APPROXIMATION THEORY:
A volume dedicated to Borislav Bojanov
(D. K. Dimitrov, G. Nikolov, and R. Uluchev, Eds.)
additional information (to be provided by the publisher)

Twelve Proofs of the Markov Inequality

Aleksei Shadrin

This is the story of the classical Markov inequality for the k-th derivative of an algebraic polynomial, and of the remarkably many attempts to provide it with alternative proofs that occurred all through the last century. In our survey we inspect each of the existing proofs and describe, sometimes briefly, sometimes not very briefly, the methods and ideas behind them. We discuss how these ideas were used (and can be used) in solving other problems of Markov type, such as inequalities with majorants, the Landau–Kolmogorov problem, error of Lagrange interpolation, etc. We also provide a bit of some less well-known historical details, and, finally, for techers and writers in approximation theory, we show that the Markov inequality is not as scary as it is made out to be and offer two candidates for the "book-proof" role on the undergraduate level.

1 Introduction

1.1 The Markov inequality

This is the story of the classical Markov inequality for the k-th derivative of an algebraic polynomial and attempts to find a simpler and better proof that occured all through the last century. Here is what it is all about.

$$\boxed{\|p^{(k)}\| \le \|T_n^{(k)}\|\,\|p\|, \quad \forall p \in \mathcal{P}_n} \tag{1.1}$$

Here (and elsewhere), $\mathcal{P}_n$ is the set of all algebraic polynomials of degree $\le n$, $\|f\| := \max\limits_{x\in[-1,1]} |f(x)|$, and $T_n(x) := \cos n \arccos x$ is the Chebyshev polynomial of degree n. Numerically, the constant is given by the formula

$$\|T_n^{(k)}\| = T_n^{(k)}(1) = \frac{n^2\,[n^2-1^2]\cdots[n^2-(k-1)^2]}{1\cdot 3\cdots(2k-1)}\,,$$

so that, for example,

$$\|p\| \le 1 \quad \Rightarrow \quad \|p'\| \le n^2, \quad \|p''\| \le \frac{n^2(n^2-1)}{3}\,, \quad \|p^{(n)}\| \le 2^{n-1} n!\,.$$

The inequality is sharp, with equality only if $p = \gamma T_n$ where $|\gamma| = 1$.

That's it, simple and elegant.

Proved originally by V. Markov in 1892 in a rather sophisticated way, this inequality plays an important role in approximation theory, and there have been remarkably many attempts to provide it with an alternative proof.

I counted twelve proofs in total which divide into four groups. Here they are to satisfy any taste: long, short, elementary, complex, erroneous, incomplete.

1) original variational proof of V. Markov (1892), which ran to 110 pages,
2) its condensed form given by Gusev (1961),
3) and its "second variation" by Dubovitsky–Milyutin (1965),
4) "small-o" arguments of Bernstein (1938),
5) its variation by Tikhomirov (1975),
6) and another variation by Bojanov (2001),
7) a pointwise majorant of Schaeffer–Duffin (1938),
8) a refinement of Duffin–Schaeffer for the discrete restrictions (1941),
9) trigonometric proof of Mohr (1963),
10) an erroneous proof for Chebyshev systems by Duffin–Karlovitz (1985),
11) a majorant of my own for the discrete restrictions (1992),
12) an incomplete proof of mine for the oscillating polynomials (1996) [which was an attempt to revive the proof of Duffin–Karlovitz].

In our survey we inspect each of the existing proofs and describe, sometimes briefly, sometimes not very briefly, the methods and ideas behind them.

We have three goals.

1) The first one is pedagogical. It is a widely held opinion that, besides the case $k=1$, there is no "book-proof" of the Markov inequality. Almost each monograph in approximation theory cites this result, but only two of them, Rivlin [52] and Schönhage [53] provides a proof, namely that of Duffin-Schaeffer. We offer two more candidates for the book-proof role (4 pages each). Also, we show that the original proof of V. Markov is not as scary as it is made out to be.

2) The second goal is methodological. There are many problems of the Markov type where we need to estimate the max-norm of the k-th derivative of a function f from a certain functional class $\mathcal{F}$; they are, in short, the problems of numerical differentiation. Examples are polynomial inequalities with majorant, Landau–Kolmogorov inequalities, error bounds of certain interpolation processes, etc. For all these mostly open problems, the classical Markov inequality is a model where a new method of the proof can be tested, or where an existing method can be taken from.

3) The final goal is historical. It was the homepage on the History of Approximation Theory (HAT), opened recently by Pinkus and de Boor [57], that formed my decision to write this survey, so that I am also eager to uncover

who proved the Bernstein inequality, why Chebyshev was the first to study Markov's inequality, and how it could happen that Voronovskaya did not read Markov's memoirs.

1.2 Prehistory

Those who try to respect historical details (e.g., Duffin–Schaeffer) call Markov's inequality the inequality of the brothers Markoff, because these details are as follows.

$$\begin{array}{llll} 1889 & \text{A. Markov}, & k=1, & \|p'\| \le n^2 \, \|p\| \,, \\ 1892 & \text{V. Markov}, & k \ge 1, & \|p^{(k)}\| \le \|T_n^{(k)}\| \, \|p\| \,. \end{array}$$

The first Markov, Andrei (1856-1922), was the famous Russian mathematician (Markov chains), while the second, Vladimir (1871-1897), was his kid brother who wrote only two papers and died from tuberculosis at age 26.

Both results appeared in Russian in (as Boas put it) not very accessible papers, so that (to cite Boas once again) they must be ones of the most cited papers and ones of least read.

A. Markov's result for $k = 1$ was published in the "Notices of Imperial Academy of Sciences" under the title "On a question by D. I. Mendeleev" [51]. In his nice survey, Boas [14] describes the chemical problem that Mendeleev was interested in and how he arrived at the question about the values of the 1-st derivative of an algebraic polynomial.

V. Markov's opus "On functions deviating least from zero in a given interval" [7] that contained (amongst others) the result for all k appeared as a small book, 110 pages of approximately A5-format, with the touching subheading "A composition of V. A. Markov, the student of St. Petersburg University", and with the stern notice "Authorized to print by the decision of the Physico-Mathematical Faculty of the Imperial St.-Petersburg University, 25 Oct 1891. Dean A. Sovetov".

Probably, it was S. Bernstein who discovered and popularized both Markov's papers in 1912 when he started his studies in approximation theory. Actually, Bernstein reproved the case $k = 1$ by himself, but the result for general k was beyond his ability (for 26 years). So, quite certain about importance and difficulty of V. Markov's achievement, he organized its translation into German which was published in "Mathematische Annalen" in 1916. Nowadays the text in German helps, perhaps, not much more than the Russian one, so that only a few lucky ones could appreciate the flavour of V. Markov's work. However, for those not very lucky, there is an exposition in English by Gusev [6] (with the flavour of Voronovskaya notations). Even though it puts the first half of V. Markov's proof in a slightly different form, it reproduces its final part almost identically.

As to the A. Markov's paper for $k=1$, it was reprinted (in modern Russian orthography) in his Selected Works (1948), but its English translation had to wait another 50 years for the enthusiasm of de Boor and Holtz (2002).

We close this section with the remark that, actually, the earliest reference to Markov's inequality must be

$$1854 \quad \text{P. Chebyshev}, \qquad k = n, \qquad \|p^{(n)}\| \le 2^{n-1}\, n!\, \|p\|\,,$$

because his result on the minimum of the max-norm of the monic polynomial,

$$\|p\| := \|x^n + c_{n-1}x^{n-1} + \cdots + c_0\| \ge \tfrac{1}{2^{n-1}}\,,$$

is nothing but the inequality

$$\|p\| \ge \tfrac{1}{2^{n-1}}\,\tfrac{1}{n!}\,\|p^{(n)}\|\,,$$

and that is exactly the Markov inequality for $k = n$.

1.3 Pointwise problem for polynomials and other functional classes

We will study the Markov inequality as the problem of finding the value

$$M_k := \sup_{\|p\| \le 1} \|p^{(k)}\|\,.$$

There are many problems of this (Markov) type where we need to estimate the max-norm of the k-th derivative of a function f from a certain functional class $\mathcal{F}$, i.e. to find

$$M_{k,\mathcal{F}} := \sup_{f \in \mathcal{F}} \|f^{(k)}\|\,,$$

and in this section we will list several of them which were (and still are) of some interest to the approximation theory community and to which our studies will be somehow related. But before we start, let us make some general remarks.

There is no way of getting a uniform bound for $\|f^{(k)}\|$ other than bounding $|f^{(k)}(z)|$ pointwise, for each particular $z \in [-1,1]$. Therefore, we have to split the original problem into two subsequent ones.

Problem 1.1. For k integer, find

$$\begin{aligned} M_{k,\mathcal{F}}(z) &:= \sup_{f\in\mathcal{F}} |f^{(k)}(z)|\,, \quad z \in [-1,1]\,,\\ M_{k,\mathcal{F}} &:= \sup_{f\in\mathcal{F}} \|f^{(k)}\| \;=\; \sup_{z\in[-1,1]} M_{k,\mathcal{F}}(z)\,. \end{aligned}$$

(The pointwise estimate is also useful in applications and is therefore of independent interest.)

The solution of both problems depends on what is being meant by a solution.

Ideally, a solution is an effective value or a reasonable upper bound for both suprema.

Another type of solution is a characterization of the function f_z that achieves the supremum in the pointwise problem for each particular z, i.e., a description of its particular properties that distinguish it from the other functions of the given class. In most cases, such a description is not constructive, and cannot help much in finding the actual quantitative value (or bound) for $M_{k,\mathcal{F}}(z)$. But sometimes it leads to conclusions about the qualitative behaviour of the function $M_{k,\mathcal{F}}(z)$, e.g., whether its maximum is attained at the endpoints ± 1, thus helping to solve the global problem. Anyway, knowing a smaller set $\{f_z\}$ where to choose from is always an advantage.

For the pointwise problem, there is always a *one-parameter family* of functions which contains extremal functions f_z for any $z \in [-1,1]$, this is the family $\{f_z\}$ itself. One needs however something more constructive, and it is not too much a surprise that, for the Markov-type problems, this something describes certain equioscillation properties of f_z. It is not so surprising either that the mostly oscillating function f_z^* is thought to be extremal for the global problem.

Below we formulate the Markov-type problems appearing in this survey and give a short description of their current status. More details are given within the text.

Problem 1.2 (Markov problem). For k integer, and $p \in \mathcal{P}_n$, find

$$\begin{aligned} M_k(z) &:= \sup_{\|p\| \le 1} |p^{(k)}(z)|\,, \quad z \in [-1,1]\,, \\ M_k &:= \sup_{\|p\| \le 1} \|p^{(k)}\| = \sup_{z \in [-1,1]} M_k(z)\,. \end{aligned}$$

V. Markov (1892) proved that, for each z, the extremal polynomial is given by

$$f_z(x) = Z_n(x, \theta_z),$$

where $Z_n(x,\theta)$ is a one-parameter family of Zolotarev polynomials having at least n equioscillations on $[-1,1]$. He made a very detailed investigation of the character of the value $M_k(z)$ when z runs through certain subintervals, and proved, using some very fine methods, that the Chebyshev polynomial T_n achieves the global maximum M_k.

Problem 1.3 (Markov problem with majorant or Turan problem). Given a majorant $\mu \ge 0$, denote by $\mathcal{P}_n(\mu)$ the set of polynomials p of degree $\le n$ such that

$$|p(x)| \le \mu(x)\,, \quad x \in [-1,1]\,.$$

For n, k integers, and $p \in \mathcal{P}_n(\mu)$, we want to find the values

$$\begin{aligned} M_{k,\mu}(z) &:= \sup_{p \in \mathcal{P}_n(\mu)} |p^{(k)}(z)|\,, \quad z \in [-1,1]\,, \\ M_{k,\mu} &:= \sup_{p \in \mathcal{P}_n(\mu)} \|p^{(k)}\| = \sup_{z \in [-1,1]} M_{k,\mu}(z)\,. \end{aligned}$$

As in the classical case $\mu \equiv 1$, the extremal polynomial is given by

$$f_z(x) = Z_{n,\mu}(x, \theta_z)$$

where $Z_{n,\mu}(x, \theta)$ is a one-parameter family of weighted Zolotarev polynomials having at least n equioscillations between $\pm\mu$ (this is a relatively simple conclusion). One can expect that the μ-weighted Chebyshev polynomial should attain the global maximum $M_{k,\mu}$, but that was proved only for a few classes of majorants.

Problem 1.4 (Markov problem for perfect splines). A piecewise polynomial function s of degree n with r knots (breakpoints) is called a perfect spline if

$$|s^{(n)}| \equiv \text{const.}$$

Denote the set of perfect splines with $\leq r$ knots by $\mathcal{P}_{n,r}$. For n, k, r integers, we want to find

$$\begin{aligned} M_{k,r}(z) \quad &:= \quad \sup_{s \in \mathcal{P}_{n,r}} |s^{(k)}(z)| \,, \quad z \in [-1,1]\,, \\ M_{k,r} \quad &:= \quad \sup_{s \in \mathcal{P}_{n,r}} \|s^{(k)}\| \quad = \sup_{z \in [-1,1]} M_{k,r}(z)\,. \end{aligned}$$

Karlin [26] was the first (and the last) to study this problem, in 1976, and he proved that an extremal perfect spline is given by

$$f_z(x) = Z_{n,r}(x, \theta_z)\,,$$

where $Z_{n,r}(x, \theta)$ is a one-parameter family of Zolotarev perfect splines in $\mathcal{P}_{n,r}$ having at least $n + r$ equioscillations on $[-1, 1]$ (thus having r knots or being $T_{n,r-1}$). Compared with polynomial cases this fact is rather nontrivial. Globally, the Chebyshev perfect spline $T_{n,r}$ with r knots and $n+1+r$ equioscillations should be a solution.

Problem 1.5 (Landau-Kolmogorov problem on a finite interval). Set

$$W_\infty^{n+1}(\sigma) := \{f : f^{(n)} \text{ abs. cont.}, \ \|f\| \leq 1, \ \|f^{(n+1)}\| \leq \sigma\}\,.$$

For n, k integers, and $\sigma > 0$, find

$$\begin{aligned} M_{k,\sigma}(z) \quad &:= \quad \sup_{f \in W_\infty^{n+1}(\sigma)} |f^{(k)}(z)| \,, \quad z \in [-1,1]\,, \\ M_{k,\sigma} \quad &:= \quad \sup_{f \in W_\infty^{n+1}(\sigma)} \|f^{(k)}(z)\| \quad := \sup_{z \in [-1,1]} M_{k,\sigma}(z)\,. \end{aligned}$$

For $\sigma = 0$ we get the classical Markov problem. In 1978, Pinkus [30] showed that an extremal function is given by

$$f_z(x) = P_{n+1,\sigma}(x, \theta_z)\,,$$

where $P_{n+1,\sigma}(x,\theta)$ is a one-parameter family of the Pinkus perfect splines. (Of course, A. Pinkus did not bestow his own name to the perfect splines he introduced. He called them "perfect splines satisfying $\|P\| = 1$, with exactly $r+1$ knots, $n+1+r$ points of equioscillation, and opposite orientation", and even though he denoted their class by $\mathcal{P}(\sigma)$, one can argue that $\mathcal{P}$ stood for "perfect". I take the credit for putting the more memorable "Pinkus" splines into use in [31].) As with Karlin's proof, the arguments are rather elaborate. In the global problem, the solution must be given by an appropriate Zolotarev spline $Z_{n+1,r}$ (this is known as Karlin's conjecture), but that was proved only in a few particular cases.

Problem 1.6 (Error bounds for Lagrange interpolation). For a continuous function f, and a knot-sequence $\delta = (t_i)_{i=0}^n \subset [-1,1]$, let ℓ_δ be the Lagrange polynomial of degree n that interpolates f on δ. For n, k integers, and any δ, find

$$\begin{aligned} M_{k,\delta}(z) &:= \sup_{\|f^{(n+1)}\|_\infty \le 1} |f^{(k)}(z) - \ell_\delta^{(k)}(z)|\,, \quad z \in [-1,1]\,, \\ M_{k,\delta} &:= \sup_{\|f^{(n+1)}\|_\infty \le 1} \|f^{(k)} - \ell_\delta^{(k)}\| \quad = \sup_{z\in[-1,1]} M_{k,\delta}(z)\,. \end{aligned}$$

This problem attracted a lot of attention, and a large number of various cases for small values of n and k were considered showing that $\omega_\delta(x) := \frac{1}{(n+1)!}\prod_{i=0}^n (x - t_i)$ achieves the global maximum. For general n, Kallioniemi [28] showed in 1976 that

$$f_z(x) = S_{n+1}(x, \theta_z)\,,$$

where $S_{n+1}(x,\theta)$ is a one-parameter family of perfect splines with just one knot θ (this is almost immediate), and established the behaviour of $M_{k,\delta}(z)$ when z runs through certain subintervals, which were surprisingly identical to those in classical Markov's problem. In 1995, a complete solution was found [41], i.e., it was proved that $M_{k,\delta} = \frac{1}{(n+1)!}\|\omega_\delta^{(k)}\|$ for all n and k. This is the only complete result among all Markov-type problems.

Problem 1.7 (Error bounds for general interpolation). We may generalize the previous problem in two different ways.

1) We may consider instead of the Lagrange interpolation any other interpolation procedure, e.g., spline interpolation of degree n at the points $\delta = (t_i)_{i=1}^N$ (with another given sequence of spline breakpoints).

2) Alternatively, we may notice that

$$M_{k,\delta}(z) = \sup_{\|f^{n+1}\|_\infty \le 1} \sup_{f|_\delta = 0} |f^{(k)}(z)|\,,$$

and consider the problem of estimating the k-th derivative of a function f that satisfies $\|f^{(n+1)}\| \le 1$ and vanishes on $\delta = (t_i)_{i=1}^N$ (which is related to the problem of optimal interpolation).

Both problems are almost untouched. We can mention only the paper by Korneichuk [40] who considered approximation of the 1-st derivative by interpolating periodic splines on the uniform knot-sequence.

1.4 Zolotarev polynomials

General properties. Here we describe some properties of Zolotarev polynomials which are solutions to the pointwise Markov problem and which bear a certain similarity with one-parameter families from the other Markov-type problems.

Definition 1.8. A polynomial $Z_n \in \mathcal{P}_n$ is called Zolotarev polynomial if it has at least n equioscillations on $[-1,1]$, i.e. if there exist n points

$$-1 \le \tau_1 < \tau_2 < \cdots < \tau_{n-1} < \tau_n \le 1$$

such that

$$(-1)^{n-i} Z_n(\tau_i) = \|Z_n\| = 1.$$

There are many Zolotarev polynomials, for example the Chebyshev polynomials T_n and T_{n-1} of degree n and $n-1$, with $n+1$ and n equioscillation points, respectively. One needs one parameter more to get uniqueness. A convenient parametrization (due to Voronovskaya) is through the value of the leading coefficient:

$$\frac{1}{n!} Z^{(n)} \equiv \theta \quad \Leftrightarrow \quad Z_n(x) := Z_n(x,\theta) := \theta x^n + \sum_{i=0}^{n-1} a_i(\theta) x^i \,.$$

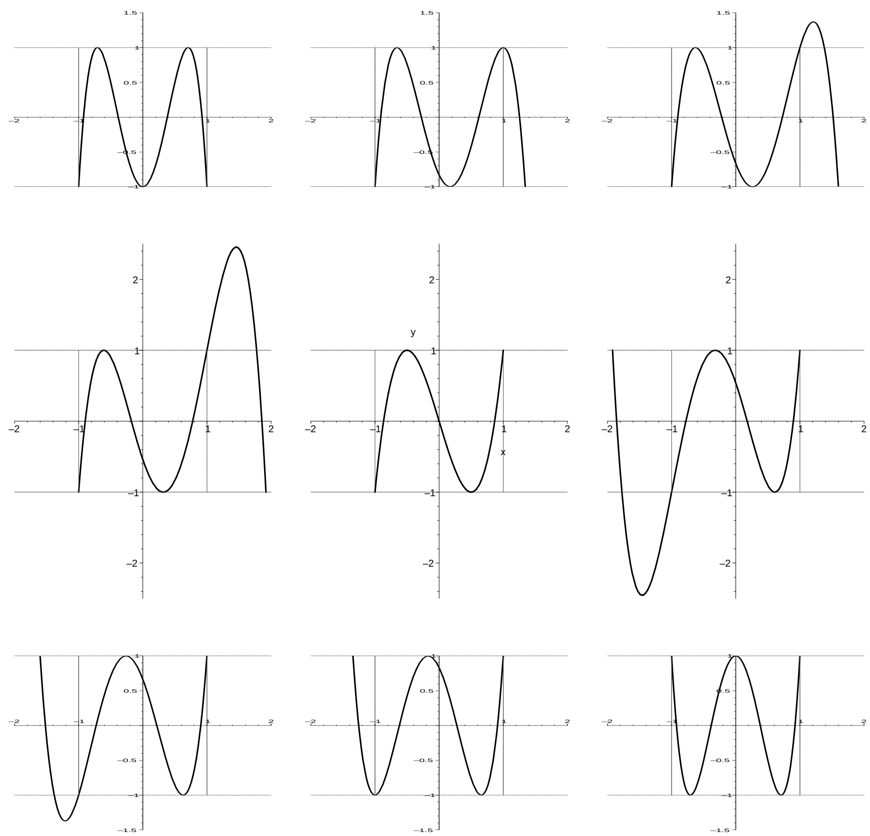
By Chebyshev's result, $\|p^{(n)}\| \le \|T_n^{(n)}\| \, \|p\|$, so the range of the parameter is

$$-2^{n-1} \le \theta \le 2^{n-1}.$$

As θ traverses the interval $[-2^{n-1}, 2^{n-1}]$, Zolotarev polynomials go through the following transformations:

$$-T_n(x) \to -T_n(ax+b) \to Z_n(x,\theta) \to T_{n-1}(x) \to Z_n(x,\theta) \to T_n(cx+d) \to T_n(x)\,.$$

The next figure illustrates it for $n = 4$.



There are many other parametrizations in use. The classical one is based on the definition of Zolotarev polynomial as the polynomial that deviates least from zero among all polynomials of degree n with two leading coefficients fixed:

$$Z_n(x,\sigma) := x^n + \sigma x^{n-1} + p_{n-2}(x) := \arg \min_{q \in \mathcal{P}_{n-2}} \|x^n + \sigma x^{n-1} + q(x)\| .$$

V. Markov used the parametrization with respect to $z \in [-1,1]$, the point where $Z_n^{(k)}(\cdot\,, z)$ attains the value $M_k(z)$ in the pointwise Markov problem.

Zolotarev polynomials subdivide into 3 groups depending on the stucture of the set $\mathcal{A} := (\tau_i)$ of their alternation points.

A) $\mathcal{A}$ contains $n+1$ points: then Z_n is the Chebyshev polynomial T_n.

B) $\mathcal{A}$ contains n points but only one of the endpoints: then Z_n is a stretched Chebyshev polynomial $T_n(ax+b)$, $|a| < 1$.

C) $\mathcal{A}$ contains n points including both endpoints: then Z_n is called a proper Zolotarev polynomial and it is either of degree n, or the Chebyshev polynomial T_{n-1} of degree $n-1$.

For a proper Zolotarev polynomial of degree n there are three points β, γ, δ to either side of $[-1,1]$ such that

$$Z_n'(\beta) = 0, \qquad Z_n(\gamma) = -Z_n(\delta) = \pm 1\,.$$

As functions of $\theta \in [-2^{n-1}, 2^{n-1}]$, the interior alternation points $(\tau_i)_{i=2}^{n-1}$ as well as β, γ, δ are monotonely increasing (the latter three go through the infinity as θ passes the zero), so that any of them may be chosen as a parameter, too.

Theorem 1.9. *For each $z \in [-1,1]$, the value $M_k(z) := \sup_{\|p\| \le 1} |p^{(k)}(z)|$ is attained by a Zolotarev polynomial Z_n. If $Z_n \ne T_n$, then*

$$M_k(z) = |Z_n^{(k)}(z)| \quad \Leftrightarrow \quad R^{(k)}(z) = 0,$$

where $R(x) = \prod_{i=1}^n (x - \tau_i)$.

This result is typical for all Markov-type problems for it says that if

$$Z_n^{(k)}(z, \theta_z) = \sup_\theta Z_n^{(k)}(z, \theta)\,,$$

then either $R^{(k)}(z) := \partial_\theta Z_n^{(k)}(z,\theta) = 0$ or θ_z is the endpoint of the θ-interval.

The structure. The structure of the (proper) Zolotarev polynomials (let alone other Zolotarev-type functions) is rather unknown. Basically, $\{Z_n\}$ satisfy the differential equation

$$1 - y(x)^2 = \frac{(1-x^2)(x-\gamma)(x-\delta)}{n^2(x-\beta)^2}\, y'(x)^2\,,$$

and Zolotarev himself provided implicit formulas for his polynomials in terms of elliptic functions, but explicit expressions for Z_n are known only for $n = 2, 3, 4$. The case $n = 2$ is trivial, and it is quite easy to construct the family $\{Z_3\}$ (it has been done already by A. Markov in 1889, and repeated thereafter in many different forms). But already for $n = 4$ it seems that nobody really believed that an explicit form can be found. As a matter of fact it was, by V. Markov in 1892. Here it is:

$$Z_4(x,t) = \frac{1}{c_0(t)} \sum_{i=0}^4 b_i(t)\, x^i\,, \qquad |t| \le \sqrt{2} - 1\,,$$

where

$$\begin{aligned}
&b_0(t) = 2t\,(-3t^6 - t^4 - t^2 + 1)\,, && b_1(t) = -t^{10} + t^8 + 2t^6 + 10t^4 + 7t^2 - 3\,,\\
&b_2(t) = 2t\,(3t^6 + t^4 + t^2 - 5)\,, && b_3(t) = 4\,(-3t^4 - 2t^2 + 1)\,,\\
&b_4(t) = 8t\,, && c_0(t) = \textstyle\sum b_i(t) = (1-t^2)(1-t^4)^2\,,
\end{aligned} \tag{1.2}$$

with the alternation points

$$\tau_1 = -1 < \tau_2 = \frac{t^3 + t - (1-t^2)}{2} < \tau_3 = \frac{t^3 + t + (1-t^2)}{2} < \tau_4 = 1\,.$$

A. Markov (1889) showed how construction of a Zolotarev polynomial in the form

$$Z_n(x,\beta) = p_0(x-\beta)^n + p_1'(x-\beta)^{n-1} + \cdots + p_{n-2}'(x-\beta)^2 + p_n'$$

can be reduced to two algebraic equations between the unknonws β, γ and δ, so that, theoretically, choosing β as a parameter it is possible to express γ and δ, and then all coeffients p_0 and p_i' in terms of β.

He also showed that Z_n can be found as a solution to a system of linear differential equations of the 1-st order, and another (non-linear) system was suggested by Voronovskaya [56, p. 97]. But as far as we know, nobody (including A. Markov and Voronovskaya themselves) has ever tried to apply these methods for constructing Z_n for any particular n.

Recently, the interest in an explicit algebraic solution of the Zolotarev problem was revived in the papers by Peherstorfer [43], Sodin-Yuditsky [44] and Malyshev [42], but it is only Malyshev who demonstrates how his theory can be applied to some explicit constructions for particular n.

From our side, we notice that there is a simple numerical procedure of constructing a polynomial p_n, say, on $[-1,1]$, with any given values (y_i) of its local maxima, i.e., such that with some $-1 = x_1 < x_2 < \cdots < x_{n+1} = 1$

$$p_n(x_i) = (-1)^i y_i, \quad i = 1..n+1, \qquad p_n'(x_i) = 0, \quad i = 2..n\,.$$

If we choose $y = (1,1,\ldots,1,y_n,1)$, then the resulting polynomial will be a proper Zolotarev polynomial parametrized by the value $Z_n(\beta) = y_n$ and squeezed to the interval $[-1,1]$.

2 Variational approach

2.1 General considerations

Maximizing M_k over the one-parameter family. The following approach is perhaps the only one that can be applied to any problem of the Markov type in the sense that, initially, it does not rely on any particular properties of polynomials or splines or whatsoever. (It is another question whether it will work or not, sometimes it does, sometimes is does not.)

Let $\{Z(x,\theta)\}$ be the one-parameter family of functions that are extremal for the pointwise Markov-type problem, i.e.,

$$M_{k,\mathcal{F}}(z) := \sup_{f\in\mathcal{F}} |f^{(k)}(z)| = |Z^{(k)}(z,\theta_z)| = \sup_{\theta} |Z^{(k)}(z,\theta)|\,.$$

Here we may assume (say, taking $\theta_z = z$) that, under our parametrization,

$$\theta_z \in [\theta_{z=-1}, \theta_{z=1}] =: [-\bar\theta, \bar\theta]\,.$$

Set

$$K(x,\theta) := Z^{(k)}(x,\theta), \qquad (x,\theta) \in [-1,1] \times [-\bar\theta, \bar\theta\,] =: \Omega\,.$$

The following statement is immediate.

Proposition 2.1. *We have*

$$M_{k,\mathcal{F}} = \sup_{z\in[-1,1]} M_{k,\mathcal{F}}(z) = \sup_{x,\theta} K(x,\theta)\,.$$

Now, take

$$T(\cdot) := Z(\cdot, \bar\theta)$$

i.e., T is the function from $\mathcal{F}$ that attains the value $M_k(z)$ at $z = 1$ (an analogue of the Chebyshev polynomial). This is our main candidate for the global solution, so we want to find whether

$$M_{k,\mathcal{F}} = \sup_{x,\theta\,\in\,\Omega} K(x,\theta) \stackrel{?}{=} \|T^{(k)}\|\,. \tag{2.1}$$

(Strictly speaking, we should have defined two functions $T_\pm(x) := Z(x, \pm\bar\theta)$, but they usually differ only in sign, or satisfy $T_-(x) = \pm T_+(-x)$ as in the Landau-Kolmogorov problem.) Notice that, directly from definition,

$$1) \quad \sup_\theta K(\pm 1, \theta) = |T^{(k)}(\pm 1)|\,, \qquad 2) \quad \sup_x K(x, \pm\bar\theta) = \|T^{(k)}\|\,,$$

i.e., on the boundary of the (x,θ)-domain Ω we have

$$\sup_{x,\theta\,\in\,\partial\Omega} K(x,\theta) = \|T^{(k)}\|\,.$$

Therefore, in order to verify (2.1), we have to deal with the following problem.

Problem 2.2. *Find whether*

$$\sup_{x,\theta\,\in\,\Omega} K(x,\theta) = \sup_{x,\theta\,\in\,\partial\Omega} K(x,\theta)\,. \tag{2.2}$$

Checking local extrema. A straightforward approach for attacking this problem is to analyze the interior extremal points of $K = K(x,\theta)$:

$$\partial_x K(x_*, \theta_*) = \partial_\theta K(x_*, \theta_*) = 0.$$

If at every such point the strict inequality

$$d := (\partial_{xx} K)(\partial_{\theta\theta} K) - (\partial_{x\theta} K)^2 < 0 \tag{2.3}$$

is valid, then (x_*, θ_*) is a saddle point, hence $|K|$ has no local maxima in the interior of domain, and therefore (2.2), hence (2.1), are true.

We mention that it makes sense to consider only those (x_*, θ_*), where the univariate functions $|K(\cdot, \theta)|$ and $|K(x, \cdot)|$ have local maxima in x and in θ respectively, i.e. such that

$$\operatorname{sgn} \partial_{xx} K = \operatorname{sgn} \partial_{\theta\theta} K = -\operatorname{sgn} K,$$

therefore the above inequality (2.3) is not trivial.

Since $K(x, \theta) := Z^{(k)}(x, \theta)$, the corresponding derivatives become

$$\partial_x K := Z^{(k+1)}, \quad \partial_\theta K := Z_\theta^{(k)},$$

and

$$\partial_{xx} K := Z^{(k+2)}, \quad \partial_{x\theta} K := Z_\theta^{(k+1)}, \quad \partial_{\theta\theta} K := Z_{\theta\theta}^{(k)},$$

so that one needs to check whether, for a given one-parameter family of functions $Z := Z(\cdot, \theta)$, the equality

$$Z^{(k+1)}(z) = Z_\theta^{(k)}(z) = 0$$

implies

$$d := Z^{(k+2)}(z) Z_{\theta\theta}^{(k)}(z) - [Z_\theta^{(k+1)}(z)]^2 < 0\,. \tag{2.4}$$

The only problem is that, as has been mentioned, there are no explicit expressions for Zolotarev polynomials or Zolotarev-type functions.

Comment 2.3. V. Markov's original approach (repeated later in [17] and [41]) had a slightly different form. Namely, he studied interior extrema of the univariate (positive or negative) function $M_k(x) = Z^{(k)}(x, \theta_x)$. In this case, if the following implication is true

$$M_k'(z) = 0 \quad \Rightarrow \quad M_k(z) M_k''(z) > 0\,, \tag{2.5}$$

then $|M_k(\cdot)|$ takes at $x = z$ a locally minimal value, and hence the global maximum of $|M_k(\cdot)|$ is attained by a polynomial (or alike) other than the Zolotarev one. In fact, (2.5) is equivalent to (2.4) for one can show [41] that the equality

$$M_k(x) := \sup_\theta Z^{(k)}(x, \theta) =: Z^{(k)}(x, \theta_x)$$

implies

$$\widehat{d} := M_k(z) M_k''(z) = \frac{Z^{(k)}(z)}{Z_{\theta\theta}^{(k)}(z)} \left(Z^{(k+2)}(z) \cdot Z_{\theta\theta}^{(k)}(z) - [Z^{(k+1)}(z)]^2 \right)\,.$$

At the point z where $Z_\theta^{(k)}(z) = 0$, the numerator and the denominator are of opposite sign, hence, $\widehat{d} > 0$ is equivalent to $d < 0$.

2.2 V. Markov's original proof

Here we show how the variational approach just described works for the Markov problem, where the extremal set for the pointwise problem consists of Zolotarev polynomials.

Let $\{Z(\cdot,\theta)\} \subset \mathcal{P}_n$ be the family of proper Zolotarev polynomials of degree n with n equioscillation points

$$-1 = \tau_1 < \tau_2 < \cdots < \tau_{n-1} < \tau_n = 1, \quad \tau_i = \tau_i(\theta)$$

such that

$$Z(1) = \|Z\| = 1, \quad Z'(\beta) = 0, \quad |\beta| = |\beta(\theta)| > 1,$$

$$Z(x,\theta) = \sum_{i=0}^{n} a_i(\theta)x^i\,, \quad a_n(\theta) = \theta \neq 0\,. \tag{2.6}$$

The following theorem is the central achievement of V. Markov's original work [7].

Theorem 2.4 (V. Markov (1892)). *If at some point* $(x,\theta) = (z,\theta_z)$

$$Z^{(k+1)}(z) = Z_\theta^{(k)}(z) = 0, \tag{2.7}$$

then

$$d := Z^{(k+2)}(z)Z_{\theta\theta}^{(k)}(z) - [Z_\theta^{(k+1)}(z)]^2 < 0\,. \tag{2.8}$$

To prove this theorem V. Markov established very fine relations between the functions involved in (2.8). Here they are.

Lemma 2.5. *For all* (x,θ)*, we have*

$$Z_\theta(x) = \prod_{i=1}^{n}(x-\tau_i) =: R(x). \tag{2.9}$$

Proof. First of all, it follows from (2.6) that $Z_\theta(x) = x^n + q_{n-1}(x,\theta)$, i.e. Z_θ is a polynomial in x with the leading coefficient equal to 1. As to its roots, differentiating the identity $Z(\tau_i) \equiv Z(\tau_i(\theta),\theta) \equiv \pm 1$ we obtain

$$Z'(\tau_i)\cdot\tau_i'(\theta) + Z_\theta(\tau_i) = Z_\theta(\tau_i) = 0, \qquad i = 1..n\,. \qquad \square$$

The next formula provides a basic relation between $Z_\theta = R$ and $Z_x = Z'$, and is decisive in further considerations.

Lemma 2.6. *For all* (x,θ)*, we have*

$$na_n\,(x-\beta)\,R(x) = (x^2-1)Z'(x)\,. \tag{2.10}$$

Proof. Both sides, as polynomials in x, have the same roots and the same leading coefficients. $\square$

Finally, an expression for $Z_{\theta\theta}$.

Lemma 2.7. *For all* (x,θ)*, we have*

$$na_n Z_{\theta\theta}(x) = -nR(x) + (x+\beta)R'(x) + (\beta^2-1)\,\psi(x)\,, \tag{2.11}$$

where

$$(x-\beta)\,\psi(x) = R'(x) - \tfrac{R'(\beta)}{R(\beta)}R(x)\,, \quad \psi \in \mathcal{P}_{n-1}\,. \tag{2.12}$$

Proof. Differentiating the identity (2.10) with respect to θ, and using (2.9) and the fact that $a_n'(\theta) = 1$, we obtain

$$\begin{aligned} &n(x-\beta)R(x) - na_n\beta'(\theta)R(x) + na_n(x-\beta)R_\theta(x) \\ &\qquad = (x^2-1)R'(x) = (x^2-\beta^2)R'(x) + (\beta^2-1)R'(x)\,, \end{aligned}$$

and division by $(x-\beta)$ and rearrangement of the terms gives

$$na_nR_\theta(x) = -nR(x) + (x+\beta)R'(x) + \tfrac{\beta^2-1}{x-\beta}R'(x) + \tfrac{na_n\beta'(\theta)}{x-\beta}R(x)\,.$$

Putting $x=\beta$ in the first equality provides $na_n\beta'(\theta) = -(\beta^2-1)\frac{R'(\beta)}{R(\beta)}$, so

$$na_nR_\theta(x) = -nR(x) + (x+\beta)R'(x) + \tfrac{\beta^2-1}{x-\beta}\left[R'(x) - \tfrac{R'(\beta)}{R(\beta)}R(x)\right].$$

In the square brackets, we have a polynomial of degree n that vanishes at $x=\beta$, hence it is of the form $(x-\beta)\psi(x)$, where $\psi \in \mathcal{P}_{n-1}$. $\square$

Proof of Theorem 2.4. We assume that

$$a_n < 0, \quad \text{hence} \quad \beta > 1, \quad \text{thus} \quad z-\beta < 0 \quad \text{if} \quad z \in [-1,1]\,. \tag{2.13}$$

We also assume that (at the point z where $Z^{(k+1)}(z) = 0$)

$$Z^{(k)}(z) > 0, \quad \text{hence} \quad Z^{(k+2)}(z) < 0\,. \tag{2.14}$$

Under these assumptions (and assumptions (2.7) of the theorem) we will show that

$$Z_\theta^{(k+1)}(z) \;>\; \tfrac{z^2-1}{na_n(z-\beta)}\,Z^{(k+2)}(z) > 0\,, \tag{2.15}$$

$$Z_\theta^{(k+1)}(z) \;>\; \tfrac{na_n(z-\beta)}{z^2-1}\,Z_{\theta\theta}^{(k)}(z)\,, \tag{2.16}$$

and that clearly proves the theorem.

1) Our starting point is again the identity (2.10)

$$na_n(x-\beta)R(x) = (x^2-1)Z'(x)\,.$$

Differentiating it $(k+1)$ times with respect to x and setting $(x,\theta)=(z,\theta_z)$ we obtain (taking into account (2.7))

$$na_n(z-\beta)\,R^{(k+1)}(z) = (z^2-1)\,Z^{(k+2)}(z) + k(k+1)\,Z^{(k)}(z)\,. \qquad (2.17)$$

Both terms on the right-hand side are positive, and also $na_n(z-\beta)>0$, so

$$R^{(k+1)}(z) > \tfrac{z^2-1}{na_n(z-\beta)}\,Z^{(k+2)}(z) > 0\,, \qquad (2.18)$$

which proves (2.15).

2a) Now we turn to (2.16). From (2.11) and (2.7), we have

$$na_n Z_{\theta\theta}^{(k)}(z) = (z+\beta)\,R^{(k+1)}(z) + (\beta^2-1)\,\psi^{(k)}(z)\,,$$

and from (2.12) and (2.7) we find $(z-\beta)\psi^{(k)}(z)+k\psi^{(k-1)}(z) = R^{(k+1)}(z)$, i.e.

$$\psi^{(k)}(z) = \frac{1}{z-\beta}\,[R^{(k+1)}(z) - k\psi^{(k-1)}(z)]\,,$$

so, putting this expression into the previous one, we obtain

$$\begin{aligned} na_n Z_{\theta\theta}^{(k)}(z) &= (z+\beta)\,R^{(k+1)}(z) + \tfrac{\beta^2-1}{z-\beta}[R^{(k+1)}(z) - k\psi^{(k-1)}(z)] \\ &= \tfrac{z^2-1}{z-\beta}R^{(k+1)}(z) - \tfrac{k(\beta^2-1)}{z-\beta}\psi^{(k-1)}(z)\,. \end{aligned}$$

Hence

$$R^{(k+1)} - \tfrac{na_n(z-\beta)}{z^2-1}Z_{\theta\theta}^{(k)}(z) = \tfrac{k(\beta^2-1)}{z^2-1}\,\psi^{(k-1)}(z)\,, \qquad (2.19)$$

and since $\frac{k(\beta^2-1)}{z^2-1}<0$, it follows that

$$(2.16) \quad \Leftrightarrow \quad \psi^{(k-1)}(z) < 0\,.$$

2b) Consider relation (2.12) for ψ:

$$(x-\beta)\psi(x) = R'(x) - \tfrac{R'(\beta)}{R(\beta)}R(x)\,.$$

For $x\in[-1,1]$, since $\beta>1$, both factors $(x-\beta)$ and $-\frac{R'(\beta)}{R(\beta)}$ are negative, hence at the zeros of R' we have

$$R'(t_i)=0 \quad \Rightarrow \quad \operatorname{sgn}\psi(t_i) = \operatorname{sgn}R(t_i)\,.$$

This means that the zeros of the polynomials ψ and R' interlace, thus, by what we know now as the Markov interlacing property,

$$R^{(k)}(z)=0 \quad \Rightarrow \quad \operatorname{sgn}\psi^{(k-1)}(z) = \operatorname{sgn}R^{(k-1)}(z)\,.$$

At the points where $R^{(k)}(z)=0$ we have $\operatorname{sgn}R^{(k-1)}(z) = -\operatorname{sgn}R^{(k+1)}(z)$, hence

$$\operatorname{sgn}\psi^{(k-1)}(z) = -\operatorname{sgn}R^{(k+1)}(z) < 0\,,$$

the last inequality by (2.18). $\square$

Comment 2.8. Compared with Markov's proof, we split the inequality (2.8) into two parts (2.15)-(2.16), and made one more simplifying assumption (2.14). We also got rid of expressions for $\tau_i'(\theta)$ and $\theta'(z)$ that were involved in Markov's arguments.

From relations (2.17) and (2.19), we find

$$\begin{aligned} R^{(k+1)}(z) &= \tfrac{(z^2-1)}{na_n(z-\beta)}\, Z^{(k+2)}(z) + \tfrac{k(k+1)}{na_n(z-\beta)}\, Z^{(k)}(z)\,, \\ R^{(k+1)}(z) &= \tfrac{na_n(z-\beta)}{z^2-1}\, Z^{(k)}_{\theta\theta}(z) + \tfrac{k(\beta^2-1)}{z^2-1}\, \psi^{(k-1)}(z)\,, \end{aligned}$$

and we can derive the exact expressions for d in (2.8)

$$\begin{aligned} -d &= [R^{(k+1)}(z)]^2 - Z^{(k+2)}(z) Z^{(k)}_{\theta\theta}(z) \\ &= \tfrac{k(\beta^2-1)}{na_n(z-\beta)}\, Z^{(k+2)}(z)\, \psi^{(k-1)}(z) + \tfrac{k(k+1)}{na_n(z-\beta)}\, Z^{(k)}(z)\, R^{(k+1)}(z) \qquad (2.20) \\ &= \frac{kZ^{(k)}(z)}{z-\beta}\, \frac{k+1 + (\beta^2-1)\, \frac{\psi^{(k-1)}(z)}{R^{(k+1)}(z)}\, \frac{Z^{(k+2)}(z)}{Z^{(k)}(z)}}{-\frac{na_n}{R^{(k+1)}(z)}}\,. \qquad (2.21) \end{aligned}$$

The last one is formula (118) of Markov's work, and he finished his proof by analyzing its sign.

Comment 2.9. An interesting fact is that, as V. Markov himself wrote in "Appendix to §34" (which was omitted in the German translation), he found the proof of Theorem 2.4 at the very last moment, when his article was already in print. Until then he had proofs of the inequality $\|p^{(k)}\| \le \|T_n^{(k)}\|\,\|p\|$ only in the cases

$$k = 1, \quad k = 2, \quad k = n-2, \quad k = n-1\,,$$

each time a different one. (He added an "Appendix" to demonstrate these proofs; they are quite interesting, by the way.)

2.3 A brief account of V. Markov's results

Markov's Theorem 2.4 (with preliminaries) reads as follows:

A) For each $z \in [-1,1]$ the value

$$M_k(z) := \sup_{\|p\| \le 1} |p^{(k)}(z)|$$

is attained either by a proper Zolotarev polynomial $Z_n(\cdot,\theta)$, or by the Chebyshev polynomial T_n, or by a transformed Chebyshev polynomial $\widetilde{T}_n(x) = T_n(ax+b)$, or by the Chebyshev polynomial T_{n-1}.

B) If a local extreme value of the (positive) function $M_k(\cdot)$ is attained by a proper Zolotarev polynomial of degree n, then it is a local minimum.

C) Hence,

$$M_k = \sup_z M_k(z) = \max\{\, \|T_n^{(k)}\|, \|\widetilde{T}_n^{(k)}\|, \|T_{n-1}^{(k)}\|, M_k(\pm 1)\,\}\,,$$

and it is not difficult to show that the last maximum is equal to $T_n^{(k)}(1)$.

Theorem 2.10 (V. Markov (1892)). *For all n, k we have*

$$\sup_{\|p\|\le 1} \|p^{(k)}\| = T_n^{(k)}(1)\,.$$

Actually, in his opus, V. Markov made a very detailed investigation of the character of the value $M_k(z)$ when z runs through certain subintervals.

0) Given k, define the points (ξ_i) and (η_i) by

$$\eta_0 := -1, \quad [(x-1)\,T_n'(x)]^{(k)} =: c_k \prod_{i=1}^{n-k}(x-\eta_i), \tag{2.22}$$

$$[(x+1)\,T_n'(x)]^{(k)} =: c_k \prod_{i=1}^{n-k}(x-\xi_i), \quad \xi_{n-k+1} =: 1. \tag{2.23}$$

Then $\eta_{i-1} < \xi_i < \eta_i$, and we define (following Voronovskaya [56])

$$\begin{array}{rl} \text{Chebyshev intervals} & e_i^T := [\eta_{i-1}, \xi_i]\,, \\ \text{Zolotarev intervals} & e_i^Z := (\xi_i, \eta_i)\,, \end{array}$$

so that the interval $[-1, 1]$ is split in the following way

$$\begin{array}{ccccccc} \text{Chebyshev} & & \text{Chebyshev} & & & & \text{Chebyshev} \\ \text{interval} & & \text{interval} & & & & \text{interval} \\ \downarrow & & \downarrow & & & & \downarrow \\ (-1, \xi_1] & (\xi_1, \eta_1) & [\eta_1, \xi_2] & (\xi_2, \eta_2) & \cdots & (\xi_{n-k}, \eta_{n-k}) & [\eta_{n-k}, 1) \\ & \uparrow & & \uparrow & & \uparrow & \\ & \text{Zolotarev} & & \text{Zolotarev} & & \text{Zolotarev} & \\ & \text{interval} & & \text{interval} & & \text{interval} & \end{array}$$

1) If z belongs to a Chebyshev interval, then

$$M_k(z) = |T_n^{(k)}(z)|, \quad z \in e_i^T\,.$$

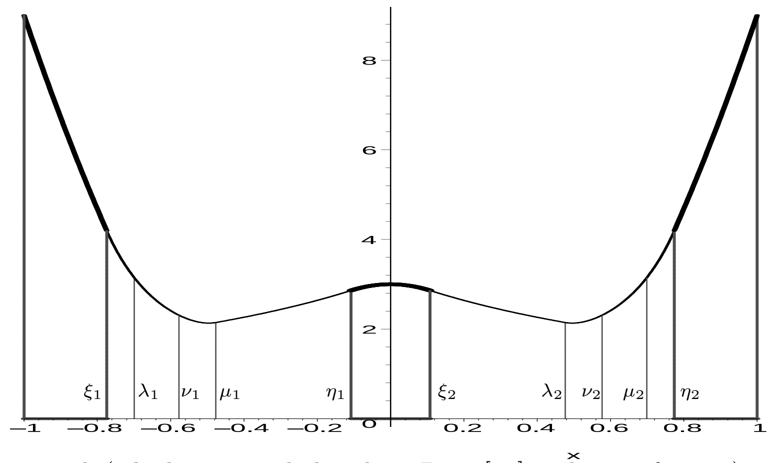
Moreover, the Chebyshev intervals contain the roots of $T_n^{(k+1)}$ (and, as a matter of interest, those of $T_n^{(k-1)}$), i.e., the *local maxima of $M_k(\cdot)$ and $|T_n^{(k)}|$ coincide.*

2) If z belongs to a Zolotarev interval e_i^Z, then the value of $M_k(z)$ is achieved either by a proper Zolotarev polynomial $Z_n(\cdot, \theta_z)$, or by a transformed Chebyshev polynomial $T_n(a_z x + b_z)$, or by the Chebyshev polynomial T_{n-1}, each time on a certain subintervals as illustrated below.

$$\begin{array}{rccccccc} & \overbrace{\text{Chebyshev interval}} & & & \overbrace{\text{Zolotarev interval}} & & & \overbrace{\text{Chebyshev interval}} \\ z \to & [\eta_{i-1}, \xi_i] & (\xi_i, \lambda_i] & (\lambda_i, \nu_i) & \nu_i & (\nu_i, \mu_i) & [\mu_i, \eta_i) & [\eta_i, \xi_{i+1}] \\ & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \text{Extr. pol.} \to & -T_n(x) & -T_n(a_z x + b_z) & Z_n(x, \theta_z) & T_{n-1}(x) & Z_n(x, \theta_z) & T_n(c_z x + d_z) & T_n(x) \\ & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ M_k(z) \to & |T_n^{(k)}(z)| & \frac{|T_n^{(k)}(\xi_i)|(1+\xi_i)^k}{(1+z)^k} & |Z_n^{(k)}(z, \theta_z)| & |T_{n-1}^{(k)}(\nu_i)| & |Z_n^{(k)}(z, \theta_z)| & \frac{|T_n^{(k)}(\eta_i)|(1-\eta_i)^k}{(1-z)^k} & |T_n^{(k)}(z)| \end{array}$$

Notice the exact behaviour of $M_k(\cdot)$ as a hyperbolic function $\frac{c}{(1\pm z)^k}$ on the intervals (ξ_i, λ_i) and (μ_i, η_i), where the extremal functions are transformed Chebyshev polynomials.

3) The next figure represents the graph of $M_k(\cdot)$ for the case of cubic polynomials ($n = 3$) and the first derivative ($k = 1$). Bold are the parts where the value is achieved by the Chebyshev polynomial $T_3(x) = 4x^3 - 3x$.



This graph (which appeared already in Boas [13] without reference) is based on the exact expressions for the functions involved computed by A. Markov [51] in 1889. Here they are (for the interval $[0, 1]$):

$$n = 3, \quad M_1(x) = \begin{cases} 3(1-4x^2), & x \in [0,\xi], \quad \xi = \frac{\sqrt{7}-2}{6}\,; \\ \frac{7\sqrt{7}+10}{9(1+x)}, & x \in [\xi,\lambda], \quad \lambda = \frac{2\sqrt{7}-1}{9}\,; \\ \frac{16x^3}{(9x^2-1)(1-x^2)}, & x \in [\lambda,\mu], \quad \mu = \frac{2\sqrt{7}+1}{9}\,; \\ \frac{7\sqrt{7}-10}{9(1-x)}, & x \in [\mu,\eta], \quad \eta = \frac{\sqrt{7}+2}{6}\,; \\ 3(4x^2-1), & x \in [\eta,1]\,. \end{cases}$$

A. Markov also provided the formula of $M_1(\cdot)$ for $n = 2$, and later, while studing the case $k > 1$, V. Markov found for $n = 3$ an exact analytic form of $M_2(\cdot)$ ($M_3(\cdot)$ is a constant). Using his expression for Z_4 (see (1.2)) it is possible to find all $M_k(\cdot)$ for $n = 4$.

4) Inside each Zolotarev interval, there is exactly one local minimum of $M_k(\cdot)$, say, at $x = \sigma_i$. A naive conjecture that $\sigma_i = \nu_i$, i.e., that these local minima $M_k(\sigma_i)$ are attained by the Chebyshev polynomial T_{n-1} is not true (as seen from the graph). V. Markov proved that this could happen only in the middle of the interval:

$$\text{a) if } \nu_i = 0, \text{ then } \sigma_i = 0\,,$$

otherwise

$$\text{b) if } \nu_i > 0, \text{ then } \sigma_i \in (\lambda_i, \nu_i), \quad \text{c) if } \nu_i < 0, \text{ then } \sigma_i \in (\nu_i, \mu_i).$$

5) In 1961, Gusev [6] provided two supplements to V. Markov's results. Firstly, he showed that while the first derivative $M_k'(\cdot)$ is continuous on $[-1,1]$ (which is rather clear and was used by V. Markov), the second derivative $M_k''(\cdot)$ has jumps at the points ξ, λ, μ, η (but not at ν) where Zolotarev polynomials change from one type to another.

His second and quite interesting observation was about the measure of Chebyshev and Zolotarev intervals, namely

$$\operatorname{mes}(e^T) = 2\,\frac{k}{n}\,, \quad \operatorname{mes}(e^Z) = 2\,\frac{n-k}{n}\,.$$

The proof is quite elementary, so we give it here. By definition,

$$\operatorname{mes}(e^Z) = \sum(\eta_i - \xi_i) = \sum \eta_i - \sum \xi_i\,,$$

where

$$p(x) := c\prod_{i=1}^{n-k}(x-\eta_i) := [(x-1)\,T_n'(x)]^{(k)}\,, \quad c\prod_{i=1}^{n-k}(x-\xi_i) := [(x+1)\,T_n'(x)]^{(k)}\,.$$

Then $\frac{1}{n-k}\sum \eta_i$ is the only root of the polynomial $p^{(n-k-1)}$ which is the polynomial $[(x-1)\,T_n'(x)]^{(n-1)}$, which has the only root $\frac{1}{n}\,[1+\sum_{i=1}^{n-1}\zeta_i]$, i.e.,

$$\sum \eta_i = \tfrac{n-k}{n}\,[1+\textstyle\sum_{i=1}^{n-1}\zeta_i] \quad (\text{where } T_n'(\zeta_i)=0).$$

Similarly, $\sum \xi_i = \frac{n-k}{n}[-1+\sum_{i=1}^{n-1}\zeta_i]$, hence the result.

6) We mention that V. Markov's results for general k were essentially of the same type as earlier results of A. Markov for the case $k=1$. Precisely, for the pointwise problem for the 1-st derivative, A. Markov showed that Zolotarev polynomials form the extremal set, proved that the value $M_1(z)$ is attained by either type of these polynomials when z belongs to certain intervals, and described the behaviour of $M_1(\cdot)$ on these intervals exactly in the same way as it is given in the cases 1)-3) of this section.

He did not get the result about the minima of $M_1(\cdot)$ as in case 4) (which was the main achievement of his kid brother), but he proved the global inequality $\|p'\| \le n^2\,\|p\|$ using what we call now Bernstein's majorant (see §4.1 for details of his proof).

2.4 Works of Voronovskaya and Gusev

Works of Voronovskaya. Voronovskaya is perhaps best known by her saturation estimate for the Bernstein polynomials,

$$B_n(f,x) - f(x) = \frac{x(1-x)}{2n^2}\,f''(x) + o(n^2)\,.$$

However, most of her studies were on extremal properties of polynomials, which she summarized in her book "The functional method and its application" [56]. Boas was very enthusiastic about Voronovskaya works. He translated her book into English in 1970, and, in his two surveys [13]-[14], made a very delightful report about her results "[which solved] a great variety of extremal problems that had previously seemed too difficult for anyone to do anything with".

In particular, Boas attributes to Voronovskaya the solution of the "point-by-point" Markov problem (for the 1-st derivative). The latter is not correct. It is true that her 1959 paper "The functional of the first derivative and improvement of a theorem of A. A. Markov" [55] does improve upon some results of A. Markov (1889). But the whole truth is that this improvement (it is about the minima of $M_1(\cdot)$) can be found in V. Markov (1892). It is only her arguments (for $k = 1$) that are a bit different (and simpler) than those of V. Markov (for general k), but the results are the same.

In this respect, astonishing is her final remark: "But neither A. A. Markov nor V. A. Markov, in studying the question of a bound for the derivatives at interior points of the fundamental interval, took advantage of the use of the Zolotarev polynomials [A. Markov, p. 64] and [V. Markov, p. 55], and hence they could not carry the problem to completion."

Since it suffices to take a brief look through either of Markov's papers in order to find that Zolotarev polynomials occupy the central place in both articles, it is all the more interesting to look at the pages pointed out by Voronovskaya. Here are the exact quotations (about the only thing they did not want to use):

A. Markov [p. 64]: "Without relying on E. I. Zolotarev's formulas, we show how it is possible to reduce our problem to three algebraic equations."

V. Markov [p. 55]: "We notice that Zolotarev in his paper expressed the solution of the equation in terms of elliptic functions, but we will not focus on that."

The only explanation for this story that I can think of is that Voronovskaya – like most of us – never read either of Markov's articles, and had no idea about their actual content. So, when her paper was about to be published, and somebody advised her to take a closer look at these works, she did not find the courage to admit that she simply rediscovered the results already 70 years old. Just another illustration of Boas' words about A. Markov's paper as "one of the most often cited, and one of the least read".

Gusev's paper. V. A. Gusev begins his paper [6] in a quite remarkable way. He is going "to study the problem considerably more completely than in Bernstein and in Duffin-Shaeffer, and in a considerably shorter way than in V. Markov". The logic of this sentence leaves open the possibility that his way is *not* shorter than those of Bernstein and Duffin-Schaeffer, and that it gives *not* more complete results than those of Markov. And this is true! (well, almost: he proved two supplementary results, as we have seen). More than this, Gusev's proof of Markov inequality is not new, it is essentially a reproduction of Markov's original proof.

There are some differences in the preliminaries, because V. Markov uses his own criterion for the norm of linear functional, while Gusev uses that of Voronovskaya (of course, both are equivalent).

But the very essence of V. Markov's treatise, the proof that

$$Z^{(k+2)}(z)Z^{(k)}_{\theta\theta}(z) - [Z^{(k+1)}_{\theta}(z)]^2 < 0,$$

hence a local extremum of $M_k(\cdot)$ if attained by a proper Zolotarev polynomial is a local minimum, hence the Markov inequality, is reproduced by Gusev almost without alterations.

"A way considerably shorter than in V. Markov" is a slight exaggeration too, especially when you find that Gusev uses without proof some of Markov's lemmas sending the reader for those to Markov's paper.

There is, however, a positive side of Gusev's paper (as well as of Voronovskaya), namely a clear and short exposition of Markov's results (provided moreover with an English translation). V. Markov's paper is rather mosaic and archaic, and this makes it a difficult (albeit pleasurable) read. Gusev squeezed it to a small set of clear theorems which give a clear picture of behaviour of the exact upper bound $M_k(\cdot)$. To a certain extent, we followed his exposition in §2.3.

2.5 Similar results

V. Markov's variational approach, based on verifying the inequality

$$d := Z^{(k+2)}(z)Z^{(k)}_{\theta\theta}(z) - [Z^{(k+1)}_{\theta}(z)]^2 < 0$$

for the one-parameter family $Z(x,\theta)$ of Zolotarev-type functions, was used in solution of two other problems of Markov type.

Theorem 2.11 (Pierre-Rahman (1976)). *For the Markov problem with the majorant*

$$\mu(x) = (1-x)^{m_1/2}(1+x)^{m_2/2}, \quad k \ge \tfrac{m_1+m_2}{2}\,,$$

we have

$$M_{k,\mu} := \sup_{|p(x)|\le\mu(x)} \|p^{(k)}\| = \max\left(\|\omega_n^{(k)}\|\,, \|\omega_{n-1}^{(k)}\|\right) \tag{2.24}$$

where $\omega_n \in \mathcal{P}_n$ is the polynomial oscillating most between $\pm\mu$.

The proof is the exact reproduction of Markov's arguments, but on a much more complicated technical level. In our notations, their final expression (which is the last equality on p. 728) has the form

$$d = \frac{Z^{(k)}(z)}{\beta - z}\left\{k(\beta^2-1)\,\frac{\psi^{(k-1)}(z)}{R^{(k+1)}(z)}\,\frac{Z^{(k+2)}(z)}{Z^{(k)}(z)} + (k+1)\left(k - \frac{m_1+m_2}{2}\right)\right\}\frac{R^{(k+1)}(z)}{na_n}\,,$$

just to compare with formula (2.21) of V. Markov.

For some reasons, Pierre & Rahman did not analyze when the maximum in (2.24) is attained by $\omega_n^{(k)}$. It seems to be so if $k > \frac{m_1+m_2}{2}$ (when it looks that $\|\omega_n^{(k)}\| = \omega_n^{(k)}(1)$).

Theorem 2.12 (Shadrin (1995)). *For the Lagrange interpolation problem on a knot-sequence $\delta = (t_i)_{i=0}^n$, we have*

$$M_{k,\delta} := \sup_{\|f^{(n+1)}\| \le 1} \|f^{(k)} - \ell_\delta^{(k)}\| = \frac{1}{(n+1)!}\|\omega_\delta^{(k)}\|,$$

where $\omega_\delta(x) := \prod_{i=0}^n (x - t_i)$.

Here, the one-parameter family $Z(x,\theta)$ consists of perfect splines with at most one knot, and details of the proof are quite different from that of Markov.

However, for the pointwise problem, there are complete analogues of the Chebyshev and Zolotarev intervals

$$e_j^T = (\eta_{j-1}, \xi_j), \quad e_j^Z = [\xi_j, \eta_j]\,.$$

Here, the endpoints of the intervals are defined via $\omega_i(x) := \frac{\omega(x)}{x-t_i}$ as

$$\begin{aligned} \eta_0 := t_1, \quad & \omega_0^{(k)}(x) =: c\prod_{j=1}^{n-k}(x - \eta_j), \\ & \omega_n^{(k)}(x) =: c\prod_{j=1}^{n-k}(x - \xi_j), \quad \xi_{n-k+1} := t_n. \end{aligned}$$

But now, it is Zolotarev intervals where $M_{k,\delta}$ and $\omega_\delta^{(k)}$ (and their local maxima) coincide:

$$M_{k,\delta}(z) := \sup_{\|f^{(n+1)}\| \le 1} |f^{(k)}(z) - \ell_\delta^{(k)}(z)| = \frac{1}{(n+1)!}|\omega_\delta^{(k)}(z)|, \quad z \in e_\delta^Z\,.$$

This pointwise estimate is due to Kallioniemi [29] who also generalized Gusev's result:

$$\mathrm{mes}\,(e_\delta^T) = \frac{k}{n}\,(t_n - t_0).$$

3 "Small-o" arguments

3.1 "Small-o" proofs of Bernstein and Tikhomirov

In 1938, in the less-known and nowadays hardly accessible "Proceedings of the Leningrad Industrial Institute", Bernstein published the article [1] where he "found it not unnecessary to point out another and simpler proof" of V. Markov's inequality. This article was reprinted in 1952 in his Collected Works, and since 1996 its English translation, thanks to Bojanov, is also available.

The proof we are going now to present is, in fact, not that of Bernstein but a mixture from different sources with the main part due to Tikhomirov, as it is given in his exposition [12, pp.111-113] for $k = 1$ (with our straightforward extension to any k). For preliminaries (where Tikhomirov used calculus of variations), we chose the more classical (and elementary) approach of Bernstein and Markov.

This is a promised "book-proof" on 4 pages, so we start from the very very beginning pretending we forgot everything dicussed before.

Book-proof. We are going to study the behaviour of the upper bounds of the k-th derivative of algebraic polynomials

$$\begin{aligned} M_k(z) &:= \sup_{\|p\|\le 1} p^{(k)}(z), \quad z \in [-1,1]\,, \\ M_k &:= \sup_{\|p\|\le 1} \|p^{(k)}\| = \sup_{z\in[-1,1]} M_k(z). \end{aligned}$$

We are going to prove that

$$M_k = \|T_n^{(k)}\| \tag{3.1}$$

by showing that, among all the polynomials p_* that are extremal for $M_k(z)$ for different z, only T_n can hope to achieve the global maximum of $M_k(z)$.

This will be done in two steps.

1) For $z = \pm 1$, we will show that $p_* = T_n$.

2) For $z \in (-1,1)$ we will show that if $p_* \ne T_n$ and

$$M_k(z) = p_*^{(k)}(z), \quad M_k'(z) = 0 \quad \big(= p_*^{(k+1)}(z)\,\big)\,,$$

then there exists a polynomial $P_\lambda \in \mathcal{P}_n$ such that, for some z_λ,

$$\|P_\lambda\| = \|p_*\| - \mathcal{O}(\lambda^2), \quad P_\lambda^{(k)}(z_\lambda) = p_*^{(k)}(z) + o(\lambda^2),$$

so that, for λ small enough,

$$M_k(z_\lambda) \ge \frac{P_\lambda^{(k)}(z_\lambda)}{\|P_\lambda\|} > \frac{p_*^{(k)}(z)}{\|p_*\|} = M_k(z)\,.$$

The latter means that the local extrema of $M_k(z)$ if attained by polynomials other than T_n are local minima, hence all local maxima of $M_k(z)$ are attained by the Chebyshev polynomial, hence the conclusion (3.1).

We start with some characterizations of the extremal polynomials.

Lemma 3.1. *Let*

$$M_k(z) := \sup_{p\in\mathcal{P}_n} \frac{p^{(k)}(z)}{\|p\|} = \frac{p_*^{(k)}(z)}{\|p_*\|}\,,$$

and let $\{\tau_i\}_{i=1}^m$ *be the set of all points for which* $|p^*(x)| = \|p^*\|$. *Then there is no polynomial* $q \in \mathcal{P}_n$ *such that*

$$q^{(k)}(z) = 0 \quad \textit{and} \quad q(\tau_i)p_*(\tau_i) < 0\,. \tag{3.2}$$

Proof. If there is such a q, then the polynomial $r := p^* + \lambda q$ will satisfy $r^{(k)}(z) = p_*^{(k)}(z)$ and $\|r\| < \|p_*\|$, a contradiction to the extremality of p_*. $\square$

Lemma 3.2. *Let $y_*, z \in (-1,1)$ and $(y_i)_{i=1}^{n-2} \in \mathbb{R}$. Then there is a unique polynomial $q \in \mathcal{P}_n$ such that*

$$q(y_i) = 0, \quad q^{(k)}(z) = q^{(k+1)}(z) = 0, \quad q(y_*) = 1,$$

and it changes its sign exactly at the points y_i.

Proof. It follows easily from Rolle's theorem that the homogeneous interpolation problem has only the trivial solution, hence existence of such a q. It also implies the sign pattern, since if there were a point x_* besides (y_i) where q vanishes, then the homogeneous problem with $y_* = x_*$ would have had a non-zero solution.

Lemma 3.3. *Let p_* be a polynomial extremal for $M_k(z)$. Then it has at least n points (τ_i) of alternation between ± 1.*

Proof. Let m be the number of alternations and let $(\tau_i)_{i=1}^m$ be the points such that

$$p_*(\tau_i) = -p_*(\tau_{i+1}) = \epsilon\,\|p_*\|, \quad |\epsilon| = 1.$$

If $m \le n-1$, then adding arbitrary $(\tau_j)_{j=m+1}^{n-1}$ with $|\tau_j| > 1$ to the list, we can apply Lemma 3.2 to construct the polynomial q such that

$$q\left(\tfrac{\tau_i+\tau_{i+1}}{2}\right) = 0, \quad q^{(k)}(z) = q^{(k+1)}(z) = 0, \quad q(\tau_1) = -\mathrm{sgn}\, p_*(\tau_1),$$

which satisfies the condition (3.2), a contradiction. $\square$

The polynomials of degree n with n alternation points in $[-1,1]$ are called Zolotarev polynomials, they divide into 3 groups depending on the stucture of the set $\mathcal{A} := (\tau_i)$ of their alternation points.

A) $\mathcal{A}$ contains $n+1$ points. Then $p_* = T_n$,

B) $\mathcal{A}$ contains n points but only one of the endpoints. Then p_* can be continued to the larger interval, say $[-1, 1+c]$, on which it has $n+1$ alternation points. Hence, it is a transformed Chebyshev polynomial, $p_*(x) = T_n(ax+b)$, $|a| < 1$. We can exclude this case from consideration since clearly $\|p_*^{(k)}\| < \|T_n^{(k)}\|$.

C) $\mathcal{A}$ contains n points including both endpoints. Then p_* is called a proper Zolotarev polynomial, and we want to show that it does not attain any local maximum of $M_k(z)$. For this, we need one more characterization property of Z.

Lemma 3.4. *Let $M_k(z) = Z^{(k)}(z)$, where Z has exactly n alternation points (τ_i). Then*

$$R^{(k)}(z) = 0, \quad R(x) := \prod_{i=1}^n (x - \tau_i).$$

Proof. By the Lagrange interpolation formula with the nodes (τ_i), any $q \in \mathcal{P}_n$ can be written in the form

$$q(x) = cR(x) + \sum_{i=1}^{n} \frac{q(\tau_i)}{R'(\tau_i)} R_i(x), \quad R_i(x) := \frac{R(x)}{x - \tau_i},$$

so that

$$q^{(k)}(z) = cR^{(k)}(z) + \sum_{i=1}^{n} \frac{q(\tau_i)}{R'(\tau_i)} R_i^{(k)}(z).$$

If $R^{(k)}(z) \neq 0$, then we may set $q(\tau_i) = -Z(\tau_i)$ and then use the freedom in choosing the constant c to annihilate the right-hand side, i.e., to obtain $q^{(k)}(z) = 0$, a contradiction to Lemma 3.1. $\square$

Remark 3.5. From the previous lemma, it follows that if $Z \neq T_n$, then it can attain some value $M_k(z)$ only for z strictly inside the interval $[-1, 1]$, whence

$$M_k(\pm 1) = |T_n^{(k)}(\pm 1)|.$$

Theorem 3.6 (Tikhomirov (1976)). *Let $Z \in \mathcal{P}_n$ be a proper Zolotarev polynomial such that*

$$Z^{(k)}(z) = M_k(z), \quad (\textit{hence } R^{(k)}(z) = 0), \quad Z^{(k+1)}(z) = 0\,.$$

Then the polynomial

$$P_\lambda := Z + \lambda R + \frac{\lambda^2}{2} c_0 R', \quad c_0 := \frac{R^{(k+1)}(z)}{Z^{(k+2)}(z)}\,,$$

satisfies for some z_λ

$$\|P_\lambda\| = \|Z\| - \mathcal{O}(\lambda^2), \quad P_\lambda^{(k)}(z_\lambda) = Z^{(k)}(z) + o(\lambda^2).$$

Lemma 3.7. *Let $f, g, h \in C^2[a, b]$ with $\|f\| = |f(x_0)|$, and let*

$$f'(x_0) = 0, \quad f(x_0) f''(x_0) < 0, \quad g(x_0) = 0, \quad g'(x_0) \neq 0.$$

Then there is an $\epsilon > 0$ such that

$$\phi(\lambda) := \left\| f + \lambda g + \tfrac{\lambda^2}{2} h \right\|_{C[x_0 - \epsilon, x_0 + \epsilon]} = \left| f(x_0) + \tfrac{\lambda^2}{2} \left(h(x_0) - \tfrac{g'(x_0)^2}{f''(x_0)} \right) \right| + o(\lambda^2).$$

Proof. Set

$$\psi(x, \lambda) := \phi'_\lambda(x) := f'(x) + \lambda g'(x) + \frac{\lambda^2}{2} h'(x).$$

Then

$$\psi(x_0, 0) = 0, \quad \partial_x \psi(x_0, 0) = f''(x_0) \neq 0, \quad \partial_\lambda \psi(x_0, 0) = g'(x_0).$$

By the implicit function theorem, there exists a function $x_\lambda = x(\lambda)$ such that

$$\psi(x,\lambda) = 0 \quad \Leftrightarrow \quad x = x_\lambda = x_0 - \frac{g'(x_0)}{f''(x_0)}\lambda + o(\lambda).$$

This means that, for small λ, the function $|f + \lambda g + \frac{\lambda^2}{2} h|$ has a unique maximum at the point $x = x(\lambda)$, and

$$\begin{aligned} \|\phi_\lambda\| &= \left| f(x_\lambda) + \lambda g(x_\lambda) + \tfrac{\lambda^2}{2} h(x_\lambda) \right| \\ &= \left| f(x_0) + \lambda^2 \tfrac{g'(x_0)^2}{f''(x_0)^2} \tfrac{f''(x_0)}{2} - \lambda^2 \tfrac{g'(x_0)}{f''(x_0)} g'(x_0) + \tfrac{\lambda^2}{2} h(x_0) + o(\lambda^2) \right| . \end{aligned}$$

Proof of Theorem 3.6 1) Firstly, let us apply the previous lemma to the functional

$$\phi(\lambda) := \|P_\lambda^{(k)}\|_{C[z-\epsilon,z+\epsilon]} .$$

In this case, $f := Z^{(k)}$, $g := R^{(k)}$, $h := R^{(k+1)}$, and the conditions of the lemma are satisfied. We obtain

$$\phi(\lambda) := \left| Z^{(k)}(z) + \tfrac{\lambda^2}{2} \left(c_0 R^{(k+1)}(z) - \tfrac{R^{(k+1)}(z)^2}{Z^{(k+2)}(z)} \right) \right| + o(\lambda^2) = |Z^{(k)}(z)| + o(\lambda^2)$$

(the expression in parentheses vanishes due to the definition of c_0).

2a) Next, we apply the lemma to the functional

$$\phi(\lambda) := \|P_\lambda\|_{C[\tau_i-\epsilon,\tau_i+\epsilon]}, \quad \tau_i \ne \pm 1.$$

Now $f := Z$, $g = R$, $h = R'$, and in a neighbourhood of each interior alternation point τ_i the norm of the polynomial P_λ is equal to the value $|Z(\tau_i) + \frac{\lambda^2}{2}\gamma_i| + o(\lambda^2)$, where

$$\gamma_i := \left[c_0 R'(\tau_i) - \tfrac{R'(\tau_i)^2}{Z''(\tau_i)} \right] = \tfrac{R'(\tau_i)}{Z''(\tau_i)} [c_0 Z''(\tau_i) - R'(\tau_i)] .$$

To prove that $\|P_\lambda\| = \|Z\| - \mathcal{O}(\lambda^2)$, it suffices to show that $\gamma_i Z(\tau_i) < 0$, and because $Z(\tau_i)Z''(\tau_i) < 0$ this is equivalent to the inequality

$$\delta_i := R'(\tau_i)[c_0 Z''(\tau_i) - R'(\tau_i)] > 0 , \quad \tau_i \ne \pm 1 . \tag{3.3}$$

Consider the polynomial

$$Q(x) := c_0 Z'(x) - R(x) . \tag{3.4}$$

It vanishes at $(\tau_i)_{i=2}^{n-1}$, and $Q^{(k)}(z) = Q^{(k+1)}(z) = 0$. Hence, by Lemma 3.2, it changes its sign only at (τ_i), and $Q'(\tau_i)$ alternate in sign. So does $R'(\tau_i)$, thus all $\delta_i := R'(\tau_i)Q'(\tau_i)$ are of the same sign. Let us show that $\delta_{n-1} > 0$. We have

$$\operatorname{sgn} Q'(\tau_{n-1}) = \operatorname{sgn} Q(t)\big|_{t\to\infty} \overset{(3.4)}{=} -\operatorname{sgn} R(t)\big|_{t\to\infty} = -1 = R'(\tau_{n-1}) .$$

The first equality is because τ_{n-2} is the rightmost zero of Q, the next one is because Z' in (3.4) is of degree $n-1$, and the last two follow because $R(x) = \prod_{i=1}^n (x - \tau_i)$.

2b) It remains to consider the endpoints, say $x = 1$, where we have

$$P_\lambda(1) = Z(1) + \frac{\lambda^2}{2} c_0 R'(1)\,.$$

As we have seen, $\operatorname{sgn} Q(1) = \operatorname{sgn} Q(t)\big|_{t\to\infty} = -1$, on the other hand, by (3.4),

$$\operatorname{sgn} Q(1) = \operatorname{sgn} c_0 Z'(1) = \operatorname{sgn} c_0 Z(1),$$

hence c_0 and $Z(1)$ are of opposite sign, and because $R'(1) > 0$

$$|P_\lambda(1)| = |Z(1)| - \mathcal{O}(\lambda^2). \qquad \square$$

Comment 3.8. The difference between Tikhomirov's and Bernstein's proofs is that, while Tikhomirov simply presents the polynomial P_λ and then proves its required properties, Bernstein moves the other way round. He considers the polynomial

$$P_1(x) = Z(x+\lambda) - \lambda\phi(x+\lambda) - \lambda^2\psi(x+\lambda),$$

where ϕ and ψ are any polynomials satisfying $\phi^{(k)}(z) = \psi^{(k)}(z) = 0$, so that

$$P_1^{(k)}(z-\lambda) = Z^{(k)}(z).$$

Then he expands P_1 with respect to λ,

$$P_1 = Z + \lambda[Z' - \phi] + \lambda^2[\tfrac{1}{2}Z'' - \phi' - \psi] + o(\lambda^2),$$

evaluates the value $\|P_1\|$, and tries to determine ϕ and ψ in order to get

$$\|P_1\| = \|Z\| - \mathcal{O}(\lambda^2)\,.$$

With that he arrives at $\phi = Z' - \frac{1}{c_0}R$ and $\psi = -\frac{1}{2}\phi'$, so that the polynomial he uses is actually the same as in Tikhomirov:

$$P_1 = Z + (\lambda/c_0)R + \tfrac{(\lambda/c_0)^2}{2}\, c_0 R'\,.$$

Comment 3.9. Lemma 3.1 is actually a criterion for a polynomial to attain the norm of the linear functional $\mu(p) = p^{(k)}(z)$ (and any other linear functional on $\mathcal{P}_n$). It was a starting point of V. Markov's studies [7, §2], and he derived from it two other criteria which were more convenient for applications. Notice the similarity between Lemma 3.1 and Kolmogorov's criterion for the element of best approximation.

Comment 3.10. The above given "book-proof" of V. Markov's inequality is not entirely complete. To bring it to the final Markov form one still needs to prove that

$$\|T_n^{(k)}\| = T_n^{(k)}(1) = \frac{n^2\,[n^2-1^2]\cdots[n^2-(k-1)^2]}{1\cdot 3\,\cdots\,(2k-1)}\,.$$

Both equalities are usually referred to as "easy to show", but it takes another half a page to really show them (we do it in §5.3)

3.2 "Small-o" proof of Bojanov

Tikhomirov provided his proof with the following comment [54, p. 285]: "This proof is not quite consistent from the point of view of theory of extremal problems. To act consistently, one should find a tangent direction (which is here unique, namely that of $R(\cdot)$), write down a general variation of the second order

$$P_\lambda(x) = Z(x) + \lambda R(x) + \frac{\lambda^2}{2} Y(x)\,,$$

and then apply again the necessary conditions of supremum. Such a plan is fulfilled in the paper by Dubovitsky–Milyutin [3]. Here we took a shorter way borrowing some parts of our arguments from Bernstein [1]".

It is not clear whether here Tikhomirov had any particular polynomial Y in mind. The paper [3] which we discuss in the next section does not make it clear either.

A version of "small-o" proof with a different polynomial Y was presented in 2002 by Bojanov [2] in his survey on Markov-type inequalities. Bojanov himself refers to his proof as "a simplification of Tikhomirov's variational approach as outlined in a private communication".

We will fit Bojanov's proof into the scheme of the previous section, and it makes our exposition quite different from his own. We discuss some of these differences in the comments below where we also show that, actually, he uses the polynomial

$$P_\epsilon(x) = Z(x) + \epsilon Z_\theta(x) + \frac{\epsilon^2}{2} Z_{\theta\theta}(x)\,, \tag{3.5}$$

which is the Taylor expansion of the Zolotarev polynomial $Z(x, \theta_z + \epsilon)$ in a neighbourhood of θ_z.

Recall that

$$R(x) := Z_\theta(x) = \prod_{i=1}^{n} (x - \tau_i), \quad \tau_i = \tau_i(\theta)\,, \tag{3.6}$$

where τ_i are the equioscillation points of the Zolotarev polynomial Z, and set

$$Y(x) := \sum_{i=2}^{n-1} \rho_i R_i(x), \quad \rho_i := \frac{R'(\tau_i)}{Z''(\tau_i)}\,, \quad R_i(x) := \frac{R(x)}{x - \tau_i}\,. \tag{3.7}$$

Theorem 3.11 (Bojanov (2002)). *Let $Z \in \mathcal{P}_n$ be a proper Zolotarev polynomial such that*

$$Z^{(k)}(z) = M_k(z) \quad (\textit{hence } R^{(k)}(z) = 0), \qquad Z^{(k+1)}(z) = 0.$$

Then the polynomial

$$P_\epsilon := Z + \epsilon R + \frac{\epsilon^2}{2} Y \tag{3.8}$$

satisfies for some z_ϵ

$$\|P_\epsilon\| = 1 + o(\epsilon^2), \qquad |P_\epsilon^{(k)}(z_\epsilon)| = |Z^{(k)}(z)| + \mathcal{O}(\epsilon^2). \tag{3.9}$$

Proof. 1) From definition (3.7) of Y, we find that

$$Y(\tau_i) = \rho_i R_i(\tau_i) = \frac{[R'(\tau_i)]^2}{Z''(\tau_i)}\,, \quad i = 2, \ldots, n-1.$$

Now, Tikhomirov's Lemma 3.7 applied to P_ϵ says that, in a neighbourhood of each interior τ_i, the local maximum of P_ϵ has the value

$$P_\epsilon(\tau_i^\epsilon) = Z(\tau_i) + \tfrac{\epsilon^2}{2}\left[Y(\tau_i) - \tfrac{[R'(\tau_i)]^2}{Z''(\tau_i)}\right] + o(\epsilon^2) = 1 + o(\epsilon^2)\,.$$

Near the endpoints of $[-1,1]$, the norm $\|P_\epsilon\|$ will not exceed 1 for small ϵ because $|Z(x)| \le 1$ and $Z'(\pm 1) \ne 0$.

2) To prove the second equality in (3.9) we apply Lemma 3.7 to $P_\epsilon^{(k)}$. So, in a neighbourhood of z, the local maximum of P_ϵ has the value

$$P_\epsilon^{(k)}(z_\epsilon) = Z^{(k)}(z) + \tfrac{\epsilon^2}{2}\left[Y^{(k)}(z) - \tfrac{[R^{(k+1)}(z)]^2}{Z^{(k+2)}(z)}\right] + o(\epsilon^2),$$

and because $Z^{(k)}(z) Z^{(k+2)}(z) < 0$ we have to deal with the inequality

$$d := Y^{(k)}(z) Z^{(k+2)}(z) - [R^{(k+1)}(z)]^2 \overset{?}{<} 0. \tag{3.10}$$

3) Since $Y = \sum_{i=2}^{n-1} \rho_i R_i$, and (trivially) $R' = \sum_{i=1}^{n} R_i$, we have

$$Y^{(k)}(x) = \textstyle\sum_{i=2}^{n-1} \frac{R'(\tau_i)}{Z''(\tau_i)} R_i^{(k)}(x), \quad R^{(k+1)}(x) = \sum_{i=1}^{n} R_i^{(k)}(x), \tag{3.11}$$

so we may write

$$\begin{aligned} d &= Z^{(k+2)}(z) \sum_{i=2}^{n-1} \tfrac{R'(\tau_i)}{Z''(\tau_i)} R_i^{(k)}(z) - R^{(k+1)}(z) \sum_{i=1}^{n} R_i^{(k)}(z) \\ &= R^{(k+1)}(z) \sum_{i=2}^{n-1} \left[\tfrac{Z^{(k+2)}(z)}{R^{(k+1)}(z)}\,\tfrac{R'(\tau_i)}{Z''(\tau_i)} - 1\right] R_i^{(k)}(z) \\ &\qquad - R^{(k+1)}(z)\,[R_1^{(k)}(z) + R_n^{(k)}(z)]\,. \end{aligned}$$

By Markov's interlacing property (since zeros of R and R_i interlace)

$$R^{(k)}(z) = 0 \quad \Rightarrow \quad \operatorname{sgn} R^{(k+1)}(z) = \operatorname{sgn} R_i^{(k)}(z) \quad \forall\, i,$$

so we are done once we prove that $\frac{Z^{(k+2)}(z)}{R^{(k+1)}(z)}\,\frac{R'(\tau_i)}{Z''(\tau_i)} - 1 < 0$, or, with the previously used notation $c_0 := \frac{R^{(k+1)}(z)}{Z^{(k+2)}(z)}$, that

$$\delta_i := \tfrac{1}{c_0 Z''(\tau_i)}[c_0 Z''(\tau_i) - R'(\tau_i)] > 0, \quad \tau_i \ne \pm 1.$$

4) The latter is proved like in Tikhomirov's proof, by considering the polynomial $Q = c_0 Z' - R$. $\square$

Comment 3.12. Bojanov wrote his polynomial (3.8) in the form

$$P_\epsilon(x) := Z(x) + \epsilon \prod_{i=1}^{n} (x - \tau_i + \tfrac{\epsilon}{2}\rho_i)$$

and dealing with (3.9) he repeated twice the arguments (of Tikhomirov's lemma) based on the implicit function theorem.

Also, he used not (3.11) but the formula

$$Y^{(k)}(z) = \textstyle\sum_{i=2}^{n-1} A_i \frac{[R'(\tau_i)]^2}{Z''(\tau_i)} \quad \big(= \sum_{i=1}^{n} A_i Y(\tau_i) \big),$$

which stems from the representation of the linear functional $\mu(p) = p^{(k)}(z)$ on $\mathcal{P}_n$,

$$p^{(k)}(z) = \sum_{i=1}^{n} A_i p(\tau_i), \quad A_i A_{i+1} < 0, \tag{3.12}$$

so that, finally, he verified not (3.10) but the inequality

$$Z^{(k)}(z) \Big[-\tfrac{[R^{(k+1)}(z)]^2}{Z^{(k+2)}(z)} + \textstyle\sum_{i=2}^{n-1} A_i \frac{[R'(\tau_i)]^2}{Z''(\tau_i)} \Big] > 0. \tag{3.13}$$

Comment 3.13. Let us show that P_ϵ has the form (3.5). We focus on the term Y in (3.7)-(3.8) and we claim that it is nothing but R_θ. Indeed, from definition (3.6) of R, since $\tau_1(\theta) \equiv -1$ and $\tau_n(\theta) \equiv 1$, we obtain

$$R_\theta(x) = \textstyle\sum_{i=2}^{n-1} (-\tau_i'(\theta)) R_i(x)\,,$$

and, by differentiating the identity $Z'(\tau_i(\theta), \theta) \equiv 0$, we find that

$$-\tau_i'(\theta) = \frac{Z_\theta'(\tau_i)}{Z''(\tau_i)} = \frac{R'(\tau_i)}{Z''(\tau_i)} = \rho_i\,.$$

Hence, $Y = R_\theta$, and Bojanov's polynomial (3.8) is

$$P_\epsilon(x) = Z(x) + \epsilon R(x) + \frac{\epsilon^2}{2} R_\theta(x)\,,$$

or, since $R = Z_\theta$,

$$P_\epsilon(x) = Z(x, \theta_z) + \epsilon Z_\theta(x, \theta_z) + \tfrac{\epsilon^2}{2} Z_{\theta\theta}(x, \theta_z) = Z(x, \theta_z + \epsilon) + o(\epsilon^2).$$

So, P_ϵ is nothing but the second order Taylor expansion of the Zolotarev polynomial $Z(x, \theta_z + \epsilon)$ with the perturbed parameter θ in a neighbourhood of θ_z. In particular, the equality $\|P_\epsilon\| = 1 + o(\epsilon^2)$ is now straightforward, and moreover, the key inequality (3.10) to be verified turns out to be

$$d := R_\theta^{(k)}(z) Z^{(k+2)}(z) - [R^{(k+1)}(z)]^2 \overset{?}{<} 0, \tag{3.14}$$

exactly the same as V. Markov considered. Basically, all three proofs – by V. Markov, Bernstein–Tikhomirov and Bojanov – deduce that

$$M_k(z) = Z^{(k)}(z, \theta_z), \quad Z^{(k+1)}(z) = 0 \quad \Rightarrow \quad |Z^{(k)}(z, \theta_z)| < |Z^{(k)}(z_\epsilon, \theta_z + \epsilon)|\,.$$

3.3 Proofs of Dubovitsky–Milyutin and Mohr

Dubovitsky–Milyutin's proof. The main goal of [3], as postulated in section 5°, is to show, "as a result of the analysis of Euler equations for the first and second variation, that the optimal polynomial [that attains the global maximum M_k] is uniquely determined and is the Chebyshev polynomial T_n". The first two pages describe some general theory, the proof itself takes another two pages.

In our notations, they start with the formula

$$\frac{p^{(k)}(z)}{Z^{(k)}(z)} = \int \frac{p(x)}{Z(x)}\, d\mu \quad \left(:= \sum_{i=1}^{n} (-1)^i \mu_i p(\tau_i) \right) \tag{6}$$

(which is the analogue of (3.12)). After a while, the proof arrives at verification of the following inequality (which is the last but one formula on the very last page):

$$\frac{[R^{(k+1)}(z)]^2}{Z^{(k)}(z)\, Z^{(k+2)}(z)} - \int \frac{[R'(x)]^2}{Z(x)\, Z''(x)}\, d\mu_{\rm ext} \overset{?}{<} 0\,. \tag{10}$$

With $\mu_{\rm ext}$ being the same measure μ from (6) but without the endpoints (so to say), this inequality is identical to inequality (3.13) considered by Bojanov, which as we showed is the same as the inequality (3.14) considered by Markov.

At this point, nothing says that we are approaching the end, but then the magic happens. The next and final expression appears like a rabbit pulled from a hat. Quotation: "Since $R(x)(x-\beta) = Z'(x)(x^2-1)$, therefore by making use of $R^{(k)}(z) = Z^{(k+1)}(z) = 0$ and identity (6), we can reduce (10) to

$$\frac{1}{R^{(k+1)}(z)} \left[\frac{R'(\beta)R(x) - R(\beta)R'(x)}{x-\beta} \right]^{(k-1)}_{x=z} + \frac{k(k+1)Z^{(k)}(z)R(\beta)}{(z-\beta)Z^{(k+2)}(z)(\beta^2-1)} > 0\,.\text{"} \tag{11}$$

The last two paragraphs swiftly show that both summands are positive (they are indeed), and that's the end of the article.

I don't think that this "proof" can be taken seriously.

First of all, both Markov and Bojanov spent more than a page on rather fine calculations before they brought their analogues of (10) to some clearer forms. It is hard to believe that Dubovitsky–Milyutin managed to do it in a few lines (which they did not even bother to present).

Secondly, no matter how you transform (10), the final relation should be still equivalent to that of Markov. In (11), the expression in square brackets is equal to what we denoted in (2.12) by $-R(\beta)\psi(x)$, so (11) is identical to

$$-\frac{R(\beta)\psi^{(k-1)}(z)}{R^{(k+1)}(z)} + \frac{k(k+1)Z^{(k)}(z)R(\beta)}{(z-\beta)Z^{(k+2)}(z)(\beta^2-1)} > 0\,. \tag{11$'$}$$

This looks very close to Markov's formula (2.20), but there is no match.

Trigonometric proof of Mohr. Mohr starts his paper [8] by making the change of variable, $x = \cos\theta$, thus switching from algebraic polynomials $p(x)$ to the cosine polynomials $\phi(\theta)$. With such a switch, the Markov problem becomes the problem of finding

$$M_k := \sup_{\|\phi\|\le 1} \|\phi^{[k]}\| \qquad \big(M_k(\xi) := \sup_{\|\phi\|\le 1} |\phi^{[k]}(\xi)|\big)\,,$$

where

$$\phi^{[1]} = -\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\,\phi\,, \qquad \|\phi\| = \max_{\theta\in[0,\pi]} |\phi(\theta)|\,.$$

Mohr wants to show that "this supremum is attained exactly for $\phi(\theta) = \cos n\theta$", so in §1.7 he assumes that

$$M_k = \Gamma^{[k]}(\xi)\,, \tag{3.15}$$

with some cosine polynomial Γ and some $\xi \in [0,\pi]$, and in §2 tries to prove that the case when Γ has less than $n+1$ equioscillation points is impossible.

1) I did not understand the reasons to move to trigonometry as Mohr considers his cosine polynomials only on the interval $[0,\pi]$, i.e., he does not make any use of periodicity (as one could expect). With such a move, nothing really changes except for complicating the matter of things.

2) At the begining, the proof develops as in the algebraic case. In particular, Mohr shows (§§2.1-2.5) that the extremal polynomial Γ has at least n points of equioscillation, and if it has exactly n points, then its resolvent satisfies $R^{[k]}(\xi) = 0$, therefore ξ is strictly inside $[0,\pi]$, hence $\Gamma^{[k+1]}(\xi) = 0$. (The latter means, by the way, that $M_k(\xi)$ is not necessarily the global maximum, but only an extreme value of $M_k(\cdot)$.)

3) However, the final part starting from §2.13 is taking more and more strange forms, and in §2.15, assuming actually that

$$M_k(\xi) = \Gamma^{[k]}(\xi), \quad R^{[k]}(\xi) = 0, \quad \Gamma^{[k+1]}(\xi) = 0, \tag{3.16}$$

Mohr managed to construct a family of polynomials ϕ such that

$$\|\phi\| \le 1, \qquad \phi^{[k]}(\xi) > \Gamma^{[k]}(\xi)\,. \tag{3.17}$$

This is of course a contradiction to the initial guess (3.15), so one might have concluded that the intermediate assumption that Γ has exactly n equioscillations was false. But it is also a contradiction to (3.16), which as we know may well be true for some Γ of Zolotarev type. I think that Mohr somehow got it wrong (in his formula (30), I suspect).

4) Even more strange is that Mohr does not consider relations (3.17) as something extraordinary, and spends two pages more in deriving further statements before he finally arrives at a contradiction.

3.4 Limitations of variational and "small-o" methods

All three authors – Bernstein, Tikhomirov and Bojanov – while using the small-o arguments, arrived actually, at exactly the same conclusion which was provided by V. Markov.

Theorem 3.14. *The local extreme values of $M_k(\cdot)$ attained by a polynomial other than T_n are local minima, or, equivalently, all local maximal values of $M_k(\cdot)$ are attained by the Chebyshev polynomial $\pm T_n$.*

The only difference is that V. Markov proved that $M_k(\cdot)$ indeed have local maxima and minima.

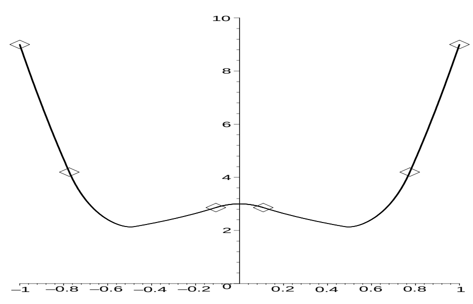
What is important in such a conclusion is that it shows that we cannot apply the variational or a "small-o" method to the Markov-type problem, unless we are sure that the local behaviour of $M_{k,\mathcal{F}}(\cdot)$ follows the pattern given by the theorem above.

Example 3.15. Consider the Landau–Kolmogorov problem

$$M_{k,\sigma}(z) = \sup_{f\in W_\infty^{n+1}(\sigma)} |f^{(k)}(z)|, \quad z \in [-1,1]\,,$$

where $W_\infty^{n+1}(\sigma) = \{f : \|f\| \le 1, \|f^{(n+1)}\| \le \sigma\}$. For $\sigma = 0$ it reduces to the Markov problem for polynomials, hence for small σ, the pointwise bound $M_{k,\sigma}(z)$ should be close to the Markov pointwise bounds $M_k(z)$.

The function $M_k(z)$ has $(n-k)$ local minima and $(n-k-1)$ local maxima as illustrated on the graph below (for $n = 3$ and $k = 1$). Now, according to Pinkus' results [30], the Chebyshev-like function $T_* \in W_\infty^{n+1}(\sigma)$ that attains the value $M_{k,\sigma}(z)$ at $z = 1$ takes other values of $M_{k,\sigma}(z)$ only at a finite set of $(n-k)$ points, and similarly for $\widehat{T}_*$ which is extremal for $z = -1$. As $\sigma \to 0$, these points will tend to the ends of Zolotarev intervals (ξ_i) and (η_i), respectively, and we see that, for small σ, there are local maxima of $M_{k,\sigma}(\cdot)$ that are achieved by functions of Zolotarev type (the maximum at $z = 0$ on the figure).



Hence, for small σ, we cannot prove that T_* is the global solution using variational or "small-o" methods. It does not mean that this is not true, most likely it is, but we certainly need other methods to prove it. In fact, the same picture is true for any $\sigma > 0$ when the extremal functions for $z = 1$ and for $z = -1$ are two proper Zolotarev splines (our Conjecture 6.1 in [31] that, for $\sigma > \sigma_0$, the function $M_{k,\sigma}(\cdot)$ is monotone on $[0,1]$ is not true, although it may still be true for $\sigma = \|T_{n+1,r}^{(n+1)}\|$).

4 Pointwise majorants

4.1 The case $k = 1$

Here we show how Andrei Markov proved the global inequality for the 1-st derivative using the fact that Zolotarev polynomials form the extremal set for the pointwise problem.

Theorem 4.1 (A. Markov (1889)). *We have*

$$\sup_{\|p\| \le 1} \|p'\| = T_n'(1) = n^2 . \tag{4.1}$$

Proof. For a fixed θ, the Zolotarev polynomial $Z_n(x, \theta)$ satisfies the differential equation

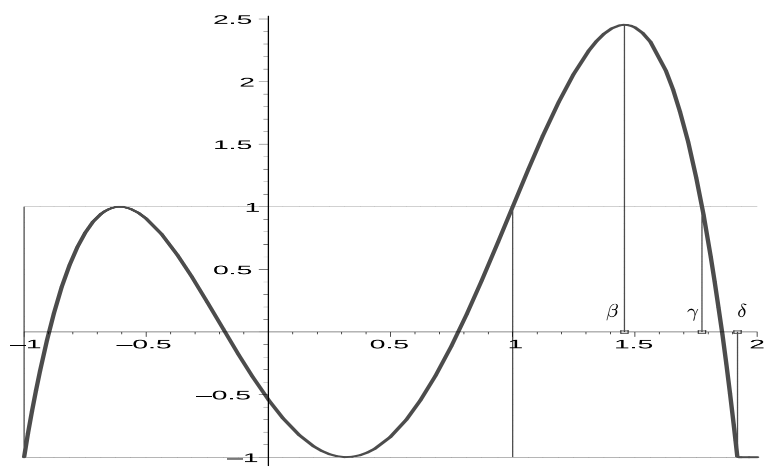
$$1 - y(x)^2 = \frac{(1-x^2)(x-\gamma)(x-\delta)}{n^2(x-\beta)^2} y'(x)^2 ,$$

or

$$y'^2 = \frac{(x-\beta)^2}{(x-\gamma)(x-\delta)} \cdot \frac{n^2(1-y^2)}{1-x^2} ,$$

where β, γ, δ are of the same sign, and

$$|x| \le 1 < |\beta| < |\gamma| < |\delta|.$$



The latter implies

$$0 < \frac{(x-\beta)^2}{(x-\gamma)(x-\delta)} < 1 ,$$

whence

$$y'^2 \le \frac{n^2(1-y^2)}{1-x^2} \le \frac{n^2}{1-x^2} .$$

The same inequality is valid for the Chebyshev polynomial T_n, for its transformations $T_n(ax+b)$ with $|a|<1$, and for T_{n-1}. Hence

$$M_1(z) \le \frac{n}{\sqrt{1-z^2}} \quad \Leftrightarrow \quad |p'(x)| \le \frac{n}{\sqrt{1-x^2}}\,\|p\|\,, \tag{4.2}$$

and we have arrived at the Bernstein inequality for algebraic polynomials (which A. Markov did not stop on).

The last step is described in every monographs:

a) if $|x| \le \cos\frac{\pi}{2n}$, then $\frac{n}{\sqrt{1-x^2}} \le n^2$,

b) if $|x| > \cos\frac{\pi}{2n}$, then $|p'(x)| \le |T_n'(x)|\|p\| \le n^2\,\|p\|$.

Comment 4.2. Nowadays, the usual way to prove the Bernstein "algebraic" inequality (4.2) (hence A. Markov's inequality (4.1)) is through the Bernstein inequality for trigonometric polynomials

$$\|t_n'\| \le n\,\|t_n\|\,, \tag{4.3}$$

since the latter has a very simple proof based on the so-called comparison lemma:

$$\|t_n\| < 1, \quad |t_n(\eta)| = |\cos n\xi| \quad \Rightarrow \quad |t_n'(\eta)| < n|\sin n\xi|\,.$$

However, Bernstein himself moved the other way round [46]. Firstly, exactly in the same way as A. Markov (see the next comment), he derived (4.2). With the substitution $x=\cos\theta$, this gives the trigonometric version (4.3) only for even polynomials $t_n(\theta)=\sum a_k\cos k\theta$, so he proved one more algebraic inequality

$$\left|\tfrac{\partial}{\partial x}(p(x)\sqrt{1-x^2})\right| \le \frac{n}{\sqrt{1-x^2}}\,\max\,|p(x)\sqrt{1-x^2}|, \quad p\in\mathcal{P}_{n-1}\,,$$

which provides (4.3) for odd $t_n(\theta)=\sum b_k\sin k\theta$. Finally, he got the general result by a tricky combination of those two.

Comment 4.3. Bernstein [46] derived the "Bernstein" inequality (4.2) exactly in the same way as A. Markov, which we have just described. He accompanied his result with the following footnote: "This is the statement of A. Markov's theorem given in his aforementioned paper. Unfortunately, I became acquainted with that paper, as well as with the composition of V. A. Markov, only when preliminary algebraic theorems, which constitute the content of the present chapter, were found and derived independently by myself. No doubt earlier acquaintance with the ideas of these scientists would have simplified my task and, probably, the presentation of this chapter. However, I considered it unnecessary to put changes into my fully accomplished proofs, because of the auxiliary character of the above-mentioned theorems ..." and there are further 2-3 lines of these beautiful poetry.

4.2 Bernstein's results

Bernstein was very enthusiastic about Markov's inequality. Not only made he Markov's results available to the western public, but he also put a lot of effort into deepening and improving them. It was not until 1938 that he managed to find a simpler proof, but meanwhile he produced several important refinements.

1) First of all, by iterating his inequality for the 1-st derivative,

$$|p'(x)| \le \frac{n}{\sqrt{1-x^2}}\,\|p\|, \tag{4.4}$$

Bernstein found [45] a *pointwise majorant for all* k:

$$|p^{(k)}(x)| \le \left(\frac{\sqrt{k}}{\sqrt{1-x^2}}\right)^k n(n-1)\cdots(n-k+1)\,\|p\|. \tag{4.5}$$

The proof for the case $k=3$ gives the general flavour:

$$\begin{aligned}
|p'''(x)| &\le \frac{n-2}{\sqrt{x_1^2-x^2}}\|p''\|_{C[-x_1,x_1]} \\
&\le \frac{n-2}{\sqrt{x_1^2-x^2}}\,\frac{n-1}{\sqrt{x_2^2-x_1^2}}\|p'\|_{C[-x_2,x_2]} \\
&\le \frac{n-2}{\sqrt{x_1^2-x^2}}\,\frac{n-1}{\sqrt{x_2^2-x_1^2}}\,\frac{n}{\sqrt{1-x_2^2}}\|p\|_{C[-1,1]},
\end{aligned}$$

where x_1, x_2 are any numbers satisfying $x^2 < x_1^2 < x_2^2 < 1$, and the choice $x_1^2 - x^2 = x_2^2 - x_1^2 = 1 - x_2^2 = \frac{1-x^2}{k}$ is clearly optimal and does the job.

The estimate (4.5) shows in particular that, for a given k, the order of the k-th derivative of $p \in \mathcal{P}_n$ inside the interval is $\mathcal{O}(n^k)$ thus differing essentially from the order $\mathcal{O}(n^{2k})$ at the endpoints.

2a) He did not stop with that and, in 1913, established [46] the *exact asymptotic bound*:

$$M_k(x) \sim \left(\frac{n}{\sqrt{1-x^2}}\right)^k.$$

For this proof, Bernstein found an *exact* form of the polynomial $q \in \mathcal{P}_{n-2}$ that deviates least from the function

$$\phi(x) = cx^n + \sigma x^{n-1} + \frac{A}{x-a}, \quad |a| > 1,$$

and, letting $A \to 0$, derived *asymptotic formulas for Zolotarev polynomial.*

2b) In the same paper [46], still bothered by complexity of V. Markov's proof, he suggested simpler arguments that provide *asymptotic form of Markov's inequality*

$$|p^{(k)}(x)| < M_k(1+\epsilon_n), \quad \epsilon_n = \mathcal{O}(1/n^2). \tag{4.6}$$

Here they are. Assuming that $\|p\|_{C[-1,1]} = 1$, it is quite easy to show that $|p^{(k)}(1)| \le T_n^{(k)}(1)$, and applying this inequality to the interval $[-1, x]$ we obtain the majorant

$$|p^{(k)}(x)| \le T_n^{(k)}(1) \left(\frac{2}{x+1}\right)^k =: F(x).$$

Comparing it with the previous majorant (4.5),

$$|p^{(k)}(x)| \le \left(\frac{\sqrt{k}}{\sqrt{1-x^2}}\right)^k \frac{n!}{(n-k)!} =: G(x),$$

we notice that, on $[0, 1]$, the functions F and G are decreasing and increasing respectively, hence the common bound for $|p^{(k)}(x)|$ is given by the value $F(x_*) = G(x_*)$ which results in (4.6).

3) Finally, in 1930, Bernstein generalized [48] his classical inequality (4.4) to the case when p is bounded by a polynomial majorant: *if*

$$|p_{n+m}(x)| \le \mu(x) = \sqrt{P^2(x) + (1-x^2)Q^2(x)}\,,$$

where P and Q are two polynomials of degree m and and $(m-1)$ respectively, which have interlacing zeros, then

$$|p'_{n+m}(x)| \le \sqrt{\frac{\left[nP(x) + xQ(x) + (x^2-1)Q'(x)\right]^2 + (1-x^2)\left[P'(x) + nQ(x)\right]^2}{1-x^2}}\,. \tag{4.7}$$

As a consequence, he concluded (without proofs) that *if $f(x) > 0$ is any continuous function, then*

$$|p_n(x)| \le f(x) \quad \Rightarrow \quad |p'_n(x)| \le \frac{nf(x)}{\sqrt{1-x^2}}\,(1 + \mathcal{O}(1/n))\,,$$

and, moreover,

$$|p_n(x)| \le f(x) \quad \Rightarrow \quad \|p_n^{(k)}\| \le T_n^{(k)}(1) f(\pm 1)(1 + \mathcal{O}(k^2/n))\,.$$

(With respect to the last two results, I have some doubts. I think that the value $E_n(f)$ of the best approximation to f should be somehow involved.)

4.3 Schaeffer–Duffin's majorant

In 1938, the same year when Bernstein produced his proof of Markov's inequality using small-o arguments, two American mathematicians, R. Duffin and A. Schaeffer, came out with another proof [9], the main part of which was a generalization of the pointwise Bernstein inequality $p'(x) \le \frac{n^2}{\sqrt{1-x^2}}\|p\|$ to higher derivatives. It is a very nice and short paper, so we only sketch briefly the main elements of the proofs.

Let T_n be the Chebyshev polynomial and $S_n(x) := \frac{1}{n}\sqrt{1-x^2}\,T'_n(x)$.

Theorem 4.4 (Schaeffer-Duffin (1938)). *Let $p \in \mathcal{P}_n$ be such that*

$$|p(x)| \le 1 \equiv |T_n(x) + iS_n(x)| \,.$$

Then

$$|p^{(k)}(x)| \le D_k(x) := |T_n^{(k)}(x) + iS_n^{(k)}(x)| \,.$$

Proof (Sketch). The formulation of the theorem is a bit misleading because what Schaeffer–Duffin really assume is that, by Bernstein's inequality,

$$|p'(x)| < D_1(x) = |T_n'(x) + iS_n'(x)| = \frac{n}{\sqrt{1-x^2}}, \quad p \ne \pm T_n$$

(and it is essential that S_n' is unbounded near the endpoints). From that it follows that, for every $\alpha \in (0,\pi)$ and for every $\lambda \in [-1,1]$, the function

$$F'(x) := \cos\alpha\, T_n'(x) + \sin\alpha\, S_n'(x) - \lambda p'(x) \quad \Big(= \tfrac{n\sin(nt-\alpha)}{\sqrt{1-x^2}} - \lambda p'(x)\Big)$$

has at least n distinct zeros in $(-1,1)$. They also prove that $F^{(n+1)} = cS^{(n+1)}$ does not change sign, hence, on $(-1,1)$, $F^{(k)}$ has exactly $(n+1-k)$ zeros all of which are simple. Finally, they show that, if one supposes that, at some $x_0 \in (-1,1)$,

$$|p^{(k)}(x_0)| \ge D_k(x_0),$$

then one can choose particular α and λ so that $F^{(k)}$ has a double zero at such x_0, a contradiction that proves the theorem. $\square$

Lemma 4.5. *For all k, we have*

a) $D_k(\cdot)$ *is a strictly increasing function on* $[0,1)$,

b) *the* $(n-k)$ *zeros of* $T_n^{(k)}$ *interlace with* $(n-k+1)$ *zeros of* $S_n^{(k)}$.

Proof (Sketch). This lemma is trivial for $k=1$ because $D_1(x) = \frac{n}{\sqrt{1-x^2}}$ and $S_1'(x) = -\frac{nT_n(x)}{\sqrt{1-x^2}}$ (hence a simple proof of A. Markov's inequality), but for general k Schaeffer–Duffin had to come through the following arguments.

Both functions $T_n^{(k)}$ and $S_n^{(k)}$ are independent solutions of the differential equation

$$(1-x^2)y''(x) - (2k+1)xy'(x) + (n^2-k^2)y(x) = 0,$$

hence, by Sturm's theorem, their zeros interlace. The latter equation may also be rewritten in the equivalent form

$$\frac{d}{dx}\left\{(1-x^2)[f_{k+1}(x)]^2 + (n^2-k^2)[f_k(x)]^2\right\} = 4kx\,[f_{k+1}(x)]^2\,, \qquad (4.8)$$

to which $[T_n^{(k)}(x)]^2$ and $[S_n^{(k)}(x)]^2$, hence also $[D_k(x)]^2$, are particular solutions. Substituting the power series of $[D_k(x)]^2$ into (4.8), they derive by induction on k that

$$[D_k(x)]^2 = \sum_{i=0}^{\infty} a_{2i} x^{2i}, \quad a_{2i} > 0. \qquad \square$$

Proof of V. Markov's inequality. From two previous results, Schaeffer–Duffin derived V. Markov's inequality

$$\|p^{(k)}\| \le \|T_n^{(k)}\| \, \|p\|$$

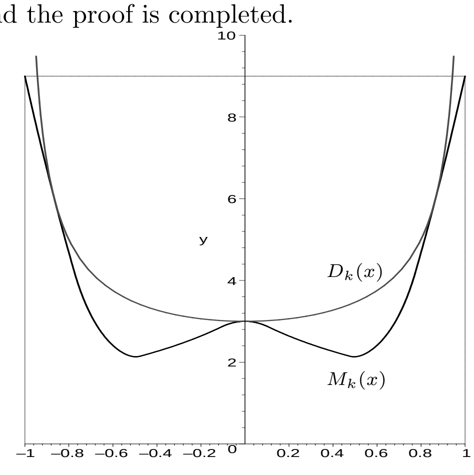
exactly in the same way as A. Markov's inequality for the 1-st derivative $\|p'\| \le n^2\|p\|$ can be derived from the Bernstein inequality $|p(x)| \le \frac{n}{\sqrt{1-x^2}}\|p\|$.

Namely, for x_* being the rightmost zero of $S_n^{(k)}$, it follows that

a) if $|x| \le x_*$, then $|p^{(k)}(x)| \le D_k(x) \le D_k(x_*) = T_n^{(k)}(x_*)$,

b) if $|x| > x_*$, then $|p^{(k)}(x)| \le |T_n^{(k)}(x)|$ (by Rolle's theorem).

and the proof is completed. $\square$



1) At the $(n-k+1)$ zeros of $S_n^{(k)}(x)$

$$M_k(x) = D_k(x) = |T_n^{(k)}(x)|\,.$$

2) Otherwise

$$M_k(x) < D_k(x)\,.$$

For $k=1$, the Schaeffer–Duffin majorant coincides with that of Bernstein, $D_1(x) = \frac{n}{\sqrt{1-x^2}}$, but they did not try to find its exact form for any other k. We performed some computations ourselves.

Lemma 4.6. *For all k, we have*

$$\frac{1}{n^2}[D_{k+1}(x)]^2 = \sum_{m=0}^{k} \frac{b_m}{(1-x^2)^{k+1+m}},$$

where $b_m = \binom{k+m}{2m} 1^2 \cdot 3^2 \cdots (2m-1)^2 \cdot (n^2-(m+1)^2) \cdots (n^2-k^2)$.

Proof. Assuming that $\frac{1}{n^2}[D_k(x)]^2 = \sum_{m=0}^{k-1} \frac{a_m}{(1-x^2)^{k+m}}$, from (4.8) we obtain

$$\frac{1}{n^2-k^2}\,b_0 = a_0, \quad \frac{1}{n^2-k^2}\,b_1 = \frac{k+1}{k-1}\,a_1, \quad \ldots, \quad \frac{1}{n^2-k^2}\,b_{k-1} = \frac{2k-1}{1}\,a_{k-1},$$

and the last coefficient $b_k = 1^2 \cdot 3^2 \cdots (2k-1)^2$ is found from

$$\sum_{m=0}^{k} b_k = [D_k(0)]^2, \quad D_k(0) = \begin{cases} n(n^2-1^2)(n^2-3^2)\cdots(n^2-(k-2)^2), & \text{odd } k; \\ n^2(n^2-2^2)(n^2-4^2)\cdots(n^2-(k-2)^2), & \text{even } k. \end{cases}$$

In particular, we get

$$\frac{1}{n^2}[D_2(x)]^2 = \frac{(n^2-1)}{(1-x^2)^2} + \frac{1}{(1-x^2)^3}\,,$$

$$\frac{1}{n^2}[D_3(x)]^2 = \frac{(n^2-1)(n^2-4)}{(1-x^2)^3} + \frac{3(n^2-4)}{(1-x^2)^4} + \frac{9}{(1-x^2)^5}\,,$$

$$\frac{1}{n^2}[D_4(x)]^2 = \frac{(n^2-1)(n^2-4)(n^2-9)}{(1-x^2)^3} + \frac{6(n^2-4)(n^2-9)}{(1-x^2)^4} + \frac{45(n^2-9)}{(1-x^2)^5} + \frac{225}{(1-x^2)^6}\,.$$

4.4 Generalization: Vidensky majorant

In 1951, Vidensky [20] extended results of Schaeffer–Duffin to the case when restrictions on p are given by an arbitrary polynomial majorant:

$$|p(x)| \le \mu(x) = \sqrt{R_{2m}(x)},$$

where R_{2m} is any polynomial of degree $\le 2m$ that is non-negative on $[-1,1]$. By Lucas theorem, for any $n \ge m$, such a polynomial can be represented in the form

$$R_{2m}(x) = P_n^2(x) + (1-x^2)Q_{n-1}^2(x),$$

where P_n and Q_{n-1} satisfy the following conditions:

a) $P_n \in \mathcal{P}_n$ and $Q_{n-1} \in \mathcal{P}_{n-1}$;

b) all zeros of P_n and Q_{n-1} lie in $[-1,1]$ and interlace;

c) the leading coefficients of P_n and Q_{n-1} are positive.

Moreover,

$$\begin{aligned} &P_{m+n}(x) + i\sqrt{1-x^2}Q_{m+n-1}(x) \\ &\qquad = \left[P_m(x) + i\sqrt{1-x^2}Q_{m-1}(x)\right]\left[T_n(x) + iS_n(x)\right] \end{aligned} \tag{4.9}$$

Theorem 4.7 (Vidensky (1951)). *Let $p \in \mathcal{P}_n$ be such that*

$$|p(x)| \le \mu(x) \equiv |P_n(x) + i\sqrt{1-x^2}Q_{n-1}(x)|\,.$$

Then

$$|p^{(k)}(x)| \le V_k(x) := \left|P_n^{(k)}(x) + i\left[\sqrt{1-x^2}Q_{n-1}(x)\right]^{(k)}\right|\,.$$

In his proof, Vidensky follows the same route as Schaeffer–Duffin, taking as the starting point the generalization of the classical Bernstein inequality (that was established by Bernstein himself, see (4.7))

$$|p'(x)| < V_1(x) = \big|P_n'(x) + i\big[\sqrt{1-x^2}Q_{n-1}(x)\big]'\big|, \quad p \ne \pm P_n\,.$$

However, it was not a straightforward journey, because the Schaeffer–Duffin arguments heavily relied on the fact that both $T_n(x)$ and $\sqrt{1-x^2}\frac{1}{n}T_n'(x)$ satisfy one and the same differential equation, whereas P_n and $\sqrt{1-x^2}Q_{n-1}(x)$ have no such property in general. One of Vidensky innovations was a statement about functions with interlacing zeros that generalized the well-known V. Markov's result about polynomials.

Lemma 4.8. *Let $f_1, f_2 \in C^1[a,b]$ be two functions such that any linear combination $c_1f_1 + c_2f_2$ has $\le n$ zeros counting multiplicity. If both f_1 and f_2 have n zeros, all simple, and these zeros interlace, then zeros of f_1' and f_2' interlace too.*

Proof (Sketch). Let $(t_i)_{i=1}^n$ be the zeros of f_1, then by the interlacing conditions the function $g = c_1f_1 + c_2f_2$ alternates in sign on the sequence $(t_i)_{i=1}^n$, hence all of its zeros are simple. The latter means that, for any x, the system

$$\begin{aligned} c_1f_1(x) + c_2f_2(x) &= 0, \\ c_1f_1'(x) + c_2f_2'(x) &= 0 \end{aligned}$$

has only the trivial solution, thus

$$f_1(x)f_2'(x) - f_1'(x)f_2(x) \ne 0, \quad \forall x \in [a,b]\,.$$

From here we get that at the points $(s_i)_{i=1}^{n-1}$ where $f_1'(s_i) = 0$, we have

$$\operatorname{sgn} f_2'(s_i) = \operatorname{sgn} f_1(s_i) = (-1)^i\gamma,$$

and the conclusion follows. □

The only result of Schaeffer–Duffin (based on the differential equation) for which Vidensky did not find an appropriate substitution was monotonicity of D_k, i.e., he did not find any general tools to verify the inequality

$$V_k(x) \overset{?}{\le} V_k(x_*).$$

This is however the crucial point in the pass from the pointwise estimate to the global one, and as a result Vidensky could not obtain the Markov-type inequality for an arbitary majorant. In a series of paper [21]-[23], he covered a number of particular cases where he suceeded to prove monotonicity of V_k using monotonicity of D_k.

Theorem 4.9 (Vidensky (1958-71)). *Let*

$$\begin{array}{lll} & \mu^2(x) = \prod_{i=1}^m (1+a_i^2 x^2), & k=1, \\ \textit{or} & \mu^2(x) = (1+a^2x^2)(1+b^2x^2), & k \ge 1, \\ \textit{or} & \mu^2(x) = 1+(a^2-1)x^2, & k \ge 1. \end{array}$$

Then $[V_k(x)]^2 = \sum_0^\infty c_{2i} x^{2i}$ *with* $c_{2i} \ge 0$, *whence*

$$M_{k,\mu} := \sup_{|p(x)| \le \mu(x)} \|p^{(k)}\| = \|\omega_\mu^{(k)}\|,$$

where $\omega_\mu = P_n$ *is the polynomial oscillating most between* $\pm\mu$.

These works of Vidensky remained largely unknown, and some of his results were rediscovered later. For example, the Markov-type inequality with circular majorant $\mu(x) = \sqrt{1-x^2}$ was reproved in 70s by Pierre and Rahman in [19], [16], [17]. Bojanov and Naidenov [24]-[25] proved the interlacing property for perfect splines and alike using arguments quite similar to those of Vidensky.

We close this section with our own observation about the explicit form of the function $V_k(x)$ which Vidensky did not try to compute. Now, there is no differential equation that allows us to find the exact formula for D_k for all n, k as in the previous section, but we can derive some recurrence relations instead.

Lemma 4.10. *For each* k

$$[V_k(x)]^2 = \frac{[p_k(x)]^2 + (1-x^2)[q_k(x)]^2}{(1-x^2)^{2k-1}},$$

where p_k *and* q_k *are polynomials from* $\mathcal{P}_{m+k-1}$ *and* $\mathcal{P}_{m+k-2}$, *respectively.*

Proof. The formula follows from the representation

$$\begin{aligned} &\left[P_{m+n}(x) + i\sqrt{1-x^2}Q_{m+n-1}(x)\right]^{(k)} \\ &\qquad = (-i)\left[\frac{p_k(x) + i\sqrt{1-x^2}q_k(x)}{(1-x^2)^{k-1/2}}\right][T_n(x) + iS_n(x)] \end{aligned}$$

which is derived from (4.9) by induction using the relation

$$T_n'(x) + iS_n'(x) = (-i)\frac{n}{\sqrt{1-x^2}}[T_n(x) + iS_n(x)].$$

Moreover, we have the recurrence formula

$$\begin{aligned} p_1(x) &= nP_m(x) + xQ_{m-1}(x) - (1-x^2)Q_{m-1}'(x), \\ q_1(x) &= nQ_{m-1}(x) + P_m'(x), \\ p_{k+1}(x) &= (1-x^2)[p_k'(x) + nq_k(x)] + (2k-1)xp_k(x); \\ q_{k+1}(x) &= (1-x^2)q_k'(x) + 2(k-1)xp_k(x) - np_k(x) \end{aligned}$$

However, it is not clear whether it is possible to extract from here information about momotonicity of V_k (assuming, say, that μ is monotone). □

5 Markov–Duffin–Schaeffer inequalities

5.1 Duffin–Schaeffer refinement for the discrete restrictions

In 1941, Duffin–Schaeffer [5] (now in alphabetical order) presented another proof of the Markov inequality that moreover strengthened Markov's result in two different directions. Namely, they showed that in order to reach the conclusion

$$|p^{(k)}(x)| \le T_n^{(k)}(1)$$

it is sufficient, instead of the uniform bound $\|p\| \le 1$, to assume that $|p(x)| \le 1$ at the $n+1$ points $x \in \{\cos\frac{\pi i}{n}\}_{i=0}^n$ only. At the same time they showed that, under this weaker assumption, the Markov inequality can be extended to the complex plane.

They started out by taking a rather general point of view. Namely, given a polynomial q with n distinct real zeros,

$$q(z) = c\prod_{\nu=1}^n (z - x_\nu), \quad q'(z) = q(z)\sum_{\nu=1}^n \frac{1}{z - x_\nu}, \tag{5.1}$$

they tried to figure out the class $\mathcal{K}$ of polynomials and the conditions on q for which the derivative q' takes the values larger than the derivative of any other polynomial $p \in \mathcal{K}$. By the Lagrange interpolating formula with the nodes (x_ν), we have

$$p'(z) = \sum_{\nu=1}^n \frac{p'(x_\nu)}{q'(x_\nu)}\frac{q(z)}{z - x_\nu}\,,$$

and it is suggestive to consider those $p \in \mathcal{P}_n$ that satisfy $|\frac{p'(x_\nu)}{q'(x_\nu)}| \le 1$, so that

$$p'(z) = q(z)\sum_{\nu=1}^n \frac{\epsilon_\nu}{z - x_\nu}\,, \quad \epsilon_\nu \in [-1, 1]. \tag{5.2}$$

It is clear that we may restrict ourselves to the polynomials for which $\epsilon_\nu = \pm 1$, in particular, for real x we obtain

$$|p'(x)| \le \sum_{\nu=1}^n \left|\frac{q(x)}{x - x_\nu}\right|\,. \tag{5.3}$$

Now, one needs to find a way to compare the two sums in (5.1) and (5.2), and Duffin–Schaeffer's choice was the following elementary lemma from complex analysis.

Lemma 5.1. *Let $p(z) = a_n z^n + \cdots a_0$ be any polynomial, and let $q(z) = b_n z^n + \cdots b_0$ be a polynomial with all its zeros lying to one side of a line ℓ in the complex plane. If*

$$|p(z)| \le |q(z)| \quad \textit{on } \ell,$$

then

$$|p^{(k)}(z)| \le |q^{(k)}(z)| \quad on\ \ell\,, \quad k = 1..n.$$

Theorem 5.2. *Let $q(z) = c\prod_{\nu=1}^n (z - x_\nu)$ with distinct $x_\nu \in \mathbb{R}$, and let $p \in \mathcal{P}_n$ be a polynomial that satisfies*

$$|p'(x)| \le |q'(x)| \quad \text{at the zeros of} \quad q.$$

If all zeros of q lie to the left of some $b \in \mathbb{R}$, and for some $\xi_0 \in \mathbb{R}$ we have

$$|q(\xi_0 + iy)| \le |q(b + iy)|, \quad \forall y \in \mathbb{R}\,, \tag{5.4}$$

then

$$|p^{(k)}(\xi_0 + iy)| \le |q^{(k)}(b + iy)|\,, \quad \forall y \in \mathbb{R}\,. \tag{5.5}$$

Proof. There is no loss of generality in assuming that $\xi_0 = 0$.

1) Set

$$\widehat{q}(z) := c\prod_{\nu=1}^n (z - |x_\nu|), \quad \text{so that} \quad \widehat{q}'(z) = \widehat{q}(z)\sum_{\nu=1}^n \frac{1}{z - |x_\nu|}.$$

Then, from (5.2), we derive

$$\begin{aligned} \left|\frac{p'(iy)}{q(iy)}\right| \quad &= \quad \left|\sum_{\nu=1}^n \frac{\epsilon_\nu}{iy - x_\nu}\right| = \left|\sum_{\nu=1}^n \frac{\epsilon_\nu(x_\nu + iy)}{x_\nu^2 + y^2}\right| = \left|\sum_{\nu=1}^n \frac{\epsilon_\nu x_\nu}{x_\nu^2 + y^2} + i\sum_{\nu=1}^n \frac{\epsilon_\nu y}{x_\nu^2 + y^2}\right| \\ &\le \quad \left|\sum_{\nu=1}^n \frac{|x_\nu|}{|x_\nu|^2 + y^2} + \sum_{\nu=1}^n \frac{iy}{|x_\nu|^2 + y^2}\right| = \left|\sum_{\nu=1}^n \frac{1}{iy - |x_\nu|}\right| = \left|\frac{\widehat{q}'(iy)}{\widehat{q}(iy)}\right|. \end{aligned}$$

Since clearly

$$|\widehat{q}(iy)| = |q(iy)|, \quad \forall y \in \mathbb{R}\,, \tag{5.6}$$

we conclude that

$$|p'(iy)| \le |\widehat{q}'(iy)|, \quad \forall y \in \mathbb{R}\,. \tag{5.7}$$

2) Now we are ready to apply Lemma 5.1. From (5.7), since all zeros of $\widehat{q}'$ lie to the right of the line $\ell = \{iy\}$, we obtain

$$|p^{(k)}(iy)| \le |\widehat{q}^{(k)}(iy)|, \quad \forall y \in \mathbb{R}\,. \tag{5.8}$$

Now we use Lemma 5.1 to evaluate $|\widehat{q}^{(k)}(iy)|$. From (5.6) and (5.4) (with $\xi_0 = 0$), it follows that $|\widehat{q}(iy)| \le |q(b + iy)|$ for all $y \in \mathbb{R}$, and because all zeros of q lie to the left of $\ell = i\mathbb{R}$, we conclude

$$|\widehat{q}^{(k)}(iy)| \le |q^{(k)}(b + iy)|, \quad \forall y \in \mathbb{R}\,, \tag{5.9}$$

and that together with (5.8) proves (5.5). □

Theorem 5.3 (Duffin–Schaeffer (1941)). *If $p \in \mathcal{P}_n$ satisfies*

$$|p(x)| \le 1, \quad x \in \{\cos \tfrac{\pi i}{n}\}_{i=0}^n ,$$

then

$$|p^{(k)}(x+iy)| \le |T_n^{(k)}(1+iy)|, \quad \forall x \in [-1,1], \quad \forall y \in \mathbb{R} . \tag{5.10}$$

Proof. Theorem 5.3 is reduced to Theorem 5.2 by means of the following statements.

Lemma 5.4. *If a polynomial $p \in \mathcal{P}_n$ satisfies*

$$|p(x)| \le |T_n(x)| \quad \textit{wherever} \quad |T_n(x)| = 1$$

then

$$|p'(x)| \le |T_n'(x)| \quad \textit{at the zeros of} \quad T_n .$$

Lemma 5.5. *We have*

$$|T_n(x+iy)| \le |T_n(1+iy)|, \quad \forall x \in [-1,1], \quad \forall y \in \mathbb{R} .$$

We omit the proofs, and make only short comments. The main remark is that, unlike the rather general Theorem 5.2, these proofs depend on specific properties of the Chebyshev polynomials.

1) The first lemma is derived by differentiating the Lagrange formula with the nodes $(\cos \frac{\pi i}{n})$,

$$p(x) = \sum_{\nu=0}^n \frac{p(t_\nu)}{\omega'(t_\nu)} \frac{\omega(x)}{x - t_\nu}, \qquad \omega(x) := (x^2-1)T_n'(x),$$

and using the differential equation $(x^2-1)T_n''(x) + xT_n(x) = n^2 T_n(x)$.

2) The second lemma is not that straightforward, and Duffin–Schaeffer's proof is a bit tricky and lengthy, where the specific form of the roots of T_n play an important role. □

Comment 5.6. Duffin–Schaeffer's original proof of Theorem 5.2 develops a bit differently from our presentation. They use Lemma 5.1 only once, in deriving the inequality (5.9) for $k = 1$, and then combine the latter with (5.7), thus proving Theorem 5.2 firstly for $k = 1$. They proceed further by induction on k, and for that they prove that

$$\begin{aligned} &\text{if} \quad |p^{(k)}(x)| \le |q^{(k)}(x)| \qquad \text{wherever} \quad q^{(k-1)}(x) = 0, \\ &\text{then} \quad |p^{(k+1)}(x)| \le |q^{(k+1)}(x)| \qquad \text{wherever} \quad q^{(k)}(x) = 0 . \end{aligned}$$

We cut this step and used Lemma 5.1 to derive both estiamtes (5.8) and (5.9) for all k at once.

Comment 5.7. Lemma 5.1 appeared originally in 1926 in Bernstein's monograph [47] as "Troisieme corollaire" on pp. 55-56, with a rather lengthy proof (if you move all the way through). Later, Bernstein also showed [48] that it is valid for a circle c instead of a line ℓ (by mapping c onto ℓ using a Möbius transform). Duffin–Schaeffer were perhaps unaware of this result and in their work gave their own short proof based on Rouche's theorem (without making it an independent statement). In 1947, de Bruijn [50] generalized the result for the boundary of any convex domain and made the proof even shorter. His proof is however too concise, so here is the one from Rivlin's book [52] on p. 142.

Proof. Since q has no zeros in the half-plane H, and $\deg p \le \deg q$, the function p/q is analytic in H, hence by the maximum principle $\max_{z\in H} |\frac{p(z)}{q(z)}| = \max_{z\in \ell} |\frac{p(z)}{q(z)}| \le 1$. Thus, for any $|\lambda| > 1$, the polynomial $p - \lambda q$ has no zeros in H, and by the Gauss-Lucas theorem the same its true for any of its derivative $p^{(k)} - \lambda q^{(k)}$, hence $|p^{(k)}(z)| \le |q^{(k)}(z)$ in $H \cup \ell$. □

Comment 5.8. The Duffin–Schaeffer inequality (5.10) makes not much sense for the points $z = x + iy$ outside the unit (or even smaller) disc, because for such z a better estimate can be obtained by simpler tools.

Let $q(x) = \prod_{\nu=1}^n (x - x_\nu)$ and let (τ_ν) satisfy $\tau_0 < x_1 < \tau_1 < \cdots < x_n < \tau_n$. Then, for any $p \in \mathcal{P}_n$ such that

$$|p(\tau_\nu)| \le |q(\tau_\nu)|,$$

we have the inequality

$$|p^{(k)}(z)| \le |q^{(k)}(z)|, \quad z \notin D,$$

where D is the open disc with (ξ, η) its diameter. Here ξ (resp. η) is the leftmost (rightmost) zero of the polynomial $\omega_n^{(k)}$ $(\omega_0^{(k)})$, where $\omega_i(x) = \frac{\omega(x)}{x-\tau_i}$ and $\omega(x) = \prod (x - \tau_i)$.

So, under the assumption of Theorem 5.3, we have

$$|p^{(k)}(x+iy)| \le |T_n^{(k)}(x+iy)|, \quad x + iy \notin D_{r_k}\,,$$

with r_k being the rightmost zero of the polynomial $[(x-1)T_n'(x)]^{(k)}$.

5.2 Duffin-Schaeffer inequalities with majorant

A natural question is whether the Duffin-Schaeffer refinement can be extended to the Markov inequalities with a majorant. Namely, given a majorant $\mu(x) \ge 0$, let $\omega_\mu \in \mathcal{P}_n$ be the polynomial oscillating most between $\pm\mu$ which is very likely to attain the supremum

$$M_{k,\mu} := \sup_{|p(x)| \le \mu(x)} \|p^{(k)}\|\,.$$

If that is the case, then for $\delta_* := (\tau_i^*)$, the set of the oscillating points of ω_μ, we define a Duffin-Schaeffer-type constant

$$D^*_{k,\mu} := \sup_{|p(x)|_{\delta_*} \le |\mu(x)|_{\delta_*}} \|p^{(k)}\| ,$$

and ask whether the two values are the same (as they are for $\mu \equiv 1$). Moreover, since for any μ we have $\|\omega_\mu^{(k)}\| \le M_{k,\mu} \le D^*_{k,\mu}$, we may try to solve the Duffin-Schaeffer problem even if the solution to the Markov problem is not known.

With Duffin-Schaeffer's general Theorem 5.2, all we have to do is to establish analogues of Lemmas 5.4-5.5 for the corresponding polynomial ω_μ. However, this turns out to be a rather difficult task.

First of all, the set $\delta_* = (\tau_i^*)$ where ω_μ touches the majorant μ becomes not that simple as with the case $\mu = 1$, or is even unknown in the explicit form. But even if you know it (say, as with $\mu(x) = (1-x^2)^{m/2}$) you have to go through a rather delicate analysis to get that

$$|p^{(m_0+1)}(x)| \le |\omega^{(m_0+1)}(x)| \quad \text{at the zeros of} \quad \omega^{(m_0)}$$

(it may also be not true for the derivatives of order $\le m_0$).

Secondly, as we mentioned, the inequality

$$|\omega^{(m_0)}(x+iy)| \le |\omega^{(m_0)}(1+iy)|$$

was not that easy to establish even for $\omega = T_n$, and this is another quite serious obstacle in getting Duffin-Schaeffer-type result even for the simplest majorants (using Theorem 5.2).

This explains why there were only two results obtained in this direction.

Theorem 5.9 (Rahman–Schmeisser (1988)).

$$\mu(x) = \sqrt{1-x^2} \;\Rightarrow\; \begin{cases} \|\omega_\mu^{(k)}\| = M_{k,\mu} < D_{k,\mu}\,, & k = 1; \\ \|\omega_\mu^{(k)}\| = M_{k,\mu} = D_{k,\mu}\,, & k > 1. \end{cases}$$

Theorem 5.10 (Rahman–Watt (1992)).

$$\mu(x) = 1 - x^2 \quad \Rightarrow \quad \|\omega_\mu^{(k)}\| = M_{k,\mu} = D_{k,\mu}\,, \quad k > 2.$$

Recently, Nikolov [38] suggested a method that allowed to prove the inequality

$$|\omega(x+iy)| \le |\omega(1+iy)|, \quad x \in [-1,1], \quad y \in \mathbb{R},$$

for a sufficiently large class of polynomials, namely the ultraspherical polynomials $\omega = P^{(\alpha,\alpha)}$ orthogonal with the weight $(1-x^2)^{\alpha-1/2}$ (where $\alpha > -1/2$). This could lead to the Duffin–Schaeffer inequalities with the majorant $\mu(x) = (1-x^2)^{-\alpha/2}$ which is however unbounded for $\alpha > 0$.

5.3 Another proof of Duffin–Schaeffer inequality

Duffin and Schaeffer derived their inequality starting with the Lagrange representation of the derivative of a polynomial p' based on the roots of an a priori given polynomial q. In 1992, we took a more natural approach [10] choosing as a starting point the Lagrange formula with exactly those points where the discrete restrictions on p are actually given. The main tool of this approach is the following lemma about polynomials with interlacing zeros.

Lemma 5.11 (V. Markov (1892)). *If the zeros of* $p(x) = \prod_{i=1}^n (x - s_i)$ *and* $q(x) = \prod_{i=1}^n (x - t_i)$ *interlace, i.e.,*

$$s_i \le t_i \le s_{i+1} \quad \textit{all } i,$$

then the zeros of $p^{(k)}(x) =: \prod_{i=1}^{n-k}(x - \xi_i)$ *and* $q^{(k)}(x) =: \prod_{i=1}^{n-k}(x - \eta_i)$ *interlace too (and, moreover, strictly):*

$$\xi_i < \eta_i < \xi_{i+1} \quad \textit{all } i.$$

There are many (short) proofs of this remarkable lemma, the simplest one is perhaps by Rivlin [52, p.125], but one may choose also from V. Markov [7, §34], Bojanov [24], or take that of Vidensky [20] that we gave in Lemma 4.8.

We will write $p \preceq q$ if the polynomials $p(x) = \prod_{i=1}^n (x - t_i)$ and $q(x) = \prod_{i=1}^n (x - s_i)$ have interlacing zeros, i.e., $t_i \le s_i \le t_{i+1}$. Then the Markov's lemma can be written as: $p \preceq q$ implies $p^{(k)} \prec q^{(k)}$.

Now we begin another version of the book-proof of Markov's inequality.

Book-proof. Given a polynomial $q \in \mathcal{P}_n$ and a sequence δ of $n+1$ points, we will study the value

$$\sup_{|p(x)|_\delta \le |q(x)|_\delta} |p^{(k)}(x)|\,, \quad x \in [-1, 1]\,,$$

and we want to find when it can be majorized by $\|q^{(k)}\|$. We obtain the Markov–Duffin–Schaeffer inequality by setting $\delta = (\cos \frac{\pi i}{n})$ and $q = T_n$.

Definition 5.12. Given $\delta = (\tau_i)_{i=0}^n$ on $[-1, 1]$, set

$$\omega(x) := \prod_{i=0}^n (x - \tau_i), \quad \omega_i(x) := \frac{\omega(x)}{x - \tau_i},$$

and let (η_j) and (ξ_j) be defined as

$$\eta_0 := -1, \quad \omega_0^{(k)}(x) =: c\prod_{j=1}^{n-k}(x - \eta_j),$$
$$\omega_n^{(k)}(x) =: c\prod_{j=1}^{n-k}(x - \xi_j), \quad \xi_{n-k+1} := +1.$$

For $k \in \mathbb{N}$, we define

$$\text{the Chebyshev intervals:} \quad e_j^T = [\eta_{j-1}, \xi_j], \quad e_\delta^T = \cup_{j=1}^{n-k+1} e_j^T,$$
$$\text{the Zolotarev intervals:} \quad e_j^Z = (\xi_j, \eta_j), \qquad e_\delta^Z = \cup_{i=1}^{n-k} e_j^Z.$$

Lemma 5.13. *For any k, δ, j, the intervals e_j^T and e_j^Z are non-empty and, on any Chebyshev interval, we have*

$$\operatorname{sgn} \omega_0^{(k)}(x) = \cdots = \operatorname{sgn} \omega_n^{(k)}(x) \quad on \quad e_j^T .$$

Proof. We have $\omega_n \preceq \cdots \preceq \omega_0$, hence $\omega_n^{(k)} \prec \cdots \prec \omega_0^{(k)}$. Thus, zeros of $\omega_n^{(k)}$ and $\omega_0^{(k)}$ strictly interlace, i.e., $\xi_j < \eta_j < \xi_{j+1}$, thus e_j^T and e_j^Z are well-defined. Further, the j-th zero of any $\omega_i^{(k)}$ is located between those of $\omega_n^{(k)}$ and $\omega_0^{(k)}$, which are ξ_j and η_j, respectively, i.e. on the Zolotarev interval, hence $\omega_i^{(k)}$ does not change its sign on the Chebyshev interval. It remains to notice that the leading coefficients of all ω_i's are equal 1, thus at their j-th zeros they change sign in the same way. □

Proposition 5.14. *Let $q(x) = \prod_{i=1}^n (x - t_i)$, and let $\delta = (\tau_i)_{i=0}^n$ be such that $\tau_{i-1} < t_i < \tau_i$, i.e., q alternates in sign on δ. If $p \in \mathcal{P}_n$ satisfies*

$$|p(x)| \le |q(x)| \quad on \quad \delta, \tag{5.11}$$

then, for any k,

$$|p^{(k)}(x)| \le |q^{(k)}(x)| \quad on \quad e_\delta^T .$$

Proof. By the Lagrange interpolation formula with nodes (τ_i),

$$p(x) = \sum_{i=0}^n \frac{p(\tau_i)}{\omega'(\tau_i)} \frac{\omega(x)}{x - \tau_i} = \sum_{i=0}^n \frac{p(\tau_i)}{\omega'(\tau_i)} \omega_i(x)$$

hence,

$$|p^{(k)}(x)| = \left| \sum_{i=0}^n \frac{p(\tau_i)}{\omega'(\tau_i)} \omega_i^{(k)}(x) \right| \le \sum_{i=0}^n \left| \frac{p(\tau_i)}{\omega'(\tau_i)} \right| \left| \omega_i^{(k)}(x) \right| \overset{(5.11)}{\le} \sum_{i=0}^n \left| \frac{q(\tau_i)}{\omega'(\tau_i)} \right| \left| \omega_i^{(k)}(x) \right| .$$

Now, both sequences $q(\tau_i)$ and $\omega'(\tau_i)$ alternate in sign, hence $\operatorname{sgn} \frac{q(\tau_i)}{\omega'(\tau_i)} =$ const for all i, and, by Corollary 5.13, on any Chebyshev interval e_j^T, we have $\operatorname{sgn} \omega_i^{(k)}(x) =$ const for all i as well. Thus,

$$\sum_{i=0}^n \left| \frac{q(\tau_i)}{\omega'(\tau_i)} \right| \left| \omega_i^{(k)}(x) \right| = \left| \sum_{i=0}^n \frac{q(\tau_i)}{\omega'(\tau_i)} \omega_i^{(k)}(x) \right| = |q^{(k)}(x)|,$$

i.e., $|p^{(k)}(x)| \le |q^{(k)}(x)|$. □

Theorem 5.15 (Shadrin (1992)). *Let q have all its zero in $[-1, 1]$. If*

$$|p(x)| \le |q(x)| \quad \text{at the zeros of} \quad (x^2 - 1)\, q'(x),$$

then

$$|p^{(k)}(x)| \le \max \left\{ |q^{(k)}(x)|,\ |\tfrac{1}{k}(x^2 - 1) q^{(k+1)}(x) + x q^{(k)}(x)| \right\}.$$

Proof. We have $\omega(x) = c\prod_{i=0}^n (x-\tau_i) = (x^2-1)q'(x)$, hence

$$\omega_0(x) = (x-1)q'(x), \quad \omega_0^{(k)}(x) =: \textstyle\prod_{j=1}^{n-k}(x-\eta_j),$$
$$\omega_n(x) = (x+1)q'(x), \quad \omega_n^{(k)}(x) =: \textstyle\prod_{j=1}^{n-k}(x-\xi_j)$$

and by the previous proposition,

$$|p^{(k)}(x)| \le |q^{(k)}(x)| \quad \text{on} \quad e_j^T = [\eta_{j-1}, \xi_j], \tag{5.12}$$

so that it is sufficient to prove that

$$|p^{(k)}(x)| < |r(x)| \quad \text{on} \quad e_j^Z = (\xi_j, \eta_j)\,,$$

where $r(x) := r_k(x) := \frac{1}{k}(x^2-1)q^{(k+1)}(x) + xq^{(k)}(x)$.

1) From the equalities $\omega_{n/0}^{(k)}(x) = (x\pm 1)q^{(k+1)}(x) + kq^{(k)}(x)$, it follows that

$$r(x) = \frac{1}{k}(x+1)\omega_0^{(k)}(x) - q^{(k)}(x) = \frac{1}{k}(x-1)\omega_n^{(k)}(x) + q^{(k)}(x).$$

From the definition of (ξ_j) and (η_j), we have $(x+1)\omega_0^{(k)}(x)\big|_{x\in\{\eta_j\}} = 0$ and $(x-1)\omega_n^{(k)}(x)\big|_{x\in\{\xi_j\}} = 0$, hence

$$r(\eta_j) = -q^{(k)}(\eta_j), \quad r(\xi_j) = +q^{(k)}(\xi_j), \quad \forall j.$$

2) Comparing these relations with (5.12) we obtain the inequalities

$$|p^{(k)}(x)| \le |r(x)| \quad \text{on} \quad (\eta_j), (\xi_j)$$

and, because q clearly does not change its sign on the Chebyshev interval $[\eta_{j-1}, \xi_j]$, we also get the following sign pattern: $\operatorname{sgn} r(\eta_{j-1}) = -\operatorname{sgn} r(\xi_j)$.

3) So, for any $\gamma \in [0,1]$, at the endpoints of $e_j^T = [\eta_{j-1}, \xi_j]$, we have

$$|\gamma p^{(k)}(x)| \le |r(x)|, \quad \operatorname{sgn} r(\eta_{j-1}) = -\operatorname{sgn} r(\xi_j),$$

hence each of the polynomials $r \pm \gamma p^{(k)} \in \mathcal{P}_{n-k+1}$ has a zero in each e_j^T, i.e., the complete set of $n-k+1$ zeros on e_δ^T.

4) Thus, for any $\gamma \in [0,1]$, there are no zeros of $r \pm \gamma p^{(k)}$ on $e_j^Z = (\xi_j, \eta_j)$, therefore, $|p^{(k)}(x)| < |r(x)|$ and we are done. □

Theorem 5.16. *Let q have all its zero in $[-1,1]$, and let*

$$|p(x)| \le |q(x)| \quad \text{at the zeros of} \quad (x^2-1)\,q'(x).$$

If

$$|\tfrac{1}{k}(x^2-1)q^{(k+1)}(x) + xq^{(k)}(x)| \le \|q^{(k)}\|\,,$$

then

$$\|p^{(k)}\| \le \|q^{(k)}\|\,.$$

Lemma 5.17. *For all n, and all $x \in [-1,1]$,*

$$|T_n^{(k)}(x)| \le T_n^{(k)}(1)\,, \tag{5.13}$$

$$\left|\tfrac{1}{k}(x^2-1)T_n^{(k+1)}(x) + xT_n^{(k)}(x)\right| \le T_n^{(k)}(1)\,, \tag{5.14}$$

$$T_n^{(k)}(1) = \tfrac{n^2[n^2-1^2]\cdots[n^2-(k-1)^2]}{1\cdot 3\cdots(2k-1)}\,. \tag{5.15}$$

Proof. We have

$$T_n(x) = \cos n\theta, \quad T_n'(x) = \frac{n\sin n\theta}{\sin\theta}, \quad x = \cos\theta.$$

1) The equality $\sin n\theta = 2\sin\theta\,[\cos(n-1)\theta + \cos(n-3)\theta + \cdots]$ implies that, for $k=1$,

$$T_n^{(k)}(x) = \sum_{i=0}^{n-k} a_{ik}T_i(x), \quad a_{ik} \ge 0. \tag{5.16}$$

Differentiating and expanding the terms on the right-hand side we arrive at the same result for all k. Obviously, the maximum of the sum occurs at $x=1$, hence (5.13)

2) The equality $\sin^2 n\theta + \cos^2 n\theta = 1$ transforms into the identity $\frac{1-x^2}{n^2}[T_n'(x)]^2 + [T_n(x)]^2 = 1$, whose differentiation gives

$$(x^2-1)T_n''(x) + xT_n'(x) = n^2T_n(x).$$

Differentiating further we obtain the formula

$$(x^2-1)T_n^{(k+1)}(x) + (2k-1)xT_n^{(k)}(x) = [n^2-(k-1)^2]T_n^{(k-1)}(x).$$

For $x=1$ this reads: $T_n^{(k)}(1) = \frac{n^2-(k-1)^2}{2k-1}T_n^{(k-1)}(1)$, and that proves (5.15).

3) If $k=1$, the left-hand side of (5.14) is evaluated as

$$|(x^2-1)T_n''(x) + xT_n'(x)| = |n^2T_n(x)| \le n^2 = T_n'(1),$$

i.e., (5.14) is true for $k=1$. If $k>1$, then we also have

$$|\tfrac{1}{k}(x^2-1)T_n''(x) + xT_n'(x)| = |\tfrac{1}{k}n^2T_n(x) + \tfrac{k-1}{k}xT_n'(x)| \le n^2 = T_n'(1),$$

and from

$$T_n^{(k)}(x) = \textstyle\sum_{i=0}^{n-k+1} b_{ik}T_i'(x)\,, \quad T_n^{(k+1)}(x) = \sum_{i=0}^{n-k+1} b_{ik}T_i''(x)\,, \quad b_{ik} \ge 0\,,$$

it follows that

$$\begin{aligned}\left|\tfrac{1}{k}(x^2-1)T_n^{(k+1)}(x) + xT_n^{(k)}(x)\right| &= \left|\textstyle\sum_i b_{ik}[\tfrac{1}{k}(x^2-1)T_i''(x) + xT_i'(x)]\right| \\ &\le \textstyle\sum_i b_{ik}\left|\tfrac{1}{k}(x^2-1)T_i''(x) + xT_i'(x)\right| \\ &\le \textstyle\sum_i b_{ik}T_i'(1) = T_n^{(k)}(1)\,.\end{aligned}$$

Theorem 5.18. *If $p \in \mathcal{P}_n$ satisfies*

$$|p(\cos \tfrac{\pi i}{n})| \le 1, \quad i = 0..n,$$

then

$$|p^{(k)}(x)| \le \max\left\{\left|T_n^{(k)}(x)\right|, |\tfrac{1}{k}(x^2-1)T_n^{(k+1)}(x) + xT_n^{(k)}(x)|\right\} \le T_n^{(k)}(1)\,.$$

Comment 5.19. Kalliomiemi [28] was the first to notice that the discrete restrictions of Duffin–Schaeffer imply the same pointwise estimate

$$|p^{(k)}(x)| \le |T_n^{(k)}(x)| \quad \text{on the Chebyshev set } e^T,$$

as in the case of the stronger uniform restrictions $\|p\| \le 1$. He derived it using the Voronovskaya criterion for the norm of the linear functional $\mu(p) = p^{(k)}(x)$, and such a result could have been easily extracted already from V. Markov's work, but nobody before paid attention to this fact.

Proposition 5.14 (which appeared in [10]) is from the same "very simple, but not noticed before" class. For example, in [52, pp. 125-127], Rivlin gives exactly the same statements as Lemmas 5.11 and 5.13 (they are going back to V. Markov), and even uses them to establish the pointwise estimate outside the interval, yet he does not notice that they also provide the pointwise estimate on the Chebyshev intervals.

Comment 5.20. Let

$$D_k^*(x) := \sup_{|p(x)|_{\delta_*} \le 1} |p^{(k)}(x)|\,. \tag{5.17}$$

where $\delta_* = (\tau_i^*) = (\cos \frac{\pi i}{n})$. This is the exact upper bound for the value of the k-th derivative of a polynomial p under the discrete Duffin–Schaeffer restrictions.

By the interlacing property, at any given point $x \in [-1,1]$, with some $i = i_x$,

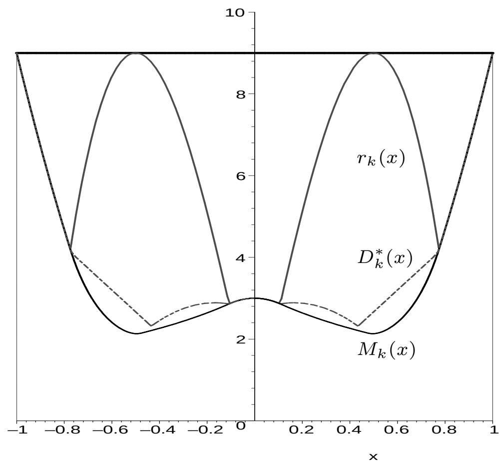
$$\operatorname{sgn} \omega_0^{(k)}(x) = \cdots = \operatorname{sgn} \omega_i^{(k)}(x) = -\operatorname{sgn} \omega_{i+1}^{(k)}(x) = \cdots = -\operatorname{sgn} \omega_n^{(k)}(x)$$

and from the Lagrange formula it follows that the set of the extremal polynomials for the pointwise problem (5.17) consists of those $p_s \in \mathcal{P}_n$ that satisfy

$$\operatorname{sgn} p_s(\tau_i) = -\operatorname{sgn} p_s(\tau_{i+1}), \quad i \ne s, \qquad \operatorname{sgn} p_s(\tau_s) = \operatorname{sgn} p_s(\tau_{s+1})\,.$$

The set of these polynomials may be viewed as a discrete analogue of the one-parameter family of Zolotarev polynomials.

The next figure illustrates how the graphs of D_k^*, M_k and our majorant r_k relate to each other.



1) On the Chebyshev intervals:

$$M_k(x) = D_k^*(x) = |T_n^{(k)}(x)|\,.$$

2) On the Zolotarev intervals

$$M_k(x) < D_k^*(x) < |r_k(x)|\,.$$

5.4 Duffin-Schaeffer inequalities for polynomials and the Landau-Kolmogorov problem

Theorem 5.16 is much more convenient for applications than Theorem 5.2 of Duffin-Schaeffer because, of the two assumptions

$$\begin{gathered}|\tfrac{1}{k}(x^2-1)q^{(k+1)}(x) + xq^{(k)}(x)| \le |q^{(k)}(1)|\,, \\ |q(x+iy)| \le |q(1+iy)|\,,\end{gathered} \tag{5.18}$$

the first one is much easier to verify.

1) Theorem 5.16 was used to obtain many other estimates, the so-called Duffin-Schaeffer (DS-) inequalities for polynomials.

Definition 5.21. The polynomial $q(x) = \prod_{i=1}^n (x - t_i)$ and the mesh $\delta = (\tau_i)_{i=0}^n$ such that $\tau_0 \le t_1 \le \tau_1 \le \cdots \le t_n \le \tau_n$ are said to admit the DS-inequality if

$$\sup_{|p(x)|_\delta \le |q(x)|_\delta} \|p^{(k)}\| = \|q^{(k)}\|\,. \tag{5.19}$$

Two typical results (see [37] for further references).

a) Bojanov and Nikolov [35] showed that (5.19) is true for the ultraspherical polynomial $q = P_n^{(\alpha,\alpha)}$, and for the mesh δ consisting of the points of its local extrema. Actually, we may take any polynomial q whose $(k-1)$-st derivative has a positive Chebyshev expansion, i.e., $q^{(k-1)} = \sum a_i T_i$ with $a_i \ge 0$.

b) Milev and Nikolov [36] obtained a refinement of Schur's inequality for the polynomials vanishing at the endpoints. Let $\widehat{T}_n(x) := T_n(x \cos \frac{\pi}{2n})$ be the Chebyshev polynomial stretched to satisfy $\widehat{T}_n(\pm 1) = 0$ and let (τ_i) be the points of its local extrema. Then

$$p(\pm 1) = 0, \quad |p(\tau_i)| \le 1 \quad \Rightarrow \quad \|p^{(k)}\| \le \widehat{T}_n^{(k)}(1)\,.$$

2) In principle, Theorem 5.15 can be applied to obtain Duffin-Schaeffer inequalities with majorant, but it is unlikely that one can get here anything more than results for $\mu(x) = (1-x^2)^{m/2}$ for small values of m.

3) Whereas Duffin-Schaeffer's inequality gives only a uniform bound, Theorem 5.18 provides also the pointwise estimate inside the interval $[-1,1]$. This was used by Eriksson [27] to derive the Landau-Kolmogorov inequality

$$\|f^{(k)}\| \le \frac{n-k}{n}\frac{\|T_n^{(k)}\|}{\|T_n\|}\,\|f\| + \frac{k}{n}\frac{\|T_n^{(k)}\|}{\|T_n^{(n)}\|}\,\|f^{(n)}\|\,.$$

5.5 Erroneous proof by Duffin–Karlovitz

In 1984, Duffin and Karlovitz [4] revisited the Duffin-Schaeffer inequality and tried to generalize it from polynomials to arbitrary Chebyshev systems.

They started with a formalization of the problem. The discrete restriction $|p(x)| \le 1$ on a set of $n+1$ points in $[-1,1]$ is equivalent to bounding by the node norm

$$\|f\|_\delta := \max_i |f(\tau_i)|$$

that is defined for any given knot-sequence $\delta = (\tau_i)_{i=0}^n$, where

$$-1 \le \tau_0 < \tau_1 < \ldots < \tau_n \le 1.$$

Problem 5.22 (Duffin–Schaeffer inequality). For integer n, k, find

$$D_k := \inf_{\delta\in[-1,1]}\ \sup_{\|p\|_\delta\le 1} \|p^{(k)}\|\,. \tag{5.20}$$

In these notations the Markov–Duffin–Schaeffer results state that

$$M_k = D_k = \|T_n^{(k)}\|,$$

and the Chebyshev polynomial T_n, which equioscillates $(n+1)$ times between ± 1, is extremal for both problems. In particular, the $n+1$ points of its equioscillation form the set δ_* giving the infimum in (5.20). Duffin and Karlovitz tried to find out which properties of T_n are crucial for such a result, and they came to the following conclusion.

Theorem 5.23 (Duffin–Karlovitz (1984)). *Let $p_* \in \mathcal{P}_n$ and $\delta_* = (\tau_i^*)$ give the infimum for D_k, i.e.*

$$D_k = \|p_*^{(k)}\|\,, \quad \|p_*\|_{\delta_*} = 1.$$

Then $\tau_0^* = -1$, $\tau_n^* = +1$, and

$$p_*'(\tau_i^*) = 0, \quad i = 2, \ldots, n-1.$$

"Proof". Duffin and Karlovitz reasoned as follows.

1) Denote by ℓ_i the Lagrange fundamental polynomials corresponding to the knot-sequence δ. Then, since $p(x) = \sum_{i=0}^n p(\tau_i)\ell_i(x)$, we have

$$D_{k,\delta} := \sup_{x\in[-1,1]} \sup_{\|p\|_\delta=1} |p^{(k)}(x)| = \sup_{x\in[-1,1]} \sum_{i=0}^n |\ell_i^{(k)}(x)| =: \sum_{i=0}^n |\ell_i^{(k)}(x_\delta)|\,,$$

and it is clear that the polynomial that attains the value $D_{k,\delta}$ is given by

$$p_\delta(x) = \sum_{i=0}^n p_\delta(\tau_i)\ell_i(x)\,, \qquad p_\delta(\tau_i) = \operatorname{sgn} \ell_i^{(k)}(x_\delta) = \pm 1, \tag{5.21}$$

so that

$$D_{k,\delta} = \|p_\delta^{(k)}\| = |p_\delta^{(k)}(x_\delta)|\,.$$

2) Two remarks. Firstly, the polynomials $\ell_i^{(k)}$ vanish of course at certain x, but one can show that, at the point $x = x_\delta$ that is a local maximum of the polynomial $p_\delta^{(k)}$, all the values $\ell_i^{(k)}(x_\delta)$ are non-zero (as we wrote in (5.21)). Secondly, an optimal δ_* contains $n+1$ disinct nodes (for if the distance between two consecutive nodes in δ tends to zero, then, the value $D_{k,\delta}$ becomes arbitrary large).

3) Let p_* and δ_* be optimal for D_k, i.e.

$$D_k = \inf_\delta \|p_\delta^{(k)}\| = \|p_*^{(k)}\| = p_*^{(k)}(x_*)\,, \quad \|p_*\|_{\delta_*} = 1.$$

Now, we perturb δ_* by an amount ϵ, i.e., $\tau_i^\epsilon = \tau_i^* \pm \epsilon_i$, and let

$$D_{k,\epsilon} = \|p_\epsilon^{(k)}\| = p_\epsilon^{(k)}(x_\epsilon)\,.$$

Then the inequality $D_k \le D_{k,\epsilon}$ reads

$$p_*^{(k)}(x_*) \le p_\epsilon^{(k)}(x_\epsilon)\,. \tag{5.22}$$

Since $p_*^{(k)}$ has a global maximum at x_*, we also have $p_*^{(k)}(x_\epsilon) \le p_*^{(k)}(x_*)$, hence

$$p_*^{(k)}(x_\epsilon) \le p_\epsilon^{(k)}(x_\epsilon). \tag{5.23}$$

4) We may assume that, if $\epsilon \to 0$, then $x_\epsilon \to x_*$, $p_\epsilon \to p_*$, $\ell_{i,\epsilon} \to \ell_{i,*}$, therefore

$$\operatorname{sgn} p_*(\tau_i^\epsilon) = \operatorname{sgn} p_*(\tau_i^*) \overset{(5.21)}{=} \operatorname{sgn} \ell_{i,*}^{(k)}(x_*) = \operatorname{sgn} \ell_{i,\epsilon}^{(k)}(x_\epsilon) \overset{(5.21)}{=} \operatorname{sgn} p_\epsilon(\tau_i^\epsilon) \ne 0. \tag{5.24}$$

5) Now suppose that, for some i_0,

$$|p_*(\tau_{i_0}^*)| = 1, \quad p_*'(\tau_{i_0}^*) > 0.$$

Take $\tau_{i_0}^\epsilon = \tau_{i_0}^* + \epsilon_{i_0}$ and $\tau_i^\epsilon = \tau_i^*$ otherwise, so that

$$|p_*(\tau_{i_0}^\epsilon)| > 1 = |p_\epsilon(\tau_{i_0}^\epsilon)|\,, \quad p_*(\tau_i^\epsilon) = p_\epsilon(\tau_i^\epsilon) = \pm 1.$$

Then, taking in account the sign pattern in (5.24), we obtain

$$|p_*^{(k)}(x_\epsilon)| = \left|\sum_{i=0}^n p_*(\tau_i^\epsilon)\ell_{i,\epsilon}^{(k)}(x_\epsilon)\right| > \left|\sum_{i=0}^n p_\epsilon(\tau_i^\epsilon)\ell_{i,\epsilon}^{(k)}(x_\epsilon)\right| = |p_\epsilon^{(k)}(x_\epsilon)|\,,$$

a contradiction to (5.23). Similarly, if $p'(\tau_{i_0}^*) < 0$, then we take $\tau_{i_0}^\epsilon = \tau_{i_0}^* - \epsilon_{i_0}$, and arrive at the same contradiction. □

It is a very nice "proof" and it is not easy to find what is wrong. I provide my explanation in the comments below, so that an interested reader may attempt this exercise.

Comment 5.24. The proof is correct if we assume that, for an optimal δ_*, there is a unique optimal polynomial p_*, and it seems that Duffin & Karlovitz overlooked that this could not be the case.

Formally, wrong is the sequel of the arguments. In Step 4, we *may* assume that if $\epsilon \to 0$, then the sign pattern (5.24) is valid, but one should add "going to a subsequence if necessary". And going to a subsequence of (ϵ) (with p_* and x_* fixed) means that, in Step 5, we are not free to choose ϵ as we want to. Say, if $p_*'(\tau_{i_0}) > 0$, then a subsequence may turn out to be with the entries $\epsilon_{i_0} < 0$.

The assumption that an optimal p_* satisfies $p_*'(\tau_{i_0}) > 0$ is not contradictory if (and only if), for the sequence of δ_ϵ defined as in Step 5, the sequences of p_ϵ and x_ϵ would tend to some other $\widehat{p}_*$ and $\widehat{x}_*$, respectively, which are also optimal for D_k. Moreover, such $\widehat{p}_*$ must satisfy $\widehat{p}_*'(\tau_{i_0}) < 0$

Comment 5.25. Theorem 5.23 is of course true for polynomials (by Duffin–Schaeffer's inequality). It may well be true for the Chebyshev systems, although Duffin–Karlovitz failed to prove it. However, its analogue for the Duffin–Schaeffer problem with majorant is no longer true. Consider the two values

$$M_{k,\mu} = \sup_{|p(x)|\le\mu(x)} \|p^{(k)}\|\,, \qquad D_{k,\mu} := \inf_{\delta\in[-1,1]} \sup_{|p(x)|_\delta\le\mu(x)_\delta} \|p^{(k)}\|$$

Then, for the majorant $\mu(x) = \sqrt{1-x^2}$, and for $k=1$ we have

$$D_{1,\mu} > M_{1,\mu} = \|\omega_\mu'\|\,.$$

This is the result of Rahman–Schmeisser mentioned in Theorem 5.9.

5.6 Inequality for the oscillating polynomials

Generally, in the Markov–Duffin–Schaeffer problem with majorant we want to find the values

$$M_{k,\mu} = \sup_{|p(x)|\le\mu(x)} \|p^{(k)}\|\,, \quad D_{k,\mu} := \inf_{\delta\in[-1,1]} \sup_{|p(x)|_\delta\le|\mu(x)|_\delta} \|p^{(k)}\|\,.$$

For any $\mu \ge 0$ there is a unique polynomial $\omega_\mu \in \mathcal{P}_n$, the so-called snake-polynomial, that oscillates $n+1$ times between $\pm\mu$, i.e., such that

$$|\omega_\mu(x)| \le \mu(x)\,, \quad x \in [-1,1]\,,$$

and on some set of $n+1$ points $\delta_* = (\tau_i^*)_{i=0}^n$ we have

$$\omega_\mu(\tau_i^*) = (-1)^i \mu(\tau_i^*)\,.$$

The question of interest is for which majorants μ it is this polynomial ω_μ that gives the supremum to both values above (as in the case $\mu \equiv 1$), in particular, whether it is the set δ_* that gives the infimum to $D_{k,\mu}$. Notice that, for any majorant μ,

$$\|\omega_\mu^{(k)}\| \le M_{k,\mu} \le D_{k,\mu} \le D_{k,\mu}^*,$$

where

$$D_{k,\mu}^* := \sup_{|p(x)|_{\delta_*} \le |\mu(x)|_{\delta_*}} \|p^{(k)}\|\,.$$

so it may well be sufficient to estimate from above the value $D_{k,\mu}^*$ only. However, even with the simplest majorants, the location of the nodes τ_i^* is not known, and we have to find some general arguments (as Duffin–Karlovitz tried to).

In 1996, we tried [11] to revive approach of Duffin-Karlovitz, where instead of varying knots along the graph of the majorant μ, we decided to vary them along the graph of ω_μ.

It is clear that the snake-polynomial ω_μ has n zeros inside the interval $[-1,1]$, $\omega_\mu(x) = c\prod_{i=1}^n (x-t_i)$, and that these zeros interlace with (τ_i^*), i.e.,

$$-1 \le \tau_0^* < t_1 < \tau_1^* < \cdots < t_n < \tau_n^* \le 1\,.$$

Denote by Δ_ω the class of knot-sequences $\delta = (\tau_i)$ with the same interlacing properties. Then

$$\begin{aligned} D_{k,\mu}^* &= \sup_{|p(x)|_{\delta_*} \le |\mu(x)|_{\delta_*}} \|p^{(k)}\| \\ &= \sup_{|p(x)|_{\delta_*} \le |\omega(x)|_{\delta_*}} \|p^{(k)}\| \le \sup_{\delta\in\Delta_\omega} \sup_{|p(x)|_{\delta} \le |\omega(x)|_{\delta}} \|p^{(k)}\| =: S_{k,\omega}\,, \end{aligned}$$

and we may try to evaluate the value $S_{k,\omega}$ in terms of $\|\omega^{(k)}\|$. It turns out that the pointwise problem,

$$S_{k,\omega}(x) := \sup_{\delta\in\Delta_\omega} \sup_{|p(x)|_{\delta} < |\omega(x)|_{\delta}} |p^{(k)}(x)|\,,$$

has a remarkable solution.

Proposition 5.26 (Shadrin (1996)). *Let* $\omega(x) = \prod_{i=1}^n (x-t_i)$, $t_i \in [-1,1]$. *Then*

$$S_{k,\omega}(x) = \max\left\{\, |\omega^{(k)}(x)|,\ \max_i |\phi_i^{(k)}(x)| \right\},$$

where

$$\phi_i(x) := \omega(x)\frac{1 - xt_i}{x - t_i}, \quad i = 1, \ldots, n.$$

Proof. Our original proof followed the idea of Duffin–Karlovits: we showed that variation of any single knot $\tau_i \in [t_i, t_{i+1}]$ does not result in a local extremum of the value $|p^{(k)}(x)|$, hence the value $S_{k,\omega}$ is achieved when τ_i is either t_i ot t_{i+1}. A simpler proof was given later by Nikolov [37]. $\square$

The polynomials ϕ_i are quite interesting. They have the same zeros as ω except one t_i, and because of the factor $\frac{1-xt}{x-t}$ they satisfy the inequalities

$$|\phi_i(z)| > |\omega(z)|, \quad z \in D_1, \qquad |\phi_i(z)| \le |\omega(z)|, \quad z \notin D_1,$$

where D_1 is the unit open disc in the complex plane.

From this proposition and considerations at the beginning of the section we obtain the statement that gives a new way of deriving Markov–Duffin-Schaeffer inequalities with a majorant.

Theorem 5.27. *Given a majorant $\mu \ge 0$, let $\omega_\mu \in \mathcal{P}_n$ be the corresponding snake-polynomial. If*

$$\max_i \|\phi_i^{(k)}\| \le \|\omega_\mu^{(k)}\|, \tag{5.25}$$

then

$$M_{k,\mu} = D_{k,\mu} = \|\omega_\mu^{(k)}\| \quad \big(= S_{k,\omega}\big). \tag{5.26}$$

An advantage of studying the inequality (5.25) is that this is purely a polynomial problem on the class of polynomials ω having all their zeros in $[-1, 1]$, with quite a simple and explicitly given polynomials ϕ_i involved. These polynomials may be viewed as the most extreme case of the Zolotarev-like polynomials.

However, it was only recently when some real improvements have been made.

1) Nikolov [39] proved that

$$\omega = T_n \quad \Rightarrow \quad (5.25) \quad (\text{hence } (5.26) \text{ for } \mu \equiv 1).$$

This gives one more proof of the classical Markov–Duffin–Schaeffer inequality.

2) Recently, in our joint paper with Nikolov [32], we extended this result:

$$\omega^{(k-1)} = \sum a_i T_i, \quad a_i \ge 0 \quad \Rightarrow \quad (5.25) \quad (\text{hence } (5.26)).$$

This allows to establish the Markov–Duffin–Schaeffer inequalities for a large class of majorants, e.g.

$$\begin{aligned} & \mu^2(x) = \textstyle\prod_{i=1}^m (1 + a_i^2 x^2), && k \ge 1; \\ \text{or} \quad & \mu^2(x) = (1 - x^2)^m, && k > m. \end{aligned}$$

This improves results of Vidensky (Theorem 4.9) and Pierre–Rahman (Theorem 2.11).

5.7 Conclusion

More than a hundred years have passed since Vladimir Markov, "a student of Sankt-Petersburg University", proved his inequality. Since then it has received a dozen alternative proofs, hundreds of generalizations and it is still a lively part of Approximation Theory. So much power in just a single line:

$$\|p^{(k)}\| \le \|T_n^{(k)}\|\,\|p\|, \quad \forall p \in \mathcal{P}_n.$$

Acknowledgements. A part of this research was performed in Bonn in 1996, supported by the Alexander von Humboldt Foundation.

It is a pleasure to thank Brad Baxter, Carl de Boor and Arieh Iserles for their comments on a draft of this paper.

References

Twelve proofs of the Markov innequality

[1] S. N. Bernstein, On the V. A. Markov theorem, Trudy Leningr. Industr. In-ta, no 5, razdel fiz-matem nauk, 1 (1938), 8–13 = Collected Works, v.1, 281–286 (Russian) = East J. Approx. 2 (1996), no. 2, 245–251.

[2] B. Bojanov, Another proof of V. Markov's inequality, in [15], §3.1, pp. 39–46.

[3] A. Ja. Dubovickii, A. A. Miljutin, Second variations in extremal problems with constraints, Dokl. Akad. Nauk SSSR 160 (1965), 18–21 (Russian) = Soviet Math. Dokl. 6 (1965), 12–16.

[4] R. J. Duffin, L. A. Karlovitz, The Markoff–Duffin–Schaeffer inequalities abstracted, Proc. Nat. Acad. Sci. USA 82 (1985), 955–957.

[5] R. J. Duffin, A. S. Shaeffer, A refinement of an inequality of the brothers Markoff, Trans. Amer. Math. Soc. 50 (1941), 517–528.

[6] V. V. Gusev, Derivative functionals of an algebraic polynomial and V. A. Markov's theorem, Izv. Akad. Nauk SSSR, Ser. Math. 25 (1961), 371-384 (Russian) = Appendix to [56], pp. 179–197.

[7] V. Markov, On functions which deviate least from zero in a given interval, St-Petersburg, 1892 (Russian), available as a pdf-file at HAT [57] = Math. Ann. 77 (1916), 213–258 (German translation with small abbreviations).

[8] E. Mohr, Elementarer Beweis einer Ungleichung von W. A. Markov, Tensor 14 (1963), 71–85 (German).

[9] A. S. Schaeffer, R. J. Duffin, On some inequalities of S. Bernstein and W. Markoff for derivatives of polynomial, Bull. Amer. Math. Soc. 44 (1938), 289–297.

[10] A. Yu. Shadrin, Interpolation with Lagrange polynomials. A simple proof of Markov inequality and some of its generalizations, Approx. Theory Applic. 8 (1992), no. 3, 51–61.

[11] A. Yu. Shadrin, On Markov-Duffin-Schaeffer inequalities with weight, a manuscript, 1996.

[12] V. M. Tikhomirov, Inequalities of Bernstein and A. A. Markov, in [54], §2.3.3, pp. 109–113 (Russian)

Surveys

[13] R. P. Boas, Inequalities for the derivatives of polynomials, Math. Mag. 42 (1969), 165–174.

[14] R. P. Boas, Extremal problems for polynomials, Amer. Math. Monthly, 85 (1978), no. 6, 473–475.

[15] B. Bojanov, Markov-type inequalities for polynomials and splines, in *Approximation Theory*, X (Nashville, TN, 2002), pp. 31–90, Vanderbilt University Press, 2002.

Markov inequalities with majorant

[16] R. Pierre, Q. I. Rahman, On a problem of Turan about polynomials, Proc. Amer. Math. Soc. 56 (1976), 231–238.

[17] R. Pierre, Q. I. Rahman, On a problem of Turan about polynomials. II, Canad. J. Math. 33 (1981), no. 3, 701–733.

[18] R. Pierre, Q. I. Rahman, G. Schmeisser, On polynomials with curved majorants, J. Approx. Theory 57 (1989), no. 2, 211–222.

[19] Q. I. Rahman, On a problem of Turan about polynomials with curved majorants, Trans. Amer. Math. Soc. 163 (1972), 447–455.

[20] V. S. Videnskii, On estimates of derivatives of a polynomial, Izvestiya Akad. Nauk SSSR, Ser. Mat. 15 (1951), 401–420 (Russian).

[21] V. S. Videnskii, A generalization of V. A. Markoff inequalities, Dokl. Akad. Nauk SSSR 120 (1958), 447–449 (Russian).

[22] V. S. Videnskii, Generalizations of Markov's theorem on the evaluation of a polynomial derivative, Doklady Akad. Nauk SSSR 125 (1959), 15–18 (Russian).

[23] V. S. Videnskii, Least upper bounds for the successive derivatives of a polynomial on an interval, Izv. Vys. Uchebn. Zaved., ser. Mat. 106 (1971), no. 3, 18–22 (Russian).

Markov inequalities for splines

[24] B. Bojanov, Markov interlacing property for perfect splines, J. Approx. Theory 100 (1999), no. 1, 183–201.

[25] B. Bojanov, N. Naidenov, Exact Markov-type inequalities for oscillating perfect splines, Constr. Approx. 18 (2002), no. 1, 37–59.

[26] S. Karlin, Generalized Markov Bernstein type inequalities for spline functions, in *Studies in Spline Functions and Approximation Theory*, pp. 461–484, Academic Press, New York, 1976.

Landau-Kolmogorov problem on a finite interval

[27] B.-O. Eriksson, Some best constants in the Landau inequality on a finite interval, J. Approx. Theory 94 (1998), no. 3, 420–454.

[28] H. Kallioniemi, On bounds for the derivatives of a complex-valued function of a compact interval, Math. Scand. 39 (1976), no. 2, 295–314.

[29] H. Kallioniemi, The Landau problem on compact intervals and optimal numerical differentiation, J. Approx. Theory 63 (1990), no. 1, 72–91.

[30] A. Pinkus, Some extremal properties of perfect splines and the pointwise Landau problem on the finite interval, J. Approx. Theory 23 (1978), no. 1, 37–64.

[31] A. Shadrin, To the Landau-Kolmogorov problem on a finite interval, in *Open Problems in Approximation Theory* (B. Bojanov, Ed.), SCT Publishing, Singapore, 1994, pp. 192-204.

Markov-Duffin-Schaeffer inequalities with majorant

[32] G. Nikolov, A. Shadrin, Markov–Duffin–Schaeffer inequalities with majorant, a manuscript, 2004.

[33] Q. I. Rahman, G. Schmeisser, Markov-Duffin-Schaeffer inequality for polynomials with a circular majorant, Trans. Amer. Math. Soc. 310 (1988), no. 2, 693–702.

[34] Q. I. Rahman, A. O. Watt, Polynomials with a parabolic majorant and the Duffin-Schaeffer inequality, J. Approx. Theory 69 (1992), no. 3, 338–354.

Duffin-Schaeffer-type inequalities for polynomials

[35] B. Bojanov, G. Nikolov, Duffin and Schaeffer type inequality for ultraspherical polynomials, J. Approx. Theory 84 (1996), no. 2, 129–138.

[36] L. Milev, G. Nikolov, On the inequality of I. Schur, J. Math. Anal. Appl. 216 (1997), no. 2, 421–437.

[37] G. Nikolov, Inequalities of Duffin-Schaeffer type, SIAM J. Math. Anal. 33 (2001), no. 3, 686–698.

[38] G. Nikolov, An extension of an inequality of Duffin and Schaeffer, a manuscript, 2003.

[39] G. Nikolov, Inequalities of Duffin-Schaeffer type. II, a manuscript, 2003.

Lagrange interpolation

[40] N. P. Korneichuk, Approximation of functions and their derivatives by interpolation splines, Dokl. Akad. Nauk SSSR 264 (1982), no. 5, 1063–1066 (Russian).

[41] A. Shadrin, Error bounds for Lagrange interpolation. J. Approx. Theory 80 (1995), no. 1, 25–49.

Zolotarev polynomials

[42] V. A. Malyshev, Algebraic solution of the Zolotarev problem, Algebra i Analiz 14 (2002), no. 4, 238–240 (Russian) = St. Petersburg Math. J. 14 (2003), no. 4, 711–712.

[43] F. Peherstorfer, Orthogonal and Chebyshev polynomials on two intervals, Acta Math. Hungar. 55 (1990), no. 3-4, 245–278.

[44] M. L. Sodin, P. M. Yuditskii, Algebraic solution of E. I. Zolotarev and N. I. Akhiezer problems on polynomials that deviate least from zero, Teor. Funktsii Funktsional. Anal. i Prilozhen. no. 56 (1991), 56–64 (Russian) = J. Math. Sci. 76 (1995), no. 4, 2486–2492.

Miscellaneous books and Markov-type results

[45] S. Bernstein, Sur l'ordre de la meilleure approximation etc., Mem. Cl. Sci. Acad. Roy. Belg. 4 (1912), 1–103 (French) = Collected Works, v.1, 8–105 (Russian).

[46] S. Bernstein, Remarques sur l'inegalite de Wladimir Markoff, Soobsch. Khark. Matem. Ob-va, 14 (1913), 81–87 (French) = Collected Works, v.1, 151–156 (Russian).

[47] S. Bernstein, *Lecons sur les proprietes extremales, etc.*, Paris, 1926 (French).

[48] S. Bernstein, Sur la limitation des derivees des polynomes, C. R. Acad. Sci, Paris 190 (1930), 338–340 (French) = Collected Works, v. 1, 497–499 (Russian)

[49] S. Bernstein, *Collected Works*, Izd-vo Akad. Nauk SSSR, v. 1, 1952 (Russian).

[50] N. G. de Bruijn, Inequalities concerning polynomials in the complex domain, Indag. Math. 9 (1947), 591–598.

[51] A. Markov, On a question by D. I. Mendeleev, Zapiski Imper. Akad. Nauk , 62 (1890), 1–24 = Selected Works, Gostechizdat, Moscow, 1948, pp. 51–75 (Russian), both articles are available as pdf-files at HAT [57], together with the English translation by de Boor–Holtz.

[52] T. J. Rivlin, *Chebyshev Polynomials*, Second edition, Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1990, xvi+249 pp

[53] A. Schönhage, *Approximationstheorie*, Walter de Gruyter & Co., Berlin New York, 1971, 212 pp (German).

[54] V. M. Tikhomirov, *Some Questions in Approximation Theory*, Izdat. Moskov. Univ., Moscow, 1976, 304 pp (Russian).

[55] E. V. Voronovskaja, The functional of the first derivative and improvement of a theorem of A. A. Markov, Izv. Akad. Nauk SSSR, ser. Matem. 23 (1959), 951–962 (Russian) ≈ [56], Chapter III, pp. 156–167 (English translation of a restructured version).

[56] E. V. Voronovskaja, *The Functional Method and its Applications*, Izdat. LEIS, Leningrad, 1963 (Russian) = Translations of Mathematical Monographs, Vol. 28 Amer. Math. Soc., Providence, R.I., 1970, 203 pp.

History of approximation theory

[57] A. Pinkus, C. de Boor, A Homepage on the History of Approximation Theory, <www.math.technion.ac.il/hat> = <www.cs.wisc.edu/∼deboor/HAT>.