coefficient

$$P_n(x) = x^n + p_1 x^{n-1} + \cdots + p_n$$

for the case of *two intervals* $-1 \le x \le \alpha$; $\beta \le x \le 1$ $(-1 < \alpha < \beta < +1)$.

Out of further results of N.I. Akhiezer we shall mention the following:

1. A solution to the problem of minimizing a polynomial $P_n(x)$ on two intervals under additional conditions of the form $P_n(x_i) = y_i$ $(i = 1, 2, \ldots, l)$ [17].

2. A significant improvement of the method used by Chebyshev for approximating a polynomial by a rational function with variable coefficients in both the numerator and denominator (Chebyshev's "third case"). Without using continued fractions, N.I. Akhiezer obtains the least deviation as a root of an algebraic equation, while the coefficients are determined from a system of linear equations [16].

As far as the methodology is concerned, N.I. Akhiezer widely uses the elliptic functions, as does Zolotarev, which is connected to the nature of the problems themselves. On the other hand, the proofs are based on the theory of functions of complex variables, including some newest results in this area.

As an example of this kind of proof, let us present a solution to Chebyshev's main problem ("the first case").

According to the necessary conditions, the minimizing polynomial

$$y \equiv P_n(x) = x^n + p_1 x^{n-1} + \cdots + p_n$$

takes on its extremal values $\pm L$ exactly $n+1$ times in the interval $-1 \le x \le +1$. Therefore, the polynomial $y^2 - L^2$ of degree $2n$ has $2n$ zeros in this interval, counting multiplicity. Let us consider the behavior of the function of a complex variable

$$\eta = \frac{y + \sqrt{y^2 - L^2}}{L}$$

in the domain (D) obtained by removing the interval $(-1, +1)$ from the extended complex plane (here the sign of the radical is chosen according to the condition $\eta = \infty$ when $y = \infty$). We then have

$$y = \frac{L}{2}(\eta + \frac{1}{\eta}).$$

The function η of the variable x is regular in the domain (D), since $y^2 - L^2 \neq 0$. It does not have any zeros in (D) and it has a single pole of order n at $x = \infty$. Its values on the boundary of the domain have absolute value 1, since it follows from $|y| \le L$ that

$$|\eta| = \left|\frac{y}{L} + \sqrt{\left(\frac{y}{L}\right)^2 - 1}\right| = \left|\frac{y}{L} + \mathrm{i}\sqrt{1 - \left(\frac{y}{L}\right)^2}\right| = 1.$$

Set

$$\xi = x + \sqrt{x^2 - 1}, \quad x = \frac{1}{2}\left(\xi + \frac{1}{\xi}\right)$$

(here the sign of the radical is taken according to the condition: $\xi = \infty$ when $x = \infty$). The domain (D) in the x-plane is mapped, under these relations, into the exterior of the disk $|\xi| \ge 1$, and, if one considers η as a function of ξ, then it turns out that the function is regular when $|\xi| > 1$, has a pole of order n at the point $\xi = \infty$, and $|\eta| = 1$ on the boundary $|\xi| = 1$. This implies

$$\eta = c^{\xi - n},$$

where c is a constant with modulus one. If one replaces ξ by $\frac{1}{\xi}$, the value of x, and, therefore, the value of y do not change, which implies that $c = \pm 1$. Thus,

$$y = \pm\frac{L}{2}(\xi^n + \xi^{-n}) = \pm\frac{L}{2}\left\{ (x+\sqrt{1-x^2})^n + (x-\sqrt{1-x^2})^n \right\} = \pm L T_n(x),$$

while the comparison of the leading coefficients shows that the sign is '+' and that

$$L = \frac{1}{2^{n-1}}.$$

9. The connection between best approximation and differential properties of a function. The works of S.N. Bernstein

A new content was given to the theory of best approximation of functions in the first decade of the current century.

It originated from a proposition established in 1885 by Weierstrass, the head of the Berlin school of mathematics. For any function $f(x)$ continuous on a given interval $a \le x \le b$ and for any $\varepsilon > 0$, there is a polynomial $P(x)$ such that

$$\max_{a\le x\le b} |P(x) - f(x)| \le \varepsilon. \tag{11}$$

By considering a sequence of values $\{\varepsilon_n\}$ converging to zero, we conclude that the function $f(x)$ is the limit of a uniformly converging sequence of polynomials. Since Weierstrass also proved that the limit of a uniformly converging sequence of continuous functions is also a continuous function, the property established by Weierstrass is exactly equivalent to the continuity of the function in the given interval. In other words, "the set of polynomials is everywhere dense in the set C", or "the system of power functions $1, x, x^2, \ldots, x^n, \ldots$ is fundamental†".

The discovery of Weierstrass, despite its deep and fundamental value, was not immediately appreciated and did not invoke immediate responses. This can be explained, on one hand, by the fact that it did not seem impressive since there were some vague ideas about representing an "arbitrary" function by an analytic formula that were prepared by Fourier, Dirichlet, and Riemann, and, on the other hand, since the Weierstrass theorem was the first stone in the fundament of functional analysis and a solid basis for further development of this new direction was not yet in place.

This all changed soon after the appearance of Lebesgue's works and Borel's monograph "Leçons sur les fonctions des variables réelle" [41]. A question arose: what is the dependence between the number ε in the inequality (11), that is, between the "deviation" of the polynomial $P(x)$ from the function $f(x)$ and the degree of $P(x)$, and it was very soon correctly pointed out that the answer to this question depends on the differential properties of the function $f(x)$. The lower bound on the numbers ε in the equation (11) for a given degree of the polynomial $P(x)$ is exactly the least deviation of the polynomial $P(x)$ from the function $f(x)$ considered by Chebyshev. Upon denoting this deviation (écart) by $E_n(f)$, we see that the sequence of numbers $E_n(f)$ is non-increasing, and by the Weierstrass theorem,

$$\lim_{n\to\infty} E_n(f) = 0.$$

While Chebyshev was interested in the exact computation of the number $E_n(f)$ for a given n, and also in constructing the corresponding uniquely defined polynomial $P(x)$, it was important from the new viewpoint of function theory to determine the rate at which the numbers $E_n(f)$ decrease. Therefore, the asymptotic side of the best approximation problem is now taking precedence.

It is not surprising that initial attention was paid to continuous functions with the simplest kinds of singularities, namely the functions whose graphs have a corner (for example, $f(x) = |x|$, at $x = 0$). Further along in the queue were more general classes of functions, namely those that are not differentiable on a given

† lit.: has the property of completeness

interval, such as $f(x) = |x|^s$ $(s > 0)$. In 1903, the Belgian Academy of Sciences, following a suggestion of its member de la Vallée-Poussin, posed the following research challenge: "To present new investigations in the area of expanding real or analytic functions into series of polynomials". De la Vallée-Poussin posed the following precise question: "Is it possible to approximate a polygonal line, or, which is the same, $|x|$, in the interval $[-1, 1]$ by a polynomial of degree n at a rate higher than $\frac{1}{n}$?" In other words, is it possible to replace the expression $E_n(|x|) = O(\frac{1}{n})$ by a more precise $E_n(|x|) = o(\frac{1}{n})$?

Research in this direction was started by de la Vallée-Poussin himself. In a work published in 1910 [74], he demonstrates a trick that allows him to get a lower bound on $E_n(|x|)$, whereas any approximating polynomial, obviously, gives an upper bound, and he uses the trick to obtain the inequality

$$E_n(|x|) > \frac{k}{n \lg^3 n} \quad (k > 0).$$

Approximation of $|x|^{1/2}$ is also considered there.

One year later, the Belgian Academy of Sciences was presented a work that completely answered the question posed. In his account, de la Vallée-Poussin said that this work is "the most valuable contribution to the area of expanding functions into polynomial series, judging by both the number and the importance of the results it contains". This was the work [19], which later became the Ph.D. dissertation of S.N. Bernstein, who is presently an academician.

Let us briefly mention the most significant results of this work.

(1) The question posed by de la Vallée-Poussin was answered negatively: there are positive numbers A and B such that

$$\frac{A}{n} < E_n(|x|) < \frac{B}{n}.$$

(2) If $E_n(f) = O(\frac{1}{n^{p+\varepsilon}})$, then the function $f(x)$ has a continuous derivative of order p.

(3) If $E_n(f) = O(\varrho^n)$, where $0 < \varrho < 1$, then the function $f(x)$ is regular not only on the basic interval $[-1, +1]$, but also in the ellipse with the foci ± 1 and with the sum of the semi-axis equal to $1/\varrho$. The converse is also true. If a function $f(x)$ is regular in the closed ellipse with the foci ± 1 and with the sum of the semi-axis equal to $1/\varrho$, then

$$E_n(f) = O(\varrho^n).$$

(4) As a tool in his proofs, S.N. Bernstein uses the following theorem, of great independent interest. *If $P_n(x)$ is a polynomial of degree n, then the inequality*

$$\max_{-1 \le x \le +1} |P_n(x)| \le 1$$

implies the inequality

$$|P_n'(x)| \le \frac{n}{\sqrt{1 - x^2}} \quad (-1 < x < +1).$$

(5) An upper bound on $E_n(f)$ is established as a function of the upper bound on $|f^{(n+1)}(x)|$ in a given interval.

(6) A regular method for estimating $E_n(f)$ is specified (a parametric method using analytic continuation).

(7) A criterion is given for the fundamentality of the system

$$x^{\alpha_1}, x^{\alpha_2}, \ldots, x^{\alpha_n}, \ldots \quad (0 \le \alpha_1 < \alpha_2 < \cdots < \alpha_n < \cdots).*$$

While the connection between differential properties of a function and the order of best approximation by polynomials of degree n was studied in France, it also gained attention, in a more general setting, in Göttingen. Here, largely due to E. Landau, another challenge problem of the same type was announced. Already in 1911, the American mathematician D. Jackson [51] presented a dissertation, in which he proved

* See the details in [61] and [68]: the system is fundamental iff the series $\sum \frac{1}{\alpha_n}$ diverges.

a series of theorems that are converses to Bernstein's theorems. From the positive assumptions on the differential properties of a function being approximated, conclusions are drawn about the rate $E_n(f)$ of best approximation.**

By 1912, a rather complete theory was formed based on the results of S.N. Bernstein and D. Jackson, that became the main part of the report presented by S.N. Bernstein at the International Mathematics Congress in Cambridge [21].

About the same time, Bernstein [20] studied Chebyshev's "second case" (weighted approximation), and found the exact value of $E_n\left(\frac{1}{x-a}\right)$, where $|a| > 1$, and the asymptotic value of $E_n\left(\frac{1}{(x-a)^k}\right)$. This gave the asymptotic value of $E_n(f)$ for an arbitrary function $f(x)$ regular in the interval $-1 \leq x \leq +1$, under the following condition: the smallest ellipse with foci ± 1 at which the function is not regular, but is still regular in the interior of the ellipse, contains only one algebraic singularity. If there are many such singularities, the picture becomes more complicated.

All the wonderful results obtained during those few years led to the idea that the rate of decrease of $E_n(f)$ could be taken as a basis for a unified classification of functions of real and complex variables. Thoughts to that effect were reflected in S.N. Bernstein's memoir "Sur la definition et les propriétés des fonctions analytiques d'une variable réelle" [22] published in 1914 in *Math. Annalen.* By considering algebraic polynomials as the major elementary basis for the theory of functions, Bernstein founded a new direction in the theory of functions, which he later named "constructive".

There is no doubt that World War I had a negative influence on the further development of approximation theory. But already in 1919, there appears a new monograph by de la Vallée-Poussin "Leçons sur l'approximation des fonctions d'une variable réelle" [75], which synthesized in a systematic way the many facts obtained, starting with Weierstrass, in functional approximation theory. Note one property of this book. It is the first book to precisely pose and study (in connection to each other) two analogous problems: approximating a function given on an interval of the real line by an algebraic polynomial of degree n and approximating a periodic function with period 2π by "trigonometric polynomials" of degree n (de la Vallée-Poussin still called them "trigonometric sums"). The trigonometric polynomials, as we have seen, can be first found, in some specific problems, in Chebyshev's work of 1881.

Another major contribution to approximation theory was a French monograph by S.N. Bernstein "Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle" [27] published a few years later, which was based on lectures given at the Sorbonne in 1923. In a revised version, it was published much later in Russian under the title "Extremal properties of polynomials" [35]. It contains, on one hand, a large number of results by Chebyshev, by the Markov brothers, and by the author himself. On the other hand, it contains a large amount of new material, which is in part discussed below. For now, we note the following particular questions: (1) the asymptotic value of the least deviation from zero of a polynomial of degree n with an arbitrary number of given highest coefficients, and an arbitrary weight; (2) finding the next terms in the asymptotic expansion for the least deviation from zero of elementary rational functions; (3) a deep investigation of best approximation of functions with an essential singularity.

Very recently, S.N. Bernstein returned to the problem of best approximation of the simplest functions that are not infinitely differentiable, $|x|^s$ $(s > 0)$, and proved in [37] an asymptotic equality, for $n \to \infty$, of the form

$$E_n(|x|^s) \sim \frac{\mu(s)}{n^s},$$

where $\mu(s)$ is a continuous function of the variable s that naturally takes the value zero at even integers s. He also established in [38] a more general relation

$$E_n(|x-c|^s) \sim (1-c^2)^{s/2}\frac{\mu(s)}{n^s} \quad (-1 < c < +1)$$

and found a method for computing $E_n(f)$ asymptotically for an arbitrary function $f(x)$ that has only a finite number of "corners" on the basic interval. Thus, the case of "algebraic-logarithmic singularities" in

** In this connection, let us note, by the way, a very recent and very precise result by N.I. Akhiezer and M.G. Krein [18], who found an upper bound on the best approximation by trigonometric polynomials of degree n for the class of periodic functions $f(x)$ satisfying the inequality $\max|f^{(r)}(x)| \leq 1$.

the basic interval was considered as thoroughly as the case when the singularities of this kind are located outside the interval.

10. Further results dealing with best approximation of functions

In this section, we will discuss some of the newest developments that are more or less close to the major line drawn by Chebyshev but can, strictly speaking, at times lie outside of the circle of his ideas, but that nevertheless carry the mark of Chebyshev's style. It is also very important to note that Chebyshev's ideas are penetrating, at an increasing rate, areas such as the theory of functions and functional analysis that were initially foreign to them.

In 1928, S.N. Bernstein [30] generalized the notion of an increasing function by introducing multiply monotone functions. To be precise, a function is called multiply monotone of order $h+1$ in a given interval if all of its derivatives up to (and including) those of order $h+1$ are non-negative on the given interval. Already for the most general multiply monotone functions $f(x)$, some extremal problems arise. In particular, Bernstein found bounds for $f^{(k)}(x_0)$, where x_0 is an internal point of the interval, while the values of $f(x)$ at the end points of the interval are given. We shall not discuss here the subject of absolutely and regularly monotone functions, which opened a new chapter in the modern theory of functions of a real variable. We shall only mention that Bernstein formulated the main problems of Chebyshev (find a polynomial that deviates least from zero when the leading coefficient is given) and Markov (find a polynomial that deviates least from zero when the derivative at some point is given) in the case of multiply monotone polynomials. As we have seen, Chebyshev had considered the first of these two problems in the case of standard monotone polynomials, and the solution turned out to be connected to Jacobi polynomials. The latter is also true in the more general case of multiply monotone polynomials. Later, a whole series of various versions of the classical problems were considered by S.N. Bernstein' students – Ya.L. Geronimus and V.F. Brzhechka – for the case of monotone and multiply monotone polynomials.

A new development was given to classical problems in the spaces $L^{(1)}$ and $L^{(1)}(\varrho)$. Chebyshev himself considered the integral of the absolute value of a function in the memoir "On interpolation in the case of a large number of data points obtained from an observation" [5]. Furthermore, in 1873, Korkin and Zolotarev (inspired by Chebyshev?) solved the problem of finding a polynomial of given degree with a given leading coefficient that minimizes the integral of the absolute value. It turned out that the polynomial of interest is, up to a scalar multiple, equal to

$$U_n(x) = \frac{\sin(n+1)\arccos x}{\sqrt{1-x^2}}.$$

The next more general result was formulated in 1927 by Bernstein [28], who found that, under rather general assumptions, the polynomial that deviates least from a given continuous function in the sense of the space $L^{(1)}$ is the interpolation polynomial with the nodes being the zeros of the polynomial $U_{n+1}(x)$. From this, by the way, it follows immediately that the polynomial $P_n(x)$ with a given leading coefficient that minimizes the total variation, that is, the integral

$$\int_{-1}^{+1} |P_n'(x)|\,\mathrm{d}x,$$

again, coincides, up to a scalar multiple, with Chebyshev's polynomial $T_n(x)$. In 1934 and later, research in this direction was continued by Ya.L. Geronimus, V.F. Brzhechka and N.I. Akhiezer.

The application of Chebyshev's ideas to generalized polynomials as well as to function spaces of more general type began with a paper written in 1907 by the American mathematician Young [77]. In 1913, Polya proved [63] that the polynomial of degree n deviating least from a given function in the space $L^{(2k)}$ (k being an integer) tends to the polynomial deviating least from that function in the Chebyshev space C as $k \to \infty$. He remarked that the idea of power-norms is due to K. Runge (but its roots are even deeper; recall the so-called "Gräffe method"). In the 20s, Jackson took up approximation theory in $L^{(s)}$-spaces and showed how best approximation depends on analytic and differential properties of the function. *

* We refer the reader to Jackson's monograph [52].

It is more or less clear that finding exact solutions to problems of Chebyshev type cannot be guaranteed in arbitrary $L^{(s)}$-spaces. The question then becomes one of existence, of estimating the error of approximation, or of developing algorithms that converge. E. Ya. Remez, in particular, worked in that last direction.

A. N. Kolmogorov [54] posed the question of the best (in the sense of Chebyshev) choice of the basis of "generalized polynomials" in an arbitrary metric space for a class of functions to be approximated. He solved the problem for a particular class of functions in the space $L^{(2)}$.

It often happened that complex analysis elucidated phenomena taking place on the real line. Approximation theory was no exception. In 1919, G. Faber [43] introduced the notion of the Chebyshev polynomial

$$T_n(z) = z^n + p_1 z^{n-1} + \cdots + p_n \tag{12}$$

that deviates least from zero on a given closed simply connected subset M of the complex plane. He established a connection between the polynomial $T_n(z)$ and a conformal mapping of the complement of M onto the exterior of a certain disk (this question is related to the Robin constant from potential theory), investigated the location of the roots of the polynomial $T_n(z)$, and found these polynomials explicitly for some specific sets M. For example, if M is an ellipse with foci ± 1, then the corresponding polynomial (12) coincides with the usual Chebyshev polynomial $\frac{1}{2^{n-1}} \cos n \arccos z$. But the most remarkable result – the solution to the Markov problem in the complex domain, i.e., finding a polynomial that deviates least from zero and has a prescribed value of its derivative at a given point – was obtained by Szegö [70]. This generalization clarifies that fact that the inequality $\max_{-1 \le x \le 1} |P_n(x)| \le 1$ implies a bound of order n for $|P_n'(x)|$ at all interior points of the interval $[-1, +1]$ as well as a bound of order n^2 at the end points (A. A. Markov's inequality). By considering domains enclosed by finitely many analytic arcs, Szegö establishes that the upper bound on $|P'(z_0)|$, where $z_0 \in M$, depends on the boundary of M and the location of z_0. For points z_0 in the interior of M, Cauchy's integral formula yields a bound of the form $\frac{O(1)}{\delta^2}$, with δ the distance from z_0 to the boundary of M. If z_0 belongs to the boundary of M and lies on only one of the analytic arcs (so that the boundary of M has a tangent line at that point), then the bound is of order $O(n)$. Finally, if z_0 is the endpoint of two of the analytic arcs and the angle between their tangent lines (external with respect to M) equals $\alpha\pi$, then the bound is of order $O(n^\alpha)$. For the end points of the interval $[-1, +1]$, the value of α is 2, which explains the increase in the order of the bound.

The systematic transfer of the methods of uniform Chebyshev approximation to the complex domain was begun in 1930 and undertaken by Jackson and Walsh. The rich material on that topic is collected in the book of the latter author "Interpolation and approximation by rational functions in the complex domain" [76] published in 1935.

Let us now turn to a question similar to the above-mentioned problem of V. A. Markov, namely to a theorem due to S. N. Bernstein that recently attracted much attention. In its most primitive formulation, the theorem deals with trigonometric polynomials of degree n and can be stated as follows. *If M_n is the maximum of the absolute value of a polynomial and M_n' is the maximum of the absolute value of its derivative, then*

$$M_n' \le n M_n. \tag{13}$$

Applying this theorem to an even polynomial (containing only cosine terms) in a variable θ and making the change of variables $\cos\theta = x$, one obtains an inequality for an algebraic polynomial of degree n:

$$M_n' \le \frac{n M_n}{\sqrt{1 - x^2}}. \tag{14}$$

Conversely, the inequality (14) is easily seen to imply the inequality (13). Of the many proofs of Bernstein's theorem, M. Riesz' proof [66] stands out as the simplest and most elegant. As was shown by Bernstein [31], the inequality (13) is preserved in the asymptotic form for weighted maxima. If $M_n = \max_{-1 \le x \le 1} \varrho(x)|P_n(x)|$, $M_n' = \max_{-1 \le x \le 1} \varrho(x)|P_n'(x)|$, then we obtain

$$M_n' \le (1 + \varepsilon_n) n M_n,$$

with ε_n depending only on n and tending to 0 as $n \to \infty$.

A. A. Markov's inequality generalizes to the setting of the space $L^{(s)}$, viz.

$$M_n' \leq A(s)n^2 M_n,$$

where

$$M_n = \{\int_{-1}^{1} |P_n(x)|^s \,\mathrm{d}x\}^{1/s}, \qquad M_n' = \{\int_{-1}^{1} |P_n'(x)|^s \,\mathrm{d}x\}^{1/s},$$

and $A(s)$ depends only on s [49].

The complex analog of Bernstein's theorem is derived by him in [27] and has the form

$$M_n' \leq nM_n,$$

where

$$M_n = \max_{|z|\leq 1} |P_n(z)|, \qquad M_n' = \max_{|z|\leq 1} |P_n'(z)|.$$

Van der Corput and Schaake [73] found an analog of inequality (13) for binary forms of a given degree. If $f(x, y)$ is a form of degree n, then the inequality (13) holds with

$$M_n = \max \frac{|f(x,y)|}{(x^2+y^2)^{\frac{n}{2}}}, \qquad M_n' = \max \frac{\sqrt{\left|\frac{\partial f}{\partial x}\right|^2 + \left|\frac{\partial f}{\partial y}\right|^2}}{(x^2+y^2)^{\frac{n-1}{2}}}.$$

They also obtained strengthened versions of the inequality (13). For example, if M_n denotes the maximum of the absolute value of a trigonometric polynomial $S_n(x)$ of degree n, then, for all x,

$$|S_n'(x)| \leq n\sqrt{M_n^2 - S_n^2(x)}.$$

S. N. Bernstein obtained a formal generalization of this original theorem by finding an upper bound of the ratio $\frac{M_n'}{M_n}$, where

$$M_n = \max_x |\sum_{m=0}^{n} (a_m \cos mx + B_m \sin mx)|,$$
$$M_n' = \max_x |\sum_{m=0}^{n} \lambda_{n-m}(a_m \cos mx + B_m \sin mx)|,$$

under some restriction on the positive constants λ_i [34]. G. Sokolov [67] considered the case $\lambda_{n-m} = m^\alpha$ $(m = 0, 1, \ldots, n)$ in detail.

The following far-reaching generalization of Bernstein's inequality was obtained by himself in 1923 [23]. Let $f(x) = \sum_0^\infty \frac{a_n}{n!}x^n$ be an entire function such that $\limsup_{n\to\infty} \sqrt[n]{|a_n|}$ is finite and equal to k. If the supremum M of $|f(x)|$ over the reals is finite, then the same is true of the supremum M' of $|f'(x)|$ over the real axis, and $M' \leq kM$.

Finally, note that trigonometric functions are eigenfunctions of a system of differential equations of a simple kind. This observation led E. Carlson, a student of Jackson, to a generalization of Bernstein's inequality in the case of a more general system of differential equations [42]. Then

$$M_n' \leq CnM_n,$$

where

$$M_n = \max |S_n(x)|, \qquad M_n' = \max |S_n'(x)|,$$

$S_n(x)$ is a sum of the form $\sum_0^\infty a_k v_k(x)$, a_k are arbitrary coefficients, $v_1, \ldots, v_k$ are eigenfunctions corresponding to the eigenvalues of the system of differential equations listed in ascending order, and C is a constant depending neither on n nor on the coefficients a_k.

There are several strengthenings of Markov's inequality $M_n' \le n^2 M_n$ under various additional assumptions.*

The most general problem of A. A. Markov type for trigonometric polynomials

$$S_n(x) = a_0 + \sum_{m=1}^{n} (a_m \cos mx + b_m \sin mx)$$

is to minimize the maximum of $|S_n(x)|$, given a linear dependence among the coefficients of S_n.

A similar problem, though of another type, was considered by L. Fejér [44]. Given a linear dependence among the coefficients of a *non-negative* trigonometric polynomial $S_n(x)$, minimize its last coefficient

$$a_n = \int_{-\pi}^{\pi} S_n(x)\,\mathrm{d}x.$$

Fejér found a solution to this problem by representing a non-negative trigonometric polynomial as the square of the absolute value of an algebraic polynomial in the variable $z = \mathrm{e}^{i\theta}$. Bernstein [32] found the solution without passing to the complex domain by applying the classical methods of Chebyshev, which turned out to be quite effective in this case.

Bernstein's works [33], [36], which appeared around 1930, are devoted to orthogonal systems of polynomials defined by a weight of the form

$$\varrho(x) = t(x)q(x),$$

where $q(x) = \frac{1}{\sqrt{1-x^2}}$ and $t(x)$ is a function bounded below and above by positive constants. These works must be mentioned in this survey, for they clarify the relationship between approximations in different function spaces. The starting point is the observation that the norm $||P_n(x)||$ of a polynomial

$$P_n(x) = x^n + p_1 x^{n-1} + \cdots + p_n \tag{15}$$

is minimized by the same Chebyshev polynomial $P_n(x) = \frac{T_n(x)}{2^{n-1}}$ in different spaces, viz. in the spaces C and $L^{(s)}(q)$ for $s \ge 1$.** In addition, if the norms of $f(x)$ in the spaces C, $C(\varrho)$, $L^{(s)}$, $L^{(s)}(\varrho)$ are denoted by

$$N(f), \quad N(\varrho, f), \quad N^{(s)}(f), \quad N^{(s)}(\varrho, f),$$

respectively, then the following equality holds for $P_n(x) = \frac{T_n(x)}{2^{n-1}}$:

$$N^{(s)}(\varrho, P_n) = \sqrt[s]{\frac{\Gamma(\frac{1}{2})\Gamma(\frac{1+s}{s})}{\Gamma(1+\frac{s}{2})}}\,N(P_n). \tag{16}$$

S.N. Bernstein generalizes this result as follows. The polynomial $P_n(x)$ of the form (15) whose $L^2(tq)$-norm is minimal (i.e., a polynomial of degree n that belongs to the orthogonal system defined by the weight $t(x)q(x)$) also *asymptotically* minimizes the norm in the spaces $L^{(s)}(tq)$ for $s > 2$ as well as in the space $C(t)$, and the asymptotic relation

$$N^{(s)}(tq, P_n) \sim \sqrt[s]{\frac{\Gamma(\frac{1}{2})\Gamma(\frac{1+s}{s})}{\Gamma(1+\frac{s}{2})}}\,N(t, P_n) \tag{17}$$

* See, for example, S. N. Bernstein [27], pp.47–50.

** Moreover, S.N. Bernstein shows that the same is true for any space where the norm is given by

$$||f|| = \int_{-1}^{1} \Phi(|f|)\frac{\mathrm{d}x}{\sqrt{1-x^2}},$$

Φ being a non-decreasing function.

holds, at least if the function has some differentiability properties similar to that of Dini-Lipschitz.

An analogous problem – to investigate properties of orthogonal polynomial systems in the complex domain – was successfully solved by G. Szegö [69], [71].

Bernstein has expanded the circle of problems of Chebyshev type in one more way. Chebyshev and his immediate successors had always taken *bounded* intervals on the real axis to be the domains of functions. But, after suitable changes in the setup of the problem, one can consider best approximations on unbounded intervals (the whole real axis or its semi-axis) as well.*

A similar opportunity exists if one considers a weight that tends to zero faster than the function to be approximated and the approximating polynomials as the argument tends to infinity. For example, in the $L^{(2)}$-space with weight e^{-x} on the interval $(0, \infty)$, this gives approximations by Laguerre polynomials. On the interval $(-\infty, +\infty)$ with weight e^{-x^2}, on obtains approximations by Hermite polynomials. It is legitimate to ask a similar question for the Chebyshev space as well, requiring the minimization of the expression

$$\max\{\varrho(x)|f(x) - P_n(x)|\}$$

on an unbounded interval, given that $\varrho(x)$ decays sufficiently fast. Another option, which was also explored by Chebyshev, is to consider approximation by rational functions on an unbounded interval [26], [27].

It is not easy to give an exhaustive list of problems that have arisen during almost a century in connections with the best approximation problems of Chebyshev. To complete our survey, we must mention two more directions of research reflected in the works of S. N. Bernstein.

1. The inverse problem of best approximation. Given a non-increasing sequence of positive numbers converging to zero,

$$a_0 \geq a_1 \geq a_2 \geq \cdots \geq a_n \geq \cdots, \qquad \lim_{n\to\infty} a_n = 0,$$

construct a function $f(x)$ defined on the basic interval that satisfies the equalities [39]

$$E_n(f) = a_n \quad (n = 0, 1, 2, \ldots).$$

2. The theory of quasi-analytic functions. We already pointed out that a classification of functions of a real variable is possible on the basis of best approximations. In particular, one can single out the classes of so-called quasi-analytic functions defined by the property that every function in each class is determined uniquely by its values on an arbitrarily small interval. The simplest of the quasi-analytic classes is that of analytic functions. It is characterized by the inequality

$$\limsup_{n\to\infty} \sqrt[n]{E_n(f)} < 1.$$

But, as was shown by Bernstein already in 1914, the much larger class of functions $f(x)$ satisfying the inequality

$$\liminf_{n\to\infty} \sqrt[n]{E_n(f)} < 1$$

turns out to be also quasi-analytic. It contains even non-differentiable functions. In the beginning of the 20-s, Carleman and A. Denjoy defined another class of quasi-analytic functions by the requirement that functions f in the class be infinitely differentiable and the series

$$\sum_n \frac{1}{\sqrt[n]{M_n}},$$

where $M_n = \max|f^{(n)}(x)|$, diverge.

As was shown later by Bernstein [24], [25], this class can be characterized in terms of best approximations, namely, by the condition

$$\sum \frac{1}{\max_{p\geq 1} p \sqrt[n]{E_p(f)}} = \infty.$$

* This is already the case in the simplest Poncelet problem.

11. References

The works of P.L. Chebyshev

1. The theory of mechanisms that are known under the name of parallelograms. Comp., v.i* pages 111–143; Théorie des mécanismes connus sous le nom de parallélogrammes. Mém. Acad. Sci. Pétersb., t.vii, 1854.
2. On continued fractions. Comp., v.i, pages 203–230; Notes of the Academy of Sciences, v.iii, 1855; Sur les fractions continues. Journ. de math. (ii), 3, 1858.
3. On constructions of geographic maps. Comp., v.i, pages 233–236; Sur la construction des cartes géographiques. Bull. Acad. Sci. Pétersb., xiv, 1856.
4. Some questions about minima connected to representing functions approximately,Comp., v.i, pages 273–378; Sure les questions des minima qui se rattachent à la représentation approximative des fonctions. Mém. Acad. Sci. Pétersb. (6), vii, 1859.
5. On interpolation in the case of a large number of observational data. Comp., v.i, pages 387–469; Sur l'interpolations dans le cas d'un grand nombre de données fournies par les observations. Mém. Acad. Sci. Pétersb. (7), i, 1859.
6. An acount by P.L. Chebyshev, an extraordinary professor from Saint Petersburg University, of his foreign trip. Comp., v.ii, pages vii-xix.
7. On a mechanism. Comp., v.ii, pages 51–57, Notes of the Academy of Sciences, xiv, 1868.
8. On functions that are similar to Legendre functions. Comp., v.ii, pages 61–68; Notes of the Academy of Sciences, xvi, 1870.
9. On functions that deviate least from zero. Comp., v.ii, pages 189–215; Attachement to Notes of the Academy of Sciences, v.xxii, No. 1, 1873; Sur les fonctions qui différent le moins possible de zéro. Journ. de math. (ii), 19, 1874.
10. On functions that deviate least from zero for some values of the variable. Comp., v.ii, pages 335–356; Attachement to Notes of the Academy of Sciences, v.xv, No. 3, 1881.
11. On approximate expressions for the square root of a variable in terms of simple fractions. Comp., v.ii, pages 543–558; Attachement to Notes of the Academy of Sciences, v.lxi, No. 1, 1889.
12. On polynomials that represent the values of the simplest rational functions best when the variable is bound to lie between two given limits. Comp., v.ii, pages 669–678; Attachement to Notes of the Academy of Sciences, v.lxxii, No. 7, 1893.
13. Chebyshev's rule for finding distances on the Earth' surface approximately. Comp., v.ii, p. 736; Mesiatselov for 1869, Acad. Sci.
14. On integration by means of logarithms. Published by recommendation of acad. A.N. Krylov in Izvestiya of USSR Acad. Sci., 8, 1930.

The works of other authors

15. N.I. Akhiezer. Über die Functionen, die in gegebenen Intervallen am wenigsten von Null abweichen. Izvestiya of Kazansk Society of Physics and Mathematics, 3, 1928.
16. –Über ein Tschebyscheffsches Extremumproblem. Math. Ann. 104, 1931.
17. –Über einige Functionen, welche in zwei gegebenen Intervallen am wenigsten von Null abweichen. Izvestiya of USSR Acad. Sci. (7), i, 9; ii, 3; iii, 4, 1932-1933.
18. N. Akhiezer and M. Krein. On best approximation of periodic differentiable functions by trigonometric sums, Doklady of USSR Acad. Sci., 15, 1937.
19. S.N. Bernstein. On best approximation of continuous functions by polynomials of a given degree. Soobshcheniya of Kharkov Society of Mathematics (2), 13, 1920.
20. – Sure la valeur asymptotique de la meilleure approximation des fonctions analytiques admettant des singularités données. Bull. Acad. Belg., 1913.

* Comp.– the Compositions by P.L. Chebyshev published under the supervision of A.A. Markov and N.Ya. Sonin, v.i, SPb., 1899; v.ii, SPb., 1907.

21. – Sur les recherches récentes relatives à la meilleure approximation des fonctions continues par des polynômes. Proc. 5 Intern. Math. Con. i, Cambridge, 1913.
22. – Sur la définition et les propriétés des fonctions analytiques d'une variable réelle. Math. Ann. 75, 1914.
23. – Sur une propriété des fonctions entiéres. C.R. Acad. Sc. Paris., 176, 1923.
24. – Sur les fonctions quasianalytiques. C.R. Acad. Sc. Paris., 177, 1923.
25. – Sur les fonctions quasianalytiques de M. Carleman. C.R. Acad. Sc. Paris., 179, 1924.
26. – Le probléme de l' approximation des fonctions continues sur tout l'axe réel et l'une de ses applications. Bull. Soc. Math. Fr., 52, 1924.
27. – Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle. Paris, 1926.
28. – Sur une propriété des polynômes de Tchébycheff. Doklady of USSR Acad. Sci., 1927.
29. – Sur les polynômes multiplement monotones. Soobshcheniya of Kharkov Society of Mathematics (4), 1, 1927.
30. – Sur les fonctions multiplement monotones. Zapiski fiz.-mat. viddilu VUAN, 3, 1928.
31. – Sur la limitation des dérivées des polynômes. C.R. Acad. Sc. Paris., 190, 1930.
32. – Sur l'application de la méthode de Tschébycheff à une class de problémes de M. Fejér. Izvestiya of USSR Acad. Sci., 1930.
33. – Sur les polynômes orthogonaux relatifs à un segment fini. i-ii. Journ. de math., 9, 1930; iii-iv, same place, 10, 1931.
34. – Sur un théoréme de M. Szegö. Prace mat.-fiz., 44, 1935.
35. – Extremal properties of polynomials and best approximation of continuous functions of one real variable. Part i, 1937.
36. – On polynomials that are orthogonal in a given interval. DNTVU, 1937.
37. – On best approximation of $|x|^p$ by polynomials of rather high degree. Izvestiya of USSR Acad. Sci., 1938.
38. – On best approximation of $|x-c|^p$. Doklady of USSR Acad. Sci., 18, 1938.
39. – Sur le problème inverse de la théorie de la meilleure approximation des fonctions continue.C.R. Acad. Sc. Paris., 206, 1938.
40. H.F. Blichfeldt. Note on the functions of the form $f(x) = \Phi(x) + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_n$ which in a given interval differ the least possible from zero. Trans. Amer. Math. Soc., 2, 1901.
41. E. Borel. Leçons sur les fonctions de variables réelles et les développements en séries de polynômes. Paris, 1905.
42. E. Carlson. Extension of Bernstein's theorem to Sturm-Liouville sums. Trans. Amer. Math. Soc., 26, 1924.
43. G. Faber. Über Tschebyscheffsche Polynome. Crelle Journ., 150, 1919.
44. L. Fejér. Über trigonometrische Polynome. Crelle Journ., 146, 1915.
45. D.A. Grave. On the basic problem in the mathematical theory of constructing geographic maps. 1896.
46. – Démonstration d'un théoréme de Tchébycheff généralisé. Crelle Journ., 140, 1911.
47. A. Haar. Die Minkowskische Geometrie und die Annäherung an stetige Funktionen. Math. Ann., 78, 1917.
48. G.H. Hardy, J.E. Littlewood, G. Polya. Inequalities. Cambridge, 1934.
49. E. Hille, G. Szegö, J.D. Tamarkin. On some generalizations of a theorem of A. Markoff. Duke Math. Journ., 3, 1937.
50. M. Horvarth. Sur les valeurs approximatives et rationnelles de la form $\sqrt{x^2+y^2+z^2}$. Bull. Soc. Philomat. Paris, 4, 1867.
51. D. Jackson. Über die Genauigkeit der Annäherung stetiger Funktionen durch ganze rationale Funktionen gegebenen Grades und trigonometrischen Summen gegebener Ordnung. Diss., Gött, 1911.
52. – The theory of approximation. New-York, 1930.
53. P. Kirchberger. Über Tschebyscheffsche Annäherungsmethoden. Diss., Gött. 1902; Math. Ann., 57.
54. A.N. Kolmogorov. Über die beste Annäherung von Funktionen einer gegebener Funktionenklasse. Ann. di Mat. (2), 37, 1936.
55. A.N. Korkin and E.I. Zolotarev. Sur un certain minimum. Nouv. Ann. Math., 12, 1873.

56. A.A. Markov. Finding the smallest and the largest values of some function that deviates least from zero. Soobshcheniya of Kharkov Society of Mathematics (1), 1, 1884.
57. – On a question by Mendeleev. Izvestiya of Acad. Sci., 62, 1889.
58. – On an algebraic theorem by Chebyshev. Izvestiya of Acad. Sci., 1903.
59. – A new case of Poncelet's problem on approximating the value of the square root of a sum of squares. Izvestiya of Acad. Sci., 24, 1906.
60. V.A. Markov. On functions that deviate least from zero in a given interval, 1892; also see Math. Ann., 77, 1916.
61. Ch. H. Müntz. Über den Approximationssatz von Weierstrass. Schwarz-Festschrift, 1914.
62. A.P. Psheborsky. On certain polynomials that deviate least from zero in a given interval. Soobshcheniya of Kharkov Society of Mathematics (2), 14, 1913.
63. G. Pólya. Sur un algorithme toujours convergent pour obtenir les polynômes de meilleure approximation de Tchébycheff pour une fonction continue quelconque. C.R. Acad. Sc. Paris., 157, 1913.
64. J.V. Poncelet. Sur la valeur approchée linéaire et rationelle des radicaux de la forme $\sqrt{a^2+b^2}$, $\sqrt{a^2-b^2}$, etc. Crelle Journ., 13, 1835.
65. M.H. Résal. Sur un théorème de Poncelet et sa géneralisation par M. Horvarth. Bull. Soc. Math. Fr., 1, 1873.
66. M. Riesz. Formule d'interpolation pour la dérivée d'un polynôme trigonométrique. C.R. Acad. Sc. Paris., 158, 1914; Eine trigonometrische Interpolationsformel und einige Ungleichungen für Polynome. Jahresb. D.M.V., 23, 1914.
67. G. Sokolov. On certain extremal properties of the trigonometric sums. Izvestiya of USSR Acad. Sci., 7, 1935.
68. O. Szász. Über die Approximation stetiger Funktionen durch lineare Aggregate von Potenzen. Math. Ann., 77, 1916.
69. G. Szegö. Über die Entwicklung einer analytischen Funktion nach den Polynomen eines Orthogonalsystems. Math. Ann., 82, 1921.
70. – Über einen Satz von A. Markoff. Math. Ztschr., 23, 1925.
71. – Orthogonal polynomials. Amer. Math. Soc. Coll. Publ., 23, 1939.
72. L. Tonelli. I polinomi d'approssimazione di Tchébycheff. Ann. di Mat. (3), 15, 1908.
73. J.G. Van-der-Corput, G. Schaake. Ungleichungen für Polynome und trigonometrische Polynome. Compos. Math., 2, 1935.
74. Ch. de la Vallée Poussin. Sur les polynômes d'approximation et la représentation approchée d'un angle. Bull. Acad. Belg., 1910.
75. – Leçons sur l'approximation des fonctions d'une variable réelle. Paris, 1919.
76. J.L. Walsh. Interpolation and approximation by rational functions in the complex domain. Amer. Math. Soc. Coll. Publ., 20, 1935.
77. J.W. Young. General theory of approximation by functions involving a given number of arbitrary parameters. Trans. Amer. Math. Soc., 8, 1907.
78. E.I. Zolotarev. An application of elliptic functions to the case of functions that deviate least from zero. Application to Notes of the Academy of Sciences, v.xxx, 1877; Collection of works by E.I. Zolotarev, issue 2.