

LEONID KANTOROVICH AND APPROXIMATION THEORY

V. M. Tikhomirov

UDC 519.65

Abstract: We give a survey of applications of the ideas and methods of Leonid Kantorovich (1912–1986) to the classical problems of approximation theory.

DOI: 10.1134/S0037446622010153

Keywords: alternance, interpolation, deviation, clean-up theorem, Landau–Kolmogorov problem

The contribution by Kantorovich had a very great influence on the mathematics of the twentieth century. His pioneering work [1] laid the foundation of the theory of convex extremal problems and applications of this theory to the optimal allocation of resources were evaluated subsequently by awarding him the Nobel Prize in Economics. Here we expose a possible rethinking of a few fragments of classical approximation theory based on some ideas, methods, and results of Kantorovich.

Kantorovich belonged to the St. Petersburg mathematical school founded by Chebyshev. Approximation theory was one of the main scientific areas of Chebyshev.

Chebyshev was the creator of the first Russian mathematical school on approximation theory. Among his followers in approximation theory, we mention such persons as Markov and Zolotarev. It is believed that this branch of mathematical analysis began to develop after Chebyshev's article [2].

In the classical approximation theory, we distinguish the three stages of development from start to 1970s.

The first, Chebyshev, stage of the newly-emerged theory was devoted to the approximation by polynomials and rational fractions.

The second stage of approximation theory was associated with the name of one of the most prominent Russian mathematicians of the last century, Bernstein, who was one of the teachers of Kantorovich. Bernstein's first fundamental work [3] on approximation theory appeared in 1912. Bernstein had obtained the initial impulse to search into relationship between smoothness and approximation rate from de la Vallée-Poussin, and the means of approximation (polynomials) and the characteristic of approximations (the uniform metric) Bernstein had taken from Chebyshev.

The third stage is connected with the contribution of Kolmogorov whom Kantorovich erstwhile called one of his teachers in a conversation with me.

Let us first single out the topic that Chebyshev, Bernstein, and their followers paid great attention to. This is the *extremal problems of approximation theory* (more precisely, approximation by polynomials, inequalities for the derivatives of polynomials, etc.). We recall some classical results on this topic:

Theorem (of alternance, Chebyshev, 1859). *For a polynomial $\hat{x}(\cdot)$ of degree n to be the best approximation in the space $C([a, b])$ to continuous functions on $[a, b]$ to a function $\xi(\cdot)$ it is necessary and sufficient that the difference $\xi(\cdot) - \hat{x}(\cdot)$ have $(n + 2)$ -alternance, i.e., take its maximal and minimal values at $(n + 2)$ consecutive points of $[a, b]$ while alternating signs.*

The polynomial of degree n least deviating from zero in the metric of $C([-1, 1])$ is the polynomial with nonzero coefficient of the highest order term and the least uniform norm on $[-1, 1]$.

Theorem (of least deviation, Chebyshev, 1854). *The polynomial least deviating from zero is unique and has the form $2^{n-1}T_n(t) = 2^{n-1} \cos n \cos^{-1} t$.*

Theorem (of polynomial extrapolation, Chebyshev, 1881). *Among all algebraic polynomials of degree at most n whose norm in $C([-1, 1])$ is at most $\delta > 0$, the greatest value at a point $\tau > 1$ is taken by the polynomial $\delta T_n(\cdot)$, where $T_n(\cdot)$ is the Chebyshev polynomial of degree n .*

Theorem (Bernstein, 1912). *The following sharp inequality holds:*

$$\|\dot{y}(\cdot)\|_{C(\mathbb{T})} \leq n \|y(\cdot)\|_{C(\mathbb{T})} \quad \forall y(\cdot) \in \mathcal{T}_n,$$

where \mathcal{T}_n is the set of trigonometric polynomials of degree n . Equality is attained at the function $t \mapsto \sin nt$.

The classical results by Chebyshev's students, the Markov brothers and Zolotarev, also belong to this topic. These are fundamental achievements of the first stage of the theory. Each of the results was obtained by special techniques and often with great effort. Now let us try to look at all these problems using the machinery laid by Minkowski and developed in the first half of the last century (especially, in the 1930s) by von Neumann, Fenchel, and Kantorovich. This area of mathematics acquired the name of convex analysis.

Convex analysis studies convex sets, convex functions, and convex extremal problems. Kantorovich was not only one of the founders of convex analysis: Throughout his creative life, he was closely associated with extremal problems, which is witnessed by his creative survey [4]. All four results we have listed belong to convex extremal problems. The first two of them are in essence convex problems without constraints, and the other two are minimization problems for convex functions with convex constraints. Classical analysis deals with necessary conditions for an extremum in smooth problems. They consist of Fermat's Theorem in a problem without constraints and of the Lagrange multiplier rule for problems with constraints. In convex analysis, the notion of differential, involved in Fermat's Theorem, is replaced by the notion of subdifferential and, instead of the necessary condition for an extremum at a point \hat{x} consisting in the equality $f'(\hat{x}) = 0$, the criterion $0 \in \partial f(\hat{x})$ arises for a minimum. For problems with constraints, the Lagrange multiplier rule with the Lagrange multiplier for the minimizing functional different from zero is also a criterion for minima of the problems.

We made this brief reminder to clarify the equivalence of the solution of a convex problem and the solution of the relations following from the minimum criteria. In order to demonstrate a unified way to proving the above results, it remains only to add three general statements of convex analysis. Without formulating the exact conditions of the theorems (the reader is referred to [5]), we formulate the essence of these general results.

Under sufficiently broad conditions, the subdifferential of the sum of convex functions is equal to the sum of their subdifferentials, and the subdifferential of the maximum of convex functions is equal to the convex hull of the union of their subdifferentials (the Fenchel–Moreau and Dubovitskii–Milyutin Theorems respectively).

The subdifferential of the norm at zero is the unit ball of the conjugate space, and at the point x_0 different from zero, it is equal to the family of unit vectors in the dual space whose value at x_0 is equal to the norm of x_0 .

Theorem (of clean-up, Levin). *The subdifferential of the maximum of a family of convex functions on \mathbb{R}^n over an upper semicontinuous parameter is equal to the subdifferential of the maximum of at most $(n + 1)$ functions in the family.*

Let us give a scheme of the proofs of the above results. We begin with the theorem of alternance.

Here we have to study the problem

$$\max_{t \in [a, b]} \left| \xi(t) - \sum_{k=0}^n x_k t^k \right| \rightarrow \min, \quad x \in \mathbb{R}^{n+1}. \quad (1)$$

By the Clean-Up Theorem, it suffices to confine consideration only to $m \leq n + 2$ points $a \leq t_1 < \dots < t_m$, and then the solution to the problem

$$f(x) = \max_{t_i \in [a, b]} \left| \xi(t_i) - \sum_{k=0}^n x_k t_i^k \right| \rightarrow \min, \quad x \in \mathbb{R}^{n+1}, \quad (1')$$

is the same vector \hat{x} as the solution to problem (1), i.e.,

$$|\xi(t_i) - \hat{x}(t_i)| = \|\xi(\cdot) - \hat{x}(\cdot)\|_{C([a,b])}.$$

Applying the minimum criterion $0 \in \partial f(\hat{x})$ and using the Dubovitskii–Milyutin criterion on the subdifferential of convex functions, we arrive at the system $\{\sum_{i=1}^m \mu_i t_i^j = 0\}_{j=0}^n$ of $n+1$ homogeneous equations with m unknowns. Moreover, $\mu_i = \lambda_i \operatorname{sgn}(\xi(t_i) - \hat{x}(t_i))$, $\lambda_i > 0$, $\sum \lambda_i = 1$. Hence, $m = n + 2$. Since the determinant of a rectangular matrix consisting of the first $(n+1)$ rows and columns is a Vandermonde one, we conclude that the signs of μ_i alternate, which proves the criterion. \square

To study the least deviation polynomial, we must consider a particular case of the problem

$$\max_{t \in [a,b]} \left| t^n - \sum_{k=0}^{n-1} x_k t^k \right| \rightarrow \min, \quad x \in \mathbb{R}^n. \quad (2)$$

Applying the Alternance Theorem to this problem, we see that, for the polynomial least deviating from zero, which we will denote by $y(\cdot)$, there exist $n+1$ points $a \leq t_1 < \dots < (t_{n+1}) \leq b$ in which the polynomial $y(\cdot)$, while alternating, takes its minimal and maximal values whose modulus is equal to γ . For finishing the proof, we must use a remarkable trick of Chebyshev. Rolle's Theorem and the Fundamental Theorem of Algebra imply that the roots of the derivative of the $y'(\cdot)$ are located inside the interval $[-1, 1]$, and the remaining extremal points are located at its endpoints. This enables us to write down the differential equation for y ; i.e., $(1-t^2)y'^2 = n^2(\gamma^2 - y^2)$, because, on the left and right sides, there are polynomials of the same degrees and the same higher coefficients. Integrating the obtained equation, we conclude that $y = 2^{n-1} \cos n \cos^{-1} t$. \square

REMARK 1. Zolotarev generalized Chebyshev's result about polynomials least deviating from zero by considering the problem

$$\max_{t \in [-1,1]} \left| t^n + \sigma t^{n-1} - \sum_{k=0}^{n-2} x_k t^k \right| \rightarrow \min, \quad x \in \mathbb{R}^{n-1}. \quad (2')$$

To Problem (2'), one must apply Chebyshev's Least Deviation Theorem and then the trick reducing the problem to a differential equation. But while in Chebyshev's case, the differential equation is integrated in trigonometric functions, in Zolotarev's case, the equations (which depend on the parameter $\sigma \in \mathbb{R}$) are integrated either in trigonometric or elliptic functions. The explicit expression for Zolotarev's polynomial is rather complicated, and we will not give it. But application of complex analysis significantly facilitates the problem.

Return to Chebyshev's Theorems. In Chebyshev's extrapolation problem, we face the extremal problem with constraints

$$f_0(x) = p(\tau, x) \rightarrow \max, \quad f_1(x) = \|p(\cdot, x)\|_{C([-1,1])} \leq \delta, \quad \tau > 1. \quad (3)$$

Here $x \in \mathbb{R}^{n+1}$, $p(t, x) = \sum_{k=0}^n x_k t^k$. The Lagrange function of the problem is as follows:

$$\mathcal{L}(x, \bar{\lambda}) = -f_0(x) + \lambda f_1(x) = -p(\tau, x) + \lambda \|p(\cdot, x)\|_{C([-1,1])}.$$

Let the solution to the problem be $p(\cdot, \hat{x})$. Here the Lagrange principle yields the criterion $0 \in \partial \mathcal{L}(\hat{x}, \bar{\lambda})$. Let $\{t_k\}_{k=1}^r$ be the points where $|p(t_k, \hat{x})| = \|p(\cdot, \hat{x})\|_{C([-1,1])} = \delta$, $r \leq n+1$. We have

$$0 \in \partial \mathcal{L}(\hat{x}, \bar{\lambda}) \Rightarrow p(\tau) = \sum_{k=1}^r \mu_k p(t_k),$$

the main identity $\forall p(\cdot) \in \mathcal{P}_n$, $\sum_{k=1}^r |\mu_k| = 1$, $r \leq n+1$, $\mu_k \neq 0$, where \mathcal{P}_n is the set of algebraic polynomials of degree n .

If we suppose that $r \leq n$ then, inserting the polynomial $p_0(t) = \prod_{k=1}^r (t - t_k)$ in the main identity, we conclude that $p = x_0(\tau) \neq 0$; a contradiction. Hence, $r = n + 1$ and, as was proved in Chebyshev's Least Deviation Theorem, the solution to the problem is proportional to the Chebyshev polynomial $T_n(\cdot)$. As a result, we have proved that the value of the problem is equal to $\delta T_n(\tau)$. \square

And, finally, let us infer Bernstein's inequality.

The proof is analogous to that of the previous result if we recall that, due to the homogeneity of the problem with respect to shifts, it suffices to solve the following "pointwise" problem:

$$f_0(x) = \dot{y}(0, x) \rightarrow \max, \quad f_1(x) = \|y(\cdot, x)\|_{C(\mathbb{T})} \leq 1. \quad (4)$$

Here

$$x \in \mathbb{R}^{2n+1}, \quad y(t, x) = \frac{x_0}{2} + \sum_{k=1}^n (x_k \cos kt + x_{n+k} \sin kt).$$

The Lagrange function of the problem is as follows: $\mathcal{L}(x, \bar{\lambda}) = -f_0(x) + \lambda f_1(x)$.

Let $y(\cdot, \hat{x})$ be the solution to the problem. Here the Lagrange principle yielding in the criterion $0 \in \partial \mathcal{L}(\hat{x}, \bar{\lambda})$. Let $\{t_k\}_{k=1}^r$ be the points where $|y(t_k, \hat{x})| = \|y(\cdot, \hat{x})\|_{C(\mathbb{T})} = 1$, $r \leq 2n + 2$. We have

$$0 \in \partial \mathcal{L}(\hat{x}, \bar{\lambda}) \Rightarrow \dot{y}(0) = \sum_{k=1}^r \mu_k y(t_k)$$

(the main property) for all $y(\cdot) \in \mathcal{T}_n$, $\sum_{k=1}^r |\mu_k| = 1$, $r \leq 2n + 2$, $\mu_k \neq 0$. The case of $r > 2n$ is impossible because of Rolle's Theorem (since otherwise the derivative of the polynomial would have more than $2n$ zeros). The assumption that $r < 2n$ leads to a contradiction either: If we consider a trigonometric polynomial with zeros at points $\{t_k\}_{k=1}^r$ (which do not contain zero), at zero, and, possibly, at some more points, so that the total number of simple zeros is equal to $2n$, and insert such a polynomial in the main identity then it turns out that $0 \neq 0$ since the derivative of the polynomial at zero cannot be zero by Rolle's Theorem. Hence, the polynomial $\hat{x}(\cdot)$ attains its maxima and minima at $2n$ points, i.e., satisfies the equation $\dot{y}^2 = n^2(1 - y^2)$ because the trigonometric polynomials on the left and right sides have the same degrees ($4n$), the same zeros (double zeros at the points $\{t_k\}_{k=1}^{2n}$), and the same coefficients at the higher term. The so-obtained equation is satisfied by the function $\sin nt$, which and the sufficiency of the necessary condition in this convex problem imply that $\sin n$ is a solution to the problem, i.e., in particular, the value of the problem is equal to n . \square

REMARK 2. The general "pointwise" problem on the k th ($0 \leq k \leq n - 1$) derivative of an algebraic polynomial is posed as follows:

$$f_0(x) = p^{(k)}(\tau, x) \rightarrow \max, \quad f_1(x) = \|p(\cdot, x)\|_{C([-1,1])} \leq \delta, \quad \tau \in \mathbb{R}. \quad (5)$$

As one can see, this is a generalization of Chebyshev's extrapolation problem. Problem (5) is dealt with in the same way as Chebyshev's extrapolation problem. It turns out that the answer is either the Chebyshev polynomial or one of the Zolotarev polynomials. An additional small variation leads to the well-known Markov brothers' inequalities for polynomials.

Another topic, which we will only briefly cover, is related to the generalization of the Landau-Kolmogorov inequalities. This is a huge set of problems depending on many parameters and various initial data. Problems about inequalities for derivatives are posed in the one-dimensional and multidimensional cases but exact solutions are obtained mainly for functions of one variable. Let us touch on this in a bit more detail.

The exact solution of one-dimensional inequalities in the simplest case of two constraints is equivalent to the exact solution of such extremal problems (where $T = \mathbb{R}, \mathbb{R}_+ \vee \mathbb{T}$)

$$\|x^{(k_0)}(\cdot)\|_{L_{p_0}(T)} \rightarrow \max, \quad \|x^{(k_1)}(\cdot)\|_{L_{p_1}(T)} \leq \gamma_1, \quad \|x^{(k_2)}(\cdot)\|_{L_{p_2}(T)} \leq \gamma_2. \quad (6)$$

The series of problems (6) has many parameters. An individual problem from this family will be denoted by $P(p, k, \gamma, T)$, where $p = (p_0, p_1, p_2)$, $k = (k_0, k_1, k_2)$, $\gamma = (\gamma_1, \gamma_2)$, $T = \mathbb{R}, \mathbb{R}_+ \vee \mathbb{T}$.

It is acknowledged that the first problem in this series, namely, Problem $P(\infty, \infty, \infty, 1, 0, 2, \gamma_1, \gamma_2, \mathbb{R}_+)$, was solved by Landau in 1912 (see [6]). The most brilliant among the solutions of the problems of this series is the solution by Kolmogorov in 1938 (see [7]) of the problem

$$P(\infty, \infty, \infty, k_0, k_1, k_2, \gamma_1, \gamma_2, \mathbb{R}), \quad k_1 = 0, \quad k_2 \geq 2, \quad 1 \leq k_0 \leq k_2 - 1.$$

(This gives the whole topic the title *Landau–Kolmogorov inequalities*.)

In [7], some more general problem was posed, when, instead of three inequalities for derivatives in the L_∞ -metric, more inequalities are used. It is hard to hope that this general Kolmogorov problem will ever be thoroughly settled but there is a metric for which complete solutions can be obtained not only in the one-dimensional case. This is the L_2 -metric on spaces with invariant structure, in particular, in \mathbb{T}^n and \mathbb{R}^n .

The crux of the matter is that, in the case of inequalities for the derivatives in the metric L_2 , after expansion in Fourier series or integrals, extremal problems are reduced to problems of linear programming whose theory was constructed by Kantorovich. The theory of linear programming makes it possible to study problems about inequalities for derivatives in the L_2 -metric thoroughly: prove existence theorems and duality, construct effective numerical methods, and often complete the exact solutions. Let us illustrate this for the one-dimensional case on the torus with several inequalities.

Consider one of the simplest cases in (6), which however is a significant share of solutions in this series. It is about Problem $P(2, 2, 2, k_0, 0, k_2, \gamma_1, \gamma_2, \mathbb{T})$; i.e., in more detail, about the problem

$$\|x^{(k_0)}(\cdot)\|_{L_2(\mathbb{T})} \rightarrow \max, \quad \|x(\cdot)\|_{L_2(\mathbb{T})} \leq \gamma_1, \quad \|x^{(k_2)}(\cdot)\|_{L_2(\mathbb{T})} \leq \gamma_2. \quad (7)$$

Using Parseval's formula, write down the hypothesis of the problem through the Fourier coefficients of $x(\cdot)$ on assuming for certain simplification that $x(\cdot)$ is equal to zero in average. Expand this function in the Fourier series

$$x(t) = \sum_{j \in \mathbb{N}} a_j \cos jt + b_j \sin jt, \quad a_j = \frac{1}{\pi} \int_{\mathbb{T}} x(t) \cos jt dt, \quad b_j = \frac{1}{\pi} \int_{\mathbb{T}} x(t) \sin jt dt.$$

If we denote $a_j^2 + b_j^2$ by x_j , $j \in \mathbb{N}$; then, by Parseval's formula, problem (7) acquires the form of the problem of linear programming whose theory was created by Kantorovich:

$$\sum_{j \in \mathbb{N}} j^{2k_0} x_j \rightarrow \max, \quad \sum_{j \in \mathbb{N}} x_j \leq \gamma_1^2, \quad \sum_{j \in \mathbb{N}} j^{2k_2} x_j \leq \gamma_2^2. \quad (8)$$

The Lagrange function of this problem leads us to the necessity of considering the function $f(z) = -z^{2k_0} + \lambda_1 + \lambda_2 z^{2k_2}$, $\lambda_1, \lambda_2 \geq 0$. By the minimum condition in the Lagrange multiplier rule, if this function is positive at an integer point j then the corresponding coefficient \hat{x}_j (of the solution to problem (8)) must be zero. The function f cannot be negative at integer points (otherwise, the minimum would be equal to minus infinity). Consequently, the coefficients of the solution can be nonzero only if f is equal to zero at integer points. There are either one or two consecutive points like that, and the solution of (8) is reduced to a very easy problem of linear programming with one or two unknowns.

Solving the general problem (7) is reduced to considering several problems with two constraints. Solving the Landau–Kolmogorov problems in their majority can be carried out by the methods we briefly described above.

In this short note, we tried to demonstrate that the most substantial part of classical approximation theory, which was developed by dozens of researchers from the schools of Chebyshev, Bernstein, Kolmogorov, Stechkin, et al. can be rethought on the basis of the ideas and methods of Leonid Kantorovich.

References

1. Kantorovich L. V., *Mathematical Methods of Organizing and Planning Production*, Leningrad State University, Leningrad (1939) (Management Sci., vol. 6, no. 4, 366–422 (1960)).
2. Chebyshev P. L., “Theory of the mechanisms known as parallelograms (1854),” in: *Collected Works*. Vol. 2, Akad. Nauk SSSR, Moscow and Leningrad (1948), 23–51 [Russian].
3. Bernstein S. N., “On the best approximation of continuous functions by polynomials of a given degree (1912),” in: *Collected Works*. Vol. 1, Akad. Nauk SSSR, Moscow (1952), 11–104 [Russian].
4. Kantorovich L. V., “My journey in science (proposed report to the Moscow Mathematical Society),” *Russian Math. Surveys*, vol. 42, no. 2, 233–270 (1987).
5. Magaril-Ilyaev G. G. and Tikhomirov V. M., *Convex Analysis: Theory and Applications*, Amer. Math. Soc., Providence (2003).
6. Landau E., “Einige Ungleichungen für zweimal differenzierbar Funktionen,” *Proc. London Math. Soc.*, vol. 2, no. 13, 43–49 (1913).
7. Kolmogorov A. N., “On inequalities between suprema of successive derivatives of functions on an infinite interval,” *Uch. Zap. Mosk. Univ., Mat.*, vol. 30, no. 3, 3–16 (1938) (*Amer. Math. Soc. Transl.*, vol. 1–2, 233–243 (1949)).

V. M. TIKHOMIROV
LOMONOSOV MOSCOW STATE UNIVERSITY, MOSCOW, RUSSIA
E-mail address: vmtikh@gmail.com