Contents

Methodological papers

ANNALES MATHEMATICAE ET INFORMATICAE 33. (2006)

ANNALES MATHEMATICAE ET INFORMATICAE

TOMUS 33. (2006)

HUNGARIA, EGER

ANNALES MATHEMATICAE ET INFORMATICAE

VOLUME 33. (2006)

INSTITUTE OF MATHEMATICS AND COMPUTER SCIENCE
ESZTERHÁZY KÁROLY COLLEGE
HUNGARY, EGER

HU ISSN 1787-5021 (Print)
HU ISSN 1787-6117 (Online)

A kiadásért felelős:
az Eszterházy Károly Főiskola rektora
Megjelent az EKF Líceum Kiadó gondozásában
Kiadóvezető: Kis-Tóth Lajos
Felelős szerkesztő: Zimányi Árpád
Műszaki szerkesztő: Tómács Tibor
Megjelent: 2007. január Példányszám: 50

Készítette: Diamond Digitális Nyomda, Eger
Ügyvezető: Hangácsi József

Annales Mathematicae et Informaticae
***33** (2006) pp. 3–13*
http://www.ektf.hu/tanszek/matematika/ami

Multiple root finder algorithm for Legendre and Chebyshev polynomials via Newton's method

Victor Barrera-Figueroa[a], Jorge Sosa-Pedroza[b], José López-Bonilla[c]

[abc]Instituto Politécnico Nacional, Escuela Superior de Ingeniería Mecánica y Eléctrica, Sección de Estudios de Postgrado e Investigación
[a]e-mail: vbarreraf@ipn.mx; [b]e-mail:jsosa@ipn.mx; [c]e-mail:jlopezb@ipn.mx

Submitted 23 June 2006; Accepted 31 October 2006

Abstract

We exhibit a numerical technique based on Newton's method for finding all the roots of Legendre and Chebyshev polynomials, which execute less iterations than the standard Newton's method and whose results can be compared with those for Chebyshev polynomials roots, for which exists a well known analytical formula. Our algorithm guarantees at least nine decimal correct ciphers in the worst case, however, when comparing with Chebyshev roots given by its formula, even eighteen decimal correct ciphers are achieved in several roots, in the best case. As a comparison guide the results are collated with those gotten by MATLAB.

Keywords: Newton's method, Legendre polynomials, Chebyshev polynomials, multiple root finder algorithm.

1. Introduction

Legendre polynomials (see [1, 2, 3, 4]) as well as Chebyshev (see [1, 2, 3, 4, 5]) ones has found countless applications in all branches of engineering and science, among which the most representatives include calculation of quadratures, electromagnetics and antenna applications, solutions for potential theory and for Schrödinger's equation, aerodynamics and mechanics applications, etc. This is due mainly because the use of their roots allows us to solve a specific problem with the best efficiency.

These polynomials are very similar to each other, not only in the form of their respective differential equation, but also in the numerical values of their roots.

However, sometimes it is better the use of Legendre polynomials instead of Chebyshev ones because by using their roots, one can reach the most optimized solution. For instance, when calculating a quadrature, Gauss established that the best way for obtaining the minimum error in a numerical integration is by dividing the integration domain in agreement with the way Legendre roots are distributed on their own domain.

So, because of the wide use of Legendre and Chebyshev roots, this paper presents a multiple root finder algorithm, which is expected to be a useful tool, not only in engineering work but also in numerical and mathematical analysis. This algorithm is based in the classical Newton's method (see [2]), however, we have made some modifications in order to find all the roots of an specific polynomial by using the improved Newton's method (see [2]).

2. Newton's method

Newton's process is a numerical tool used to find the zero x_e of a real-valued function $f(x)$, continuously differentiable and whose derivative does not vanish at $x = x_e$. The method consists in proposing an initial root x_0 which must be in the neighborhood of x_e, as shown in Figure 1. Once made this, we draw a tangent line to $f(x)$ at $x = x_0$, and determine its intersection with the x axis. The equation for the line is:

$$y = f(x_0) + f'(x_0)(x - x_0), \tag{2.1}$$

which means that locally, in the small neighborhood of x_0, $f(x)$ can be considered as a linear equation, even if $f(x)$ is not linear. We should notice that this equation corresponds to the truncate Taylor series with center at $x = x_0$ by neglecting the remainder term. The intersection, named x_1, is then:

$$x_1 - x_0 = -\frac{f(x_0)}{f'(x_0)}, \tag{2.2}$$

Now, x_1 is closer to the real zero x_e, and it will be used to draw the next tangent line to $f(x)$ at $x = x_1$ which produces the intersection x_2. So, in the n_{th} iteration, the intersection of the tangent line with the x axis will be [2]:

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}, \quad n = 1, 2, 3, \ldots \tag{2.3}$$

After a number of iterations x_n will be close enough to x_e. In this way, once we have established the allowed error ϵ for the zero, we can set the condition for stopping the method, which is reached when:

$$\mid x_n - x_{n-1} \mid \leqslant \epsilon. \tag{2.4}$$

Newton's technique will usually converge provided the initial guess is close enough to true root. Furthermore, for a zero of multiplicity 1, the convergence

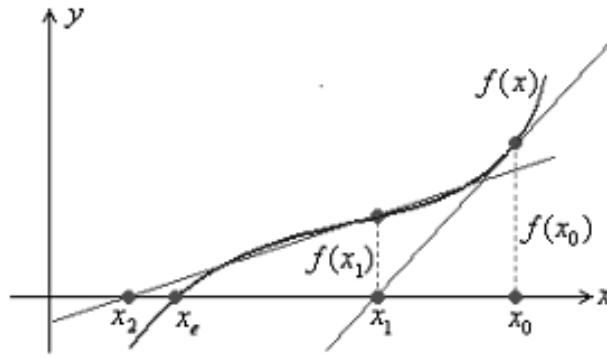

Figure 1: Tangent lines to $f(x)$ at $x = x_1$.

is at least quadratic in a neighborhood of x_e, which intuitively means that the number of correct digits roughly at least doubles in every step.

Newton's method acquire a bigger convergence if in Taylor series we consider the quadratic term:

$$y = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2, \tag{2.5}$$

that is, in the small neighborhood of x_0, $f(x)$ can be considered as a quadratic equation, even if $f(x)$ is not quadratic. The intersection x_1 between (2.5) and the x axis can be calculated from:

$$x_1 - x_0 = -\frac{f(x_0)}{f'(x_0) + \frac{1}{2}f''(x_0)(x_1 - x_0)}, \tag{2.6}$$

by substituting (2.2) into (2.6) we get:

$$x_1 - x_0 = -\frac{2f(x_0)f'(x_0)}{2\left[f'(x_0)\right]^2 - f(x_0)f''(x_0)}. \tag{2.7}$$

So, in the n_{th} iteration, we have the next recursive formula (see [2]):

$$x_n = x_{n-1} - \frac{2f(x_{n-1})f'(x_{n-1})}{2\left[f'(x_{n-1})\right]^2 - f(x_{n-1})f''(x_{n-1})}, \quad n = 1, 2, 3, \ldots \tag{2.8}$$

This is the improved Newton's method which converges faster than the original one. As well as we have to know the first derivative of $f(x)$ in both methods, in the improved one, we must know the second derivative of $f(x)$ in each point. This could be a problem if $f(x)$ is not an analytical formula but a set of data points, in which both derivatives could be calculated from standard numerical techniques.

3. Multiple root finder algorithm based on Newton's process

The application of Newton's method to a polynomial in the quest for all its roots could be easily implemented if we express it according to the fundamental theorem of Algebra:

$$f(x) = k(x - x_1)(x - x_2)\dots(x - x_N), \tag{3.1}$$

where N is the polynomial order, k is a proportionality constant and $\{x_1, x_2, \dots, x_N\}$ are all the polynomial roots, which are not necessarily put in order. In the following, the subindex j in the root x_j represents the number of root and it should not be confused with the number of iterations used in the previous sections which will be omitted from here. By applying Newton's algorithm to the polynomial (3.1) for searching the first root x_1, we have the next recursive relation:

$$x_1 = x_1 - \frac{f(x)}{f'(x)}. \tag{3.2}$$

Once x_1 is found, we built the polynomial $g(x)$ of order $N - 1$ which has not x_1 as one of its roots:

$$g(x) = \frac{f(x)}{x - x_1} = k(x - x_2)(x - x_3)\dots(x - x_N), \tag{3.3}$$

and by applying Newton's technique to $g(x)$ we determine the next root x_2 and so on. So, in the k_{th} step, we built the following polynomial:

$$g(x) = \frac{f(x)}{\prod_{i=1}^{k-1}(x - x_i)}, \tag{3.4}$$

and, according to Newton's method, the k_{th} root is:

$$x_k = x_k - \frac{g(x_k)}{g'(x_k)}, \tag{3.5}$$

which implies the calculation for the derivative of $g(x)$ which could not be so evident because of the need to find the derivative of the product term. Such derivative is:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\prod_{i=1}^{k-1}(x - x_i) &= \left[\frac{1}{x - x_1} + \frac{1}{x - x_2} + \dots + \frac{1}{x - x_{k-1}}\right]\prod_{i=1}^{k-1}(x - x_i) = \\ &= \prod_{i=1}^{k-1}(x - x_i)\sum_{i=1}^{k-1}\frac{1}{x - x_i}, \end{aligned} \tag{3.6}$$

therefore the derivative for $g(x)$ is:

$$g'(x) = \frac{1}{\prod_{i=1}^{k-1}(x - x_i)} \left[f'(x) - f(x) \sum_{i=1}^{k-1} \frac{1}{x - x_i} \right]. \tag{3.7}$$

By substituting (3.7) and (3.4) into (3.5) we reach the recursive relation for getting all the roots for the polynomial $f(x)$:

$$x_k = x_k - \frac{f(x_k)}{f'(x_k) - f(x_k) \sum_{i=1}^{k-1} \frac{1}{x - x_i}}, \quad k = 1, 2, \ldots, N. \tag{3.8}$$

On the basis of (3.8) we express the multiple root finder algorithm with the following diagram:

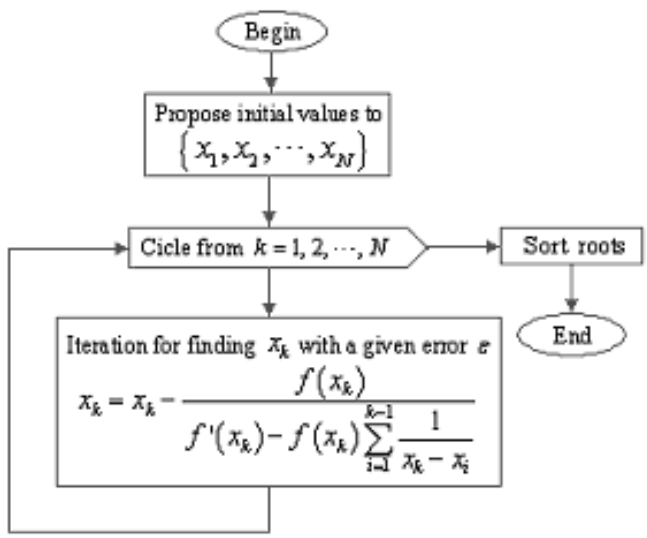

Figure 2: Multiple root finder algorithm.

In order to sort all the roots given by the multiple root finder process, we can introduce the well known bubble sort method which by successively interchanging the N roots can provide us the list of expected results. The bubble sort method is drawn in the following diagram:

4. Multiple root finder algorithm based on improved Newton's method

In order to improve the algorithm's convergence, we can introduce Newton's technique in the quest for all the polynomial roots. In first term we look at the

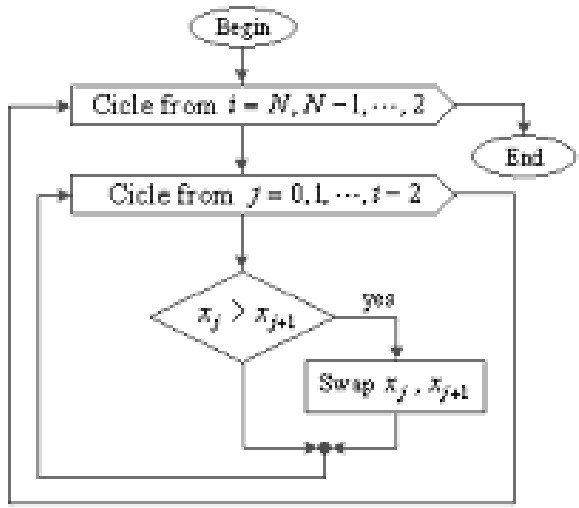

Figure 3: Bubble sort method for putting in order the polynomial's roots.

root x_1 of the polynomial (3.1) with the recursive relation (2.8). Once x_1 is found, we built the polynomial $g(x)$, (3.3). By applying improved Newton's method to $g(x)$ we found the next root x_2 and so on. So, in the k_{th} step, we built the polynomial (3.4), from which the k_{th} root is:

$$x_k = x_k - \frac{2g(x_k)g'(x_k)}{2\left[g'(x_k)\right]^2 - g(x_k)g''(x_k)}. \tag{4.1}$$

Iterative formula (4.1) implies the calculation for the second derivative of $g(x)$ which could a hard task, however, the result is not as difficult as one could expect:

$$g''(x) = \frac{f''(x) - 2f'(x)\sum\limits_{i=1}^{k-1}\frac{1}{x-x_i} + f(x)\left[\sum\limits_{i=1}^{k-1}\frac{1}{(x-x_i)^2} + \left(\sum\limits_{i=1}^{k-1}\frac{1}{x-x_i}\right)^2\right]}{\prod\limits_{i=1}^{k-1}(x-x_i)}. \tag{4.2}$$

So by substituting (3.4),(3.7) and (4.2) into (4.1) we get:

$$x_k = x_k - \frac{2f(x_k)B(x_k)}{B^2(x_k) + \left[f'(x_k)\right]^2 - f(x_k)\left[f''(x_k) + f(x_k)\sum\limits_{i=1}^{k-1}\frac{1}{(x_k-x_i)^2}\right]}, \tag{4.3}$$

$$B(x_k) = f'(x_k) - f(x_k) \sum_{i=1}^{k-1} \frac{1}{x_k - x_i}, \quad k = 1, 2, \ldots, N.$$

Notice the similarity between (4.1) and (4.3) where $f(x_k)$ is similar to $g(x_k)$, and $g'(x_k)$ is analogous to $B(x_k)$. Obviously both formulas are not completely analogous, but the existent similarity is notorious. On the basis of (4.3) we express the improved multiple root finder algorithm with the following diagram, Figure 4, again, the bubble sort method could be used in order to sort the obtained roots:

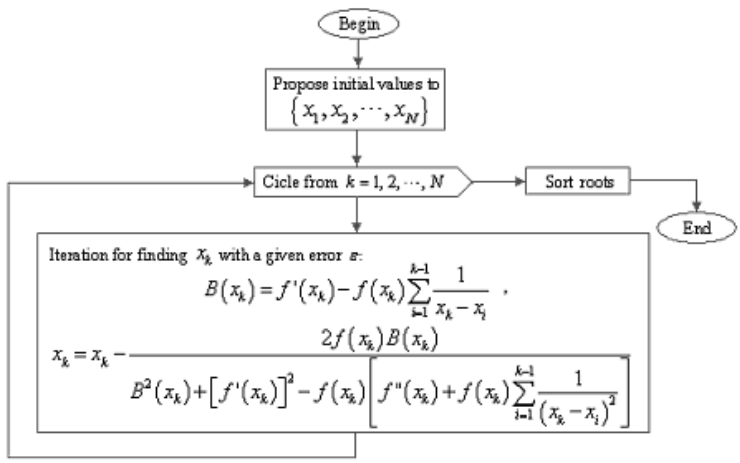

Figure 4: Improved multiple root finder algorithm.

5. Legendre and Chebyshev polynomials roots calculation

In this section we present the results reached by both algorithms applied to Legendre and Chebyshev polynomials; also, it is shown the number of performed iteration. Such results are compared with those gotten by MATLAB. In the case of Chebyshev's roots, we provide the analytical results gotten by the formula [2]:

$$x_k = \cos\left(\frac{2k-1}{2N}\pi\right), \quad k = 1, 2, \ldots, N, \tag{5.1}$$

where N is the polynomial degree. Legendre polynomials $P_n(x)$ are the solution of the Legendre differential equation:

$$(1 - x^2)\frac{\mathrm{d}^2 P_n(x)}{\mathrm{d}x^2} - 2x\frac{\mathrm{d}P_n(x)}{\mathrm{d}x} + n(n+1)P_n(x) = 0, \quad n = 0, 1, 2, \ldots, \tag{5.2}$$

while Chebyshev polynomials $T_n(x)$ are the solution for its respective differential equation:

$$(1-x^2)\frac{\mathrm{d}^2T_n(x)}{\mathrm{d}x^2} - x\frac{\mathrm{d}T_n(x)}{\mathrm{d}x} + n^2T_n(x) = 0, \quad n = 0, 1, 2, \dots \ . \tag{5.3}$$

Both $P_n(x)$ and $T_n(x)$ can be defined by mean of recursive relations, which is an important issue in the numerical point of view. For Legendre polynomials, the recursive relations are:

$$P_0(x) = 1,\ P_1(x) = x,\ P_{n+2} = \frac{2n+3}{n+2}xP_{n+1} - \frac{n+1}{n+2}P_n,\ n = 0, 1, 2, \dots \ , \tag{5.4}$$

while for Chebyshev polynomials the recursive relations are:

$$T_0(x) = 1,\ T_1(x) = x,\ T_{n+2} = 2xT_{n+1} - T_n, \quad n = 0, 1, 2, \dots \ , \tag{5.5}$$

Such polynomials are plotted in Figure 5. It is notable how the roots are clustered in the ends of the domain $[-1, 1]$ in both polynomials.

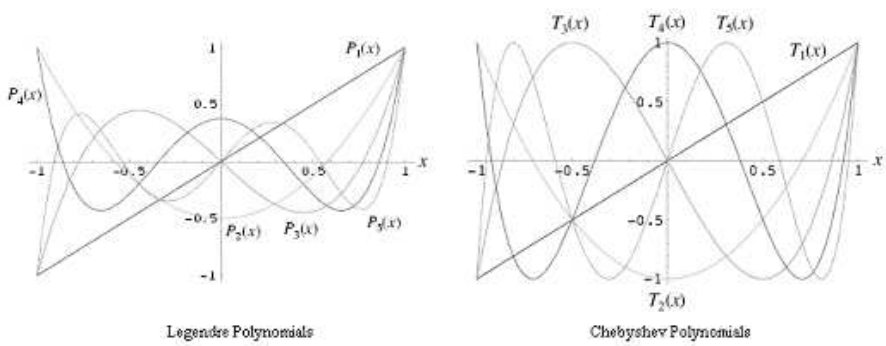

Figure 5: Legendre and Chebyshev polynomials.

In Table 1 are shown the results for $P_{19}(x)$ roots reached by both algorithms, with a permissible error of $\epsilon = 1 \times 10^{-12}$ for each of them.

In Table 2 the roots for $P_{19}(x)$ obtained using MATLAB are presented for comparison. In Table 3 are shown the results for $T_{19}(x)$ roots reached by both algorithms, with a permissible error of $\epsilon = 1 \times 10^{-12}$ for each of them.

In Table 4 the roots for $T_{19}(x)$ obtained using (5.1) and MATLAB are presented for comparison.

6. Conclusions

Both algorithms have shown to be effective in finding all the roots for a specific polynomial. Such polynomial must have all its roots real and with simple multiplicity. These conditions are satisfied by Legendre and Chebyshev polynomials.

k	Multiple root finder algorithm, x_k	Number of iterations	Improved multiple root finder algorithm, x_k	Number of iterations
1	-0.992406843843584352	2	-0.992406843843584350	3
2	-0.960208152134830018	12	-0.960208152134830020	4
3	-0.903155903614817901	15	-0.903155903614817900	7
4	-0.822714656537142819	6	-0.822714656537142820	4
5	0.720966177335229386	6	-0.720966177335229390	4
6	-0.600545304661680990	6	-0.600545304661680990	5
7	-0.464570741375960938	7	-0.464570741375960940	5
8	-0.316564099963629830	8	-0.316564099963629830	6
9	-0.160358645640225367	9	-0.160358645640225390	6
10	0.000000000000000000	10	0.000000000000000000	6
11	0.160358645640225367	10	0.160358645640225390	7
12	0.316564099963629830	11	0.316564099963629830	7
13	0.464570741375960938	11	0.464570741375960940	7
14	0.600545304661680990	12	0.600545304661680990	8
15	0.720966177335229386	11	0.720966177335229390	8
16	0.822714656537142819	11	0.822714656537142820	8
17	0.903155903614817901	10	0.903155903614817900	7
18	0.960208152134830018	8	0.960208152134830020	7
19	0.992406843843584352	8	0.992406843843584460	5

Table 1: Roots for $P_{19}(x)$ with both algorithms.

However, for other type of polynomials, another process is being developed in order to get not only their real roots but also their complex ones, taking into account their own multiplicity.

The first algorithm is faster than the second one, because it performs fewer operations than the improved one, which must calculate the second derivative and several sums, among other operations. However, the second algorithm executes less iteration than the first one, and this fact could be useful in finding the roots for polynomials of big order. Both algorithms provide us several correct ciphers when comparing with the results gotten by MATLAB, however, both are faster than the routines used by MATLAB in doing the same action.

It is expected that these algorithms can be employed as a part for several programs which must involve quadratures or interpolations (among other numerical issues) while looking for engineering or science solutions, because of its speed and ease for programming. We the authors recommend an error of $\epsilon = 1 \times 10^{-12}$ for each root, which has shown to provide correct results for all the possible polynomial degrees. However, for an error of $\epsilon = 1 \times 10^{-18}$ in some degrees, both algorithms perform a great number of iterations in which the loop becomes endless.

k	MATLAB, x_k
1	-0.992406843843584350
2	-0.960208152134830020
3	-0.903155903614817900
4	-0.822714656537142820
5	-0.720966177335229390
6	-0.600545304661680990
7	-0.464570741375960940
8	-0.316564099963629830
9	-0.160358645640225370
10	0.000000000000000000
11	0.160358645640225370
12	0.316564099963629830
13	0.464570741375960940
14	0.600545304661680990
15	0.720966177335229390
16	0.822714656537142820
17	0.903155903614817900
18	0.960208152134830020
19	0.992406843843584350

Table 2: Roots for $P_{19}(x)$ with MATLAB.

k	Multiple root finder algorithm, x_k	Number of iterations	Improved multiple root finder algorithm, x_k	Number of iterations
1	-0.996584493006669847	2	-0.996584493006669850	3
2	-0.969400265939330374	12	-0.969400265939330370	5
3	-0.915773326655057396	15	-0.915773326655057400	7
4	-0.837166478262528546	4	-0.837166478262528550	4
5	-0.735723910673131587	5	-0.735723910673131590	4
6	-0.614212712689667817	6	-0.614212712689667820	4
7	-0.475947393037073563	7	-0.475947393037073560	5
8	-0.324699469204683511	8	-0.324699469204683510	5
9	-0.164594590280733893	9	-0.164594590280733890	6
10	0.000000000000000000	10	0.000000000000000000	6
11	0.164594590280733893	10	0.164594590280733890	7
12	0.324699469204683511	11	0.324699469204683460	7
13	0.475947393037073563	11	0.475947393037073560	7
14	0.614212712689667817	12	0.614212712689667820	8
15	0.735723910673131587	12	0.735723910673131590	8
16	0.837166478262528546	11	0.837166478262528550	8
17	0.915773326655057396	10	0.915773326655057400	7
18	0.969400265939330374	8	0.969400265939330370	7
19	0.996584493006669847	8	0.996584493006669850	5

Table 3: Roots for $T_{19}(x)$ with both algorithms.

k	Analytical formula: $x_k = \cos\left(\frac{2k-1}{2N}\pi\right)$	MATLAB, x_k
1	-0.996584493006669847	-0.996584493006669850
2	0.969400265939330486	-0.969400265939330370
3	-0.915773326655057507	-0.915773326655057510
4	-0.837166478262528546	-0.837166478262528550
5	-0.735723910673131587	-0.735723910673131590
6	-0.614212712689667817	-0.614212712689667820
7	-0.475947393037073563	-0.475947393037073560
8	-0.324699469204683455	-0.324699469204683460
9	-0.164594590280733838	-0.164594590280734060
10	0.000000000000000061	0.000000000000000061
11	0.164594590280733977	0.164594590280733980
12	0.324699469204683566	0.324699469204683570
13	0.475947393037073618	0.475947393037073620
14	0.614212712689667817	0.614212712689667820
15	0.735723910673131698	0.735723910673131590
16	0.837166478262528657	0.837166478262528550
17	0.915773326655057396	0.915773326655057400
18	0.969400265939330374	0.969400265939330370
19	0.996584493006669847	0.996584493006669850

Table 4: Roots for $T_{19}(x)$ with analytical formula and MATLAB.

References

[1] M. Abramowitz and I. A. Stegun, Handbook of mathematical functions, Wiley and Sons, N. Y. (1972).

[2] C. Lanczos, Applied analysis, Dover N. Y. (1988).

[3] J. B. Seaborn, Hypergeometric functions and their applications, Springer-Verlag (1991).

[4] C. Lanczos, Linear differential operators, Dover N. Y. (1997).

[5] J. C. Mason and D. C. Handscomb, Chebyshev polynomials, Chapman & Hall - CRC Press (2002).

V. Barrera-Figueroa, J. Sosa-Pedroza, J. López-Bonilla
UPALM. Edif. Z-4, 3er. piso, col. Lindavista, C.P. 07738, México D.F.

Annales Mathematicae et Informaticae
***33** (2006) pp. 15–21*
http://www.ektf.hu/tanszek/matematika/ami

On Heron triangles

Andrew Bremner

Department of Mathematics and Statistics, Arizona State University
e-mail: bremner@asu.edu

Submitted 30 August 2006; Accepted 15 November 2006

Abstract

There has previously been given a one-parameter family of pairs of Heron triangles with equal perimeter and area. In this note, we find two two-parameter families of such triangle pairs, one of which contains the known one-parameter family as a special case. Second, for an arbitrary integer $n \geqslant 2$ we show how to find a set of n Heron triangles in two parameters such that all triangles have equal perimeter and area.

MSC: 11D72, 14G05.

1. Introduction

A.-V. Kramer and F. Luca [3] investigate several problems related to Heron triangles (triangles with integral sides and integral area; *rational* triangles are those with rational sides and rational area, which by scaling thus become Heron triangles). They give a one-parameter family of pairs of such triangles having equal perimeter and area (curiously, Aassila [1] in a paper that gives the appearance of plagiarism produces exactly the same parametric family). This note shows first in completely elementary manner how to construct a doubly infinite family of such triangle pairs. In fact we produce two such parametrizations, in three (homogeneous) parameters, containing the Kramer-Luca family as a special case.

Recently, van Luijk [4] answers a question posed by Kramer and Luca by showing that there exist arbitarily many Heron triangles having equal perimeter and area, and gives a method whereby a one-parameter family may be written down for n such triangles for a given integer n. We use the same ideas in showing how to produce a set of n Heron triangles in two parameters with the property of equal perimeter and area.

2. Pairs of Heron triangles

Brahmagupta gave a parametrization for all Heron triangles, with sides proportional to

$$(v+w)(u^2-vw), \quad v(u^2+w^2), \quad w(u^2+v^2),$$

where the semi-perimeter is equal to $u^2(v+w)$, and the area is equal to $uvw(v+w)(u^2-vw)$. Thus to find a pair of Heron triangles with equal perimeter and area, we take the two triangles with independent parameters u, v, w and r, s, t and demand solutions of the system

$$u^2(v+w) = mr^2(s+t), \qquad uvw(v+w)(u^2-vw) = m^2rst(s+t)(r^2-st),$$

for a scaling factor m. The general solution will be difficult to obtain. However, we focus on the situation $u = r$ and consider two cases.
First, $m = 1$. Then $w = s + t - v$ and equality of the area demands

$$(s+t)u(s-v)(t-v)(st-u^2+sv+tv-v^2) = 0.$$

For non-trivial solutions, we thus have $st - u^2 + sv + tv - v^2 = 0$, and this quadric surface is birationally equivalent to the projective plane under the mapping

$$s : t : u : v = b(b+c) : (a^2 - bc + c^2) : a(b+c) : c(b+c),$$

(with inverse $a : b : c = u : s : v$). Accordingly,

$$r : s : t : u : v : w = a(b+c) : b(b+c) : a^2 - bc + c^2 : a(b+c) : c(b+c) : a^2 + b^2 - bc,$$

leading to the triangle-pair

$$\begin{aligned} &b(a^2-bc+c^2)(a^2+b^2+c^2),\\ &c(a^4+3a^2b^2+b^4-2b^3c+a^2c^2+b^2c^2),\\ &(b+c)(a^2+b^2-bc)(a^2+c^2), \end{aligned} \tag{2.1}$$

and

$$\begin{aligned} &c(a^2+b^2-bc)(a^2+b^2+c^2),\\ &b(a^4+a^2b^2+3a^2c^2+b^2c^2-2bc^3+c^4),\\ &(a^2+b^2)(b+c)(a^2-bc+c^2) \end{aligned} \tag{2.2}$$

with the common semi-perimeter $a^2(b+c)(a^2+b^2+c^2)$ and the common area $abc(b+c)(a^2+b^2-bc)(a^2-bc+c^2)(a^2+b^2+c^2)$. As an example, at $(a, b, c) = (2, 3, 4)$, we obtain the triangles $(174, 197, 35)$ and $(29, 195, 182)$, both with perimeter 406 and area 2436.

Second, we assume $m \neq 1$ and restrict to $v = ms$, $w = mt$, when the perimeters become equal. Equality of the area demands

$$-r^2 + st + mst + m^2st = 0,$$

and considered as a quadric curve over $\mathbf{Q}(m)$ we have a birational correspondence with the projective line given by

$$r : s : t = (1 + m + m^2)\pi\rho : (1 + m + m^2)\rho^2 : \pi^2, \qquad \pi : \rho = r : s.$$

Thus

$$\begin{aligned}
&r : s : t : u : v : w = \\
&(1 + m + m^2)\pi\rho : (1 + m + m^2)\rho^2 : \pi^2 : (1 + m + m^2)\pi\rho : m(1 + m + m^2)\rho^2 : m\pi^2 = \\
&P(Q^2 + QR + R^2) : R(Q^2 + QR + R^2) : P^2R : P(Q^2 + QR + R^2) : Q(Q^2 + QR + R^2) : P^2Q,
\end{aligned}$$

on setting $\pi/\rho = P/R$, $m = Q/R$ (so P, Q, R independent parameters). This leads to the triangle pair

$$\begin{aligned}
&Q(Q + R)(P^2 + Q^2 + QR + R^2), \\
&Q^4 + 2Q^3R + P^2R^2 + 3Q^2R^2 + 2QR^3 + R^4, \\
&(P^2 + R^2)(Q^2 + QR + R^2)
\end{aligned} \tag{2.3}$$

and

$$\begin{aligned}
&R(Q + R)(P^2 + Q^2 + QR + R^2), \\
&P^2Q^2 + Q^4 + 2Q^3R + 3Q^2R^2 + 2QR^3 + R^4, \\
&(P^2 + Q^2)(Q^2 + QR + R^2).
\end{aligned} \tag{2.4}$$

If we put $(P, Q, R) = (t(3 + 3t^2 + t^4), 1, 1 + t^2)$, then the resulting triangle pair is the one-parameter family of Kramer and Luca [3].

3. Sets of Heron triangles

Kramer and Luca [3] essentially ask whether one can find sets of k Heron triangles with equal perimeter and area, for a given positive integer k. Relatedly, for a given triangle with rational sides a_0, b_0, c_0 of perimeter $2s$ and area A, we can ask to find other triangles with the same perimeter and area. If such a triangle has sides a, b, c then

$$a + b + c = 2s, \qquad s(s - a)(s - b)(s - c) = A^2.$$

Equivalently,

$$C : s(s - a)(s - b)(a + b - s) = A^2, \tag{3.1}$$

the equation of a cubic curve in the a, b-plane. Certainly C contains the points at infinity $(0, 1, 0), (1, 0, 0), (-1, 1, 0)$, so is an elliptic curve. Fixing one of these points

as the zero of the group law, then the other two points become torsion points of order 3. Moreover, C contains the rational points at $(a,b) = (a_0,b_0)$, (b_0,a_0), (b_0,c_0), (c_0,b_0), (c_0,a_0), (a_0,c_0), the sextet comprising the points $\pm(a_0,b_0)$ + 3-torsion in the group $C(\mathbf{Q})$. In general, the point (a_0,b_0) will be of infinite order, allowing arbitrarily large sets of rational points (a,b) to be determined, each in turn defining a triangle with sides $(a,b,2s-a-b)$, having the perimeter $2s$ and area A. The triangle may of course not be geometrically realisable if $a<0$, $b<0$, or $2s<a+b$, or if the triangle inequality is violated; but since (a_0,b_0) corresponds to a genuine triangle, a density argument of points on the elliptic curve (dating back to Hurwitz: see Theorem 13 of [2]) guarantees the existence of arbitrarily many (a,b) corresponding to genuine triangles. (van Luijk [4] makes this argument explicit: if points P_i correspond to real triangles, then $\sum_{i=1}^{i=k} n_i P_i$ corresponds to a real triangle if and only if $\sum_{i=1}^{i=k} n_i$ is odd). Scaling will now produce arbitrarily large sets of Heron triangles with equal perimeter and area.

Remark 3.1. The isosceles triangle (a_0,a_0,c_0) with $b_0 = a_0$ has corresponding curve C with (homogeneous) equation

$$2ab(a+b)-(2a_0+c_0)(a^2+3ab+b^2)d+(2a_0+c_0)^2(a+b)d^2-a_0(2a_0^2+3a_0c_0+2c_0^2)d^3=0,$$

and the points (a_0,a_0), (a_0,c_0), and (c_0,a_0) have the property that doubling them results in a torsion point at infinity: so the points are either of order 2 or of order 6. The point (a_0,b_0) may also be of finite order for a non-isosceles triangle, for example the triangle $(a_0,b_0,c_0) = (13,27,34)$, where (a_0,b_0) has order 12. If the rational rank of C is 0 (as is the case for example with the (non-Heron) triangles given by $(a_0,b_0,c_0) = (1,1,1)$ or $(13,27,34)$) then there are at most finitely many rational-sided triangles with same perimeter and area, arising from the torsion points on C. When (a_0,b_0) is a torsion point therefore, to determine arbitrarily many triangles with equal perimeter and area we require C to have an additional rational non-torsion point (corresponding to a real triangle) in order to start the above construction. For instance, the Heron triangle $(14,25,25)$ has (a_0,b_0) a torsion point, but the respective curve C exhibits the additional non-torsion point $\left(\frac{39}{2},\frac{136}{5}\right)$, leading to the triangle $\left(\frac{39}{2},\frac{136}{5},\frac{173}{10}\right)$, with same perimeter and area.

As illustration of the above construction of sets of points, take as example the Heron triangle $(3,4,5)$, with semi-perimeter 6 and area 6. The construction of taking multiples of the point $(3,4)$ on C provides the triangles

$$\left(\frac{156}{35},\frac{41}{15},\frac{101}{21}\right),\quad \left(\frac{81831}{16159},\frac{27689}{8023},\frac{35380}{10153}\right),\quad \left(\frac{678541575}{151345267},\frac{683550052}{142637329},\frac{221167193}{81180907}\right),\quad \ldots$$

with perimeter 6 and area 6. The numbers grow rapidly because the underlying elliptic curve here has rank 1, and the heights on an elliptic curve of multiples of a fixed point are rapidly increasing. When the underlying curve has higher rank, then by taking linear combinations of the generators there is expectation of a greater supply of rational points with relatively small height, and accordingly an

expectation of a more plentiful supply of triangles, as for example in the table of van Luijk [4], where 20 triangles are generated with same area and perimeter as the triangle $(75, 146, 169)$; in this instance, the underlying elliptic curve has rank 4 (independent points on C are $(111, 104)$, $(125, 91)$, $(146, 75)$, and $(265, 203)$).

Of course, we can use as our initial triangle one given by a one- or two-parameter family, and construct arbitrarily many triangles in the corresponding number of parameters, all having the same perimeter and area. The formulae rapidly become lengthy, and we give as example a three (homogeneous) parameter family of only four such triangles, arising from the parametrizations at (2.3), (2.4). Denote the points on C corresponding to the parametrizations (2.3), (2.4), by S and T respectively. Then the parametrizations corresponding to the points S, T, $2S+T$, $S+2T$ are given by:

$$
\begin{aligned}
&Q(Q+R)(2Q+R)(Q+2R)(P^2+Q^2+QR+R^2)(P^2Q+Q^3+P^2R+Q^2R+QR^2)(P^2Q+P^2R+\\
&\quad Q^2R+QR^2+R^3)(P^2Q+Q^3+2Q^2R+2QR^2+R^3)(Q^3+P^2R+2Q^2R+2QR^2+R^3),\\
&(2Q+R)(Q+2R)(P^2Q+Q^3+P^2R+Q^2R+QR^2)(P^2Q+P^2R+Q^2R+QR^2+R^3)(P^2Q+Q^3+\\
&\quad 2Q^2R+2QR^2+R^3)(Q^3+P^2R+2Q^2R+2QR^2+R^3)(Q^4+2Q^3R+P^2R^2+3Q^2R^2+2QR^3+R^4),\\
&(2Q+R)(Q+2R)(P^2+R^2)(Q^2+QR+R^2)(P^2Q+Q^3+P^2R+Q^2R+QR^2)(P^2Q+P^2R+\\
&\quad Q^2R+QR^2+R^3)(P^2Q+Q^3+2Q^2R+2QR^2+R^3)(Q^3+P^2R+2Q^2R+2QR^2+R^3),
\end{aligned}
$$

$$
\begin{aligned}
&R(Q+R)(2Q+R)(Q+2R)(P^2+Q^2+QR+R^2)(P^2Q+Q^3+P^2R+Q^2R+QR^2)(P^2Q+P^2R+\\
&\quad Q^2R+QR^2+R^3)(P^2Q+Q^3+2Q^2R+2QR^2+R^3)(Q^3+P^2R+2Q^2R+2QR^2+R^3),\\
&(2Q+R)(Q+2R)(P^2Q+Q^3+P^2R+Q^2R+QR^2)(P^2Q+P^2R+Q^2R+QR^2+R^3)(P^2Q+Q^3+\\
&\quad 2Q^2R+2QR^2+R^3)(Q^3+P^2R+2Q^2R+2QR^2+R^3)(P^2Q^2+Q^4+2Q^3R+3Q^2R^2+2QR^3+R^4),\\
&(P^2+Q^2)(2Q+R)(Q+2R)(Q^2+QR+R^2)(P^2Q+Q^3+P^2R+Q^2R+QR^2)(P^2Q+P^2R+\\
&\quad Q^2R+QR^2+R^3)(P^2Q+Q^3+2Q^2R+2QR^2+R^3)(Q^3+P^2R+2Q^2R+2QR^2+R^3),
\end{aligned}
$$

$$
\begin{aligned}
&(2Q+R)(P^2Q+Q^3+P^2R+Q^2R+QR^2)(P^2Q+Q^3+2Q^2R+2QR^2+R^3)(Q^3+P^2R+2Q^2R+\\
&\quad 2QR^2+R^3)(P^4Q^4+P^2Q^6+4P^4Q^3R+6P^2Q^5R+6P^4Q^2R^2+13P^2Q^4R^2+4P^4QR^3+\\
&\quad 16P^2Q^3R^3+P^4R^4+13P^2Q^2R^4+Q^4R^4+6P^2QR^5+2Q^3R^5+2P^2R^6+3Q^2R^6+2QR^7+R^8),\\
&(2Q+R)(P^2Q+Q^3+P^2R+Q^2R+QR^2)(P^2Q+P^2R+Q^2R+QR^2+R^3)(P^2Q+Q^3+2Q^2R+\\
&\quad 2QR^2+R^3)(P^2Q^6+Q^8+6P^2Q^5R+6Q^7R+13P^2Q^4R^2+17Q^6R^2+16P^2Q^3R^3+30Q^5R^3+\\
&\quad P^4R^4+13P^2Q^2R^4+36Q^4R^4+6P^2QR^5+30Q^3R^5+2P^2R^6+17Q^2R^6+6QR^7+R^8),\\
&R(Q+R)(2Q+R)(Q+2R)(Q^2+QR+R^2)(P^2+Q^2+QR+R^2)(P^2Q+Q^3+P^2R+Q^2R+QR^2)\\
&\quad (P^2Q+Q^3+2Q^2R+2QR^2+R^3)(P^4-P^2Q^2+Q^4+2P^2QR+2Q^3R+2P^2R^2+3Q^2R^2+2QR^3+R^4),
\end{aligned}
$$

$$
\begin{aligned}
&(Q+2R)(P^2Q+P^2R+Q^2R+QR^2+R^3)(P^2Q+Q^3+2Q^2R+2QR^2+R^3)(Q^3+P^2R+2Q^2R+\\
&\quad 2QR^2+R^3)(P^4Q^4+2P^2Q^6+Q^8+4P^4Q^3R+6P^2Q^5R+2Q^7R+6P^4Q^2R^2+13P^2Q^4R^2+3Q^6R^2+\\
&\quad 4P^4QR^3+16P^2Q^3R^3+2Q^5R^3+P^4R^4+13P^2Q^2R^4+Q^4R^4+6P^2QR^5+P^2R^6),\\
&(Q+2R)(P^2Q+Q^3+P^2R+Q^2R+QR^2)(P^2Q+P^2R+Q^2R+QR^2+R^3)(Q^3+P^2R+2Q^2R+\\
&\quad 2QR^2+R^3)(P^4Q^4+2P^2Q^6+Q^8+6P^2Q^5R+6Q^7R+13P^2Q^4R^2+17Q^6R^2+16P^2Q^3R^3+\\
&\quad 30Q^5R^3+13P^2Q^2R^4+36Q^4R^4+6P^2QR^5+30Q^3R^5+P^2R^6+17Q^2R^6+6QR^7+R^8),\\
&Q(Q+R)(2Q+R)(Q+2R)(Q^2+QR+R^2)(P^2+Q^2+QR+R^2)(P^2Q+P^2R+Q^2R+QR^2+R^3)\\
&\quad (Q^3+P^2R+2Q^2R+2QR^2+R^3)(P^4+2P^2Q^2+Q^4+2P^2QR+2Q^3R-P^2R^2+3Q^2R^2+2QR^3+R^4).
\end{aligned}
$$

Remark 3.2. The family of elliptic curves at (3.1) is actually one-dimensional parameterized by $t = A/s^2$, namely

$$(1-x)(1-y)(x+y-1) = t^2, \tag{3.2}$$

where $(x, y) = (a/s, b/s)$, $t = A/s^2$. For the triangles at (2.1), (2.2), we have

$$t = \frac{bc(a^2+b^2-bc)(a^2+c^2-bc)}{a^3(b+c)(a^2+b^2+c^2)}, \tag{3.3}$$

which for general t defines a curve in the (a, b, c)-plane of genus 5. Thus by Falting's proof of the Mordell Conjecture, only finitely many a, b, c give rise to the same t. Specialization of a, b, c therefore in general produces n-tuples of triangles each corresponding to a different elliptic curve. A similar remark holds for the triangles at (2.3), (2.4), where

$$t = \frac{PQR(Q+R)}{(Q^2+QR+R^2)(P^2+Q^2+QR+R^2)}$$

defines for general t a curve of genus 2 in the (P, Q, R)-plane.

Remark 3.3. The curve (3.2) comprises a bounded component lying within the region $0 < x < 1,\ 0 < y < 1,\ x + y > 1$, and an unbounded component in the region $x > 1$. Real triangles correspond to points on the bounded component, and it is immediately apparent from the geometrical definition of addition on the curve (and straightforward to prove) that if points P_i lie on the bounded component, then $\sum_{i=1}^{i=k} n_i P_i$ lies on the bounded component if and only $\sum_{i=1}^{i=k} n_i$ is odd, recovering the density argument mentioned above.

Remark 3.4. The triangles at (2.1), (2.2) give rise to points S' and T' on the elliptic curve (3.2), with t given by (3.3); and by specialization, S', T' are seen to be generically linearly independent in the Mordell-Weil group. Similarly the two points S and T arising from triangles (2.3), (2.4) are independent in the corresponding Mordell-Weil group. It may well be possible to specialize to polynomials in one variable so that the Mordell-Weil group acquires further independent points, so will have rank at least 3. As remarked previously, the larger the rank, the greater the expectation of a supply of points of small height, and hence the expectation of providing parametrizations of smaller degree.

We refine the question of Kramer and Luca by asking how many distinct *primitive* Heron triangles may be found (those with sides having no non-trivial common divisor), with equal perimeter and area. It is straightforward to find pairs with this property, and there is the triple

$$(75, 146, 169), \qquad (91, 125, 174), \qquad (104, 111, 175)$$

(implicit in the table of van Luijk) with perimeter 390 and area 5460; but I am not aware of a quadruple of such triangles.

Acknowledgements. My thanks to the referee for helpful criticism of the first draft of this paper.

References

[1] M. Aassila, Some results on Heron triangles, *Elem. Math.*, 56 (2001), 143-146.

[2] A. Hurwitz, Über ternäre diophantische Gleichungen dritten Grades, *Viertel Jahrschrift Naturforsch. Gesellsch.* Zürich 62 (1917) 207-229.

[3] A.-V. Kramer and F. Luca, Some results on Heron triangles, *Acta Acad. Paed. Agriensis, Sectio Math.* 26 (2000), 1-10.

[4] R. van Luijk, An elliptic K3 surface associated to Heron triangles, (see `http://arxiv.org/abs/math/0411606`) *J. Number Theory*, to appear.

Andrew Bremner
Department of Mathematics and Statistics
Arizona State University
Tempe
AZ 85287-1804
USA

Annales Mathematicae et Informaticae
***33** (2006) pp. 23–34*
http://www.ektf.hu/tanszek/matematika/ami

Cyclic quadrangles from squares

Zvonko Čerin[a], Gian Mario Gianella[b]

[a]Department of Mathematics, University of Zagreb
e-mail: cerin@math.hr

[b]Dipartimento di Matematica, Universita di Torino
e-mail: gianella@dm.unito.it

Submitted 6 April 2006; Accepted 7 April 2006

Abstract

In this paper we show how to use computers to discover appearence of cyclic quadrangles in geometric configurations based on a fixed square $ABCD$ and a variable point P in the plane. The idea is to consider various central points (like the orthocenters) of the four triangles ABP, BCP, CDP and DAP or their orthogonal projections to the lines AP, BP, CP and DP. This is done in Maple V by describing basic functions for the analytic plane geometry and applying them to these configurations. The figures are realized in The Geometer's Sketchpad, Mathematica, and Maple V.

Keywords: square, triangle, orthocenter, circumcenter, area, Steiner point, cyclic quadrangle

MSC: 51N20, 51M04, 14A25,14Q05

1. Introduction

Consider in the plane a positively oriented (in the counterclockwise sense) square $ABCD$ with the center O and a point P which is not on any of the four lines AB, BC, CD and DA. Let H_a, H_b, H_c and H_d be the orthocenters (i.e., the intersections of altitudes) of the triangles ABP, BCP, CDP and DAP, respectively. Let J_a, J_b, J_c and J_d denote the orthogonal projections of H_a, H_b, H_c and H_d onto the lines AP, BP, CP and DP, respectively. (See Figures 1 and 2.)

In this paper we want to show how one can use computers to explore properties of the quadrangles $H_aH_bH_cH_d$ and $J_aJ_bJ_cJ_d$. The first property of the quadrangle $H_aH_bH_cH_d$ that its diagonals H_aH_c and H_bH_d are perpendicular is obvious because points H_a and H_c are on the perpendicular through P onto lines AB and CD while H_b and H_d are on the perpendicular through P onto lines BC and DA.

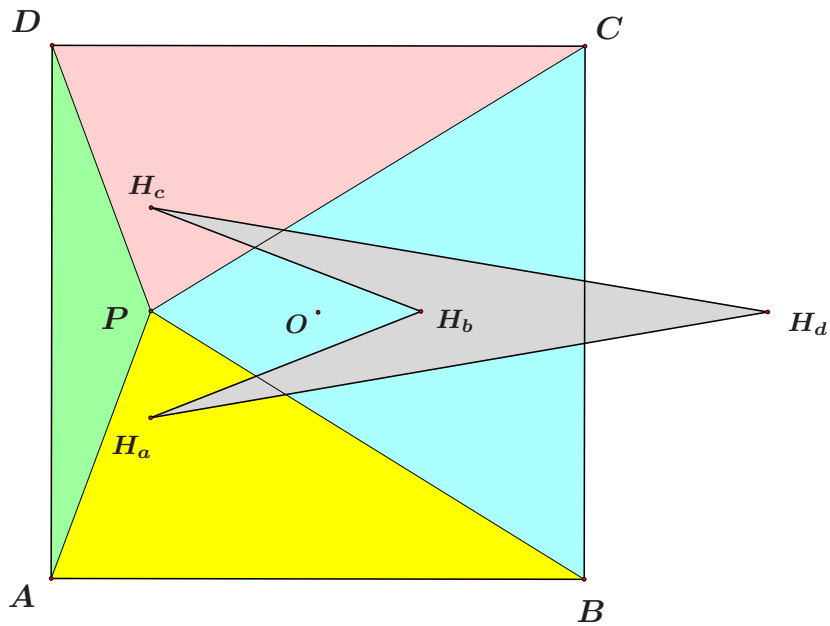

Figure 1: The quadrangle $H_aH_bH_cH_d$ from orthocenters.

The second property is more difficult to establish. We shall do this later using analytic geometry. Ever since Decartes this is the most simple and the most effective method to transfer geometric problems into algebraic setting. The reduction usually leads to equations whose solutions give answers. Since the software for symbolic computation (like Derive, Maple V and Mathematica) excels in solving equations, in this way we get the possibility to use computers in our explorations.

Property 2. *The points O, P, J_a, J_b, J_c and J_d lie on a circle. In particular, the quadrangle $J_aJ_bJ_cJ_d$ is cyclic.*

We can say more about the circle m which appears in the Property 2. It is the circumcircle of the negatively oriented square $PONM$ built on the segment $\overline{PO}$. Hence, if $|PO| = \delta$, then the radius of the circle m is $\frac{\delta\sqrt{2}}{2}$ (see Figure 2).

The third property describes the following surprising connection of the quadrangles $H_aH_bH_cH_d$ and $J_aJ_bJ_cJ_d$ (see Figure 3).

Property 3. *The lines H_aJ_a, H_bJ_b, H_cJ_c and H_dJ_d intersect in the point N and go through the points B, C, D and A, respectively.*

Now one can wonder when is the quadrangle $H_aH_bH_cH_d$ cyclic and when will the quadrangle $J_aJ_bJ_cJ_d$ have perpendicular diagonals J_aJ_c and J_bJ_d. The answers give the following two theorems.

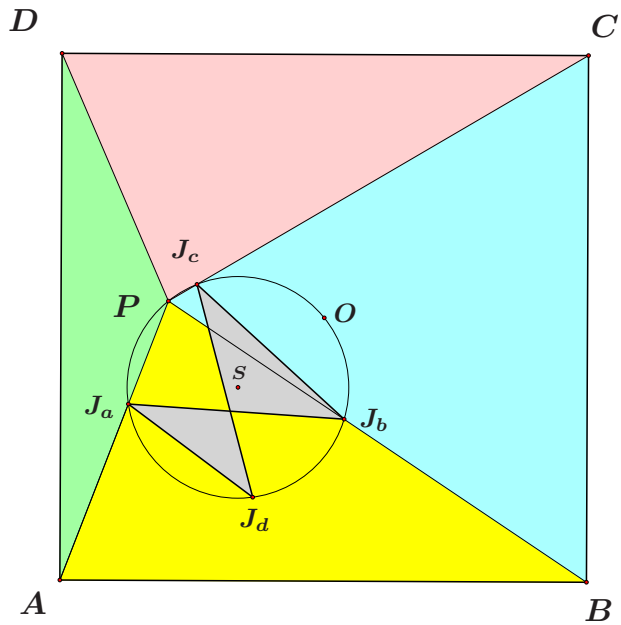

Figure 2: The cyclic quadrangle $J_aJ_bJ_cJ_d$.

Theorem 1.1. *The quadrangle $H_aH_bH_cH_d$ is cyclic if and only if the point P is either on the line AC, on the line BD or on the circumcircle k of the square $ABCD$.*

Theorem 1.2. *In the quadrangle $J_aJ_bJ_cJ_d$ the lines J_aJ_c and J_bJ_d are perpendicular if and only if the point P is either on the line AC, on the line BD or on the circumcircle k of the square $ABCD$.*

More precisely, $H_a = H_c$ and/or $H_b = H_d$ if and only if P is either on the line AC or on the line BD. Hence, the first two parts of the locus from Theorems 1.1 and 1.2 correspond to the case when the quadrangle $H_aH_bH_cH_d$ degenerates to a segment or a point and either $J_a = J_c$ or $J_b = J_d$. The role of the third part (the circumcircle k) is explained better by the following statement: If P is on the circumcircle k, then

(a) $H_aH_bH_cH_d$ is a square of side equal to the diagonals of $ABCD$ with P as the center whose diagonals H_aH_c and H_bH_d are parallel to the lines BC and AB,

(b) $J_aJ_bJ_cJ_d$ is also a square of side equal to the half of the diagonals of $ABCD$,

(c) $J_aJ_bJ_cJ_d$ and $H_aH_bH_cH_d$ are related by the homothety $h(N, 2)$ where N is the vertex of the square on the segment PO (see Figure 4).

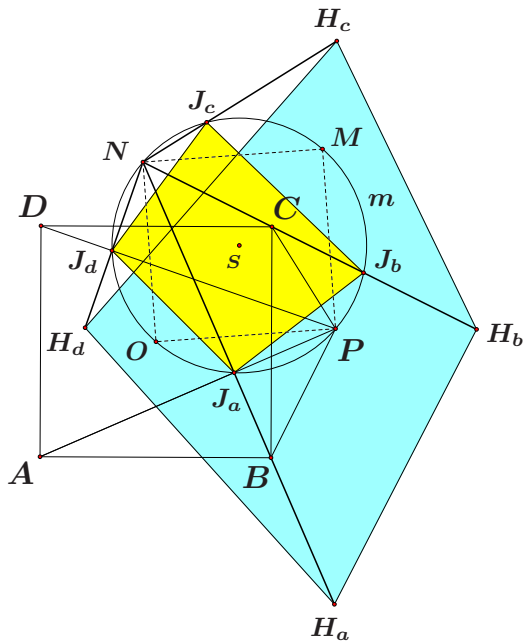

Figure 3: The lines H_aJ_a, H_bJ_b, H_cJ_c and H_dJ_d concur in the point N.

An interesting path is to explore how the areas of the quadrangles $H_aH_bH_cH_d$ and $J_aJ_bJ_cJ_d$ compare to the area Ω of the square $ABCD$. We define the area $|WXYZ|$ of the quadrangle $WXYZ$ as the sum $|WXY| + |WYZ|$ of (oriented) areas of the triangles WXY and WYZ.

For every real number m let P_m and Q_m denote the loci of all points P such that $|H_aH_bH_cH_d| = m\,\Omega$ and $|J_aJ_bJ_cJ_d| = m\,\Omega$, respectively.

Theorem 1.3. *For $m < 0$, $m = 0$, $0 < m < 2$, $m = 2$ and $m > 2$ the set P_m is the union of two hyperbolas, the union of lines AC and BD, the empty set, the circumcircle k of the square $ABCD$, and the union of two ellipsis, respectively. If the axes of one conic are φ and ψ then the axes of the other are ψ and φ.*

The Figures 5 and 6 show the sets P_m for $m = -1$ and $m = 3$ and the set $Q_{\frac{1}{2}}$ together with the square $ABCD$. Notice that $Q_{\frac{1}{2}}$ is the union of the circumcircle k of the square $ABCD$ and a symmetric curve of order six that touches k in the vertices of the square.

Let us conclude this description of our results with some comments on what else one can do with this approach. Instead of orthocenters we can consider other central points of the triangle (like the centroid, the circumcenter, the center of the nine-point circle – see the references [3] and [4] for the list of more than thou-

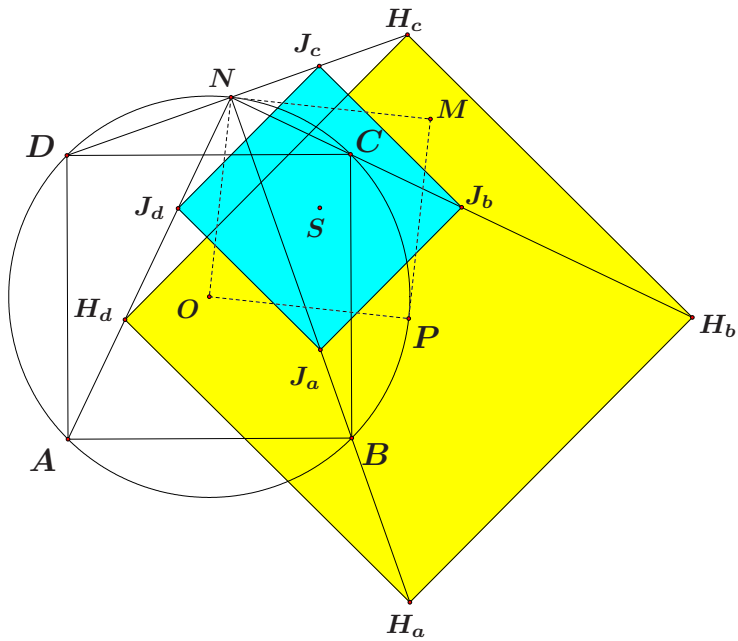

Figure 4: The squares $H_aH_bH_cH_d$ and $J_aJ_bJ_cJ_d$ when the point P is on the circumcircle of $ABCD$.

sand such points). For example, we have the following analogue of Property 2 for circumcenters.

Let O_a, O_b, O_c and O_d be the circumcenters of the triangles ABP, BCP, CDP and DAP, respectively. Let N_a, N_b, N_c and N_d denote the orthogonal projections of O_a, O_b, O_c and O_d onto the lines AP, BP, CP and DP, respectively.

Property 4. *The points N_a, N_b, N_c and N_d are vertices of the square which is related to the square $ABCD$ by the homothety $h(P, \frac{1}{2})$.*

A computer search reveals that the Steiner point gives the following result also similar to the Property 2.

Recall that the Steiner point is denoted as $X(99)$ in [3] and on page 120 of [1] it is noted that the Steiner point of a triangle is the center of mass of the system obtained by suspending at each vertex a mass equal to the magnitude of the exterior angle at that vertex. It is also the intersection of the circumcircle with the Steiner ellipse and the point of concurrency of parallels through the vertices to the corresponding sides of its first Brocard triangle (see [2]).

Let S_a, S_b, S_c and S_d be the Steiner points of the triangles ABP, BCP, CDP and DAP, respectively.

Property 5. *The points P, S_a, S_b, S_c and S_d lie on a circle whose center is*

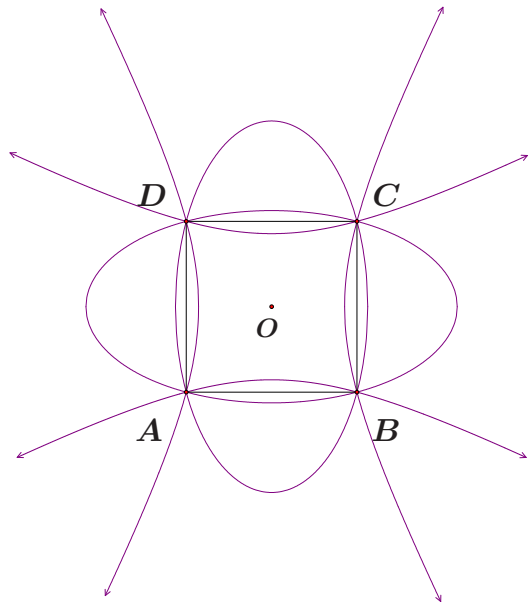

Figure 5: The loci P_{-1} (hyperbolas) and P_3 (ellipses).

on the line PO. In particular, the quadrangle $S_aS_bS_cS_d$ is cyclic.

In this article we took the square $ABCD$ as the underlying figure. Of course, it is possible to take instead any quadrangle or any triangle and perform similar constructions. The possibilities are numerous here but it remains to explore which of these choices give interesting results.

2. Primer on analytic plane geometry in Maple V

The key idea of the analytic geometry is to associate algebraic entities with geometric objects and then investigate them using algebraic methods.

The input of points on the plane in Maple V is quite simple because they are just ordered pairs of real numbers (their rectangular coordinates). For example, the input

```
tA:=[2, 3]:  tB:=[5, 7]:  tC:=[-2, 0]:  tT:=[x, y]:
```

defines four points on the plane $A(2,3)$, $B(5,7)$, $C(-2,0)$, $T(x,y)$.

We shall now list definitions of basic functions in Maple V for analytic geometry in the plane in rectangular coordinates. The first group of these functions are `FS`

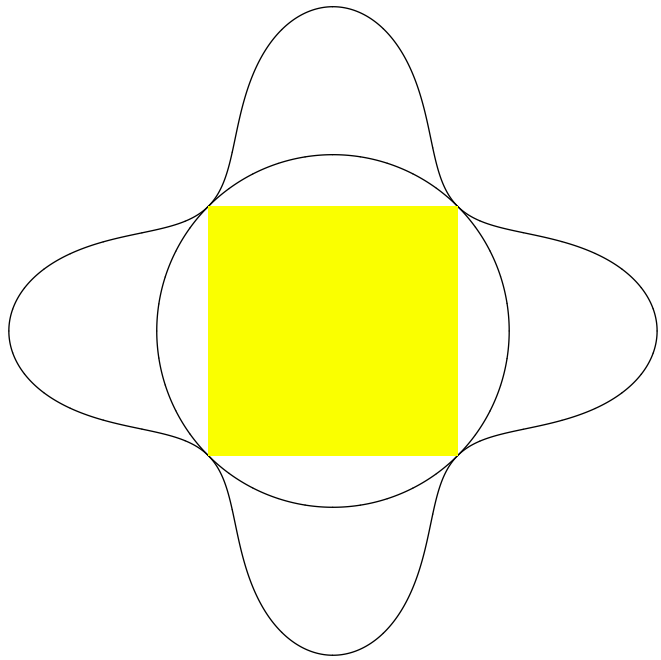

Figure 6: The locus $Q_{\frac{1}{2}}$.

(shortcut for composition of commands `simplify` and `factor`), `dis` (the distance of two points), `mid` (the midpoint of two points), `rat` (the point which divides two given points in given ratio $k \neq -1$), `rat2` (the point which divides two given points in given ratio $\frac{m}{n}$ where $m + n \neq 0$).

```
FS:=a->factor(simplify(a)):
dis:=(a,b)->FS(sqrt((a[1]-b[1])^2+(a[2]-b[2])^2)):
mid:=(a,b)->FS([(a[1]+b[1])/2,(a[2]+b[2])/2]):
rat:=(a,b,k)->FS([(a[1]+k*b[1])/(1+k),
                  (a[2]+k*b[2])/(1+k)]):
rat2:=(a,b,m,n)->FS([(n*a[1]+m*b[1])/(m+n),
                  (n*a[2]+m*b[2])/(m+n)]):
```

The lines in the program Maple V are represented as ordered triples $[a, b, c]$ of coefficients of their linear equations. For example, the input
`pX:=[1,0,0]: pY:=[0,1,0]: pD:=[1,-1,0]: pG:=[-1,2,2]:`
define the y-axis, the x-axis, the bisector of the first and the third quadrant and the line $-x + 2\,y + 2 = 0$.

We continue with functions `li1` (for a line through a given point with a given slope), `li2` (for a line through two given different points), `olQ` (to test if a point is

on a line), `clQ` (to test if three given points are collinear), and `ins` (the intersection of two lines or the information that they are parallel).

```
li1:=(a,k)->FS([k,-1,a[2]-k*a[1]]):
li2:=(a,b)->FS([a[2]-b[2],b[1]-a[1],a[1]*b[2]-b[1]*a[2]]):
olQ:=(t,p)->FS(t[1]*p[1]+t[2]*p[2]+p[3]):
clQ:=(a,b,c)->FS(a[1]*b[2]-a[1]*c[2]-b[1]*a[2]+
                    b[1]*c[2]+c[1]*a[2]-c[1]*b[2]):

ins:=(p,q)->FS([(q[3]*p[2]-q[2]*p[3])/(q[2]*p[1]-q[1]*p[2]),
                (q[1]*p[3]-q[3]*p[1])/(q[2]*p[1]-q[1]*p[2])]):
```

Functions `par` and `per` for the parallel and the perpendicular through a point to a line and tests `paQ` and `peQ` if two lines are parallel or perpendicular and the test `ccQ` for concurrency of three lines (i.e., whether they are parallel or intersect in a point) are next.

```
par:=(t,p)->FS([p[1],p[2],-t[1]*p[1]-t[2]*p[2]]):
per:=(t,p)->FS([p[2],-p[1],t[2]*p[1]-t[1]*p[2]]):
paQ:=(p,q)->FS(q[1]*p[2]-p[1]*q[2]):
peQ:=(p,q)->FS(p[1]*q[1]+p[2]*q[2]):

ccQ:=(a,b,c)->FS(a[1]*b[2]*c[3]-a[1]*b[3]*c[2]-b[1]*a[2]*
        c[3]+b[1]*a[3]*c[2]+c[1]*a[2]*b[3]-c[1]*a[3]*b[2]):
```

We conclude with the functions `pro` and `ar` for the orthogonal projection of a point onto a line and for the oriented area of a triangle on three given points.

```
pro:=(a,p)->FS([
  (p[2]*(a[1]*p[2]-a[2]*p[1])+p[1]*p[3])/(p[1]^2+p[2]^2),
  (p[1]*(a[2]*p[1]-a[1]*p[2])-p[2]*p[3])/(p[1]^2+p[2]^2)]):

ar:=(a,b,c)->FS((a[2]*c[1]-b[1]*a[2]-a[1]*c[2]+
                    a[1]*b[2]+b[1]*c[2]-c[1]*b[2])/2):
```

3. Central points functions

In this continuation of the previous section we shall describe functions for the central points that are mentioned in the introduction: the circumcenter, the orthocenter, and the Steiner point. On the way to define the Steiner point we also need functions for the symmedian point and the vertices of the first Brocard triangle.

First define the functions for the perpendicular bisector of a segment and the triangle circumcenter and orthocenter.

```
bis:=(a,b)->per(mid(a,b),li2(a,b)):
O_:=(a,b,c)->ins(bis(a,b),bis(a,c)):
H_:=(a,b,c)->ins(per(a,li2(b,c)),per(b,li2(c,a))):
```

The following function for the symmedian point is using the fact (see [1, p. 60g]) that symmedians bisect sides of the triangle from the projections A_h, B_h, C_h of vertices on opposite sidelines.

```
Ah_:=(a,b,c)->pro(a,li2(b,c)):
K_:=(a,b,c)->ins(li2(a,mid(Ah_(b,c,a),Ah_(c,a,b))),
                 li2(b,mid(Ah_(c,a,b),Ah_(a,b,c)))):
```

The vertices A_b, B_b, C_b of the first Brocard triangle are the orthogonal projections of the symmedian point onto the perpendicular bisectors of sides (see [1, p. 110] and Figure 7).

```
Ab_:=(a,b,c)->pro(K_(a,b,c),bis(b,c)):
```

Hence, the Steiner point is defined as follows (see Figure 8):

```
S_:=(a,b,c)->ins(par(a,li2(Ab_(b,c,a),Ab_(c,a,b))),
                 par(b,li2(Ab_(c,a,b),Ab_(a,b,c)))):
```

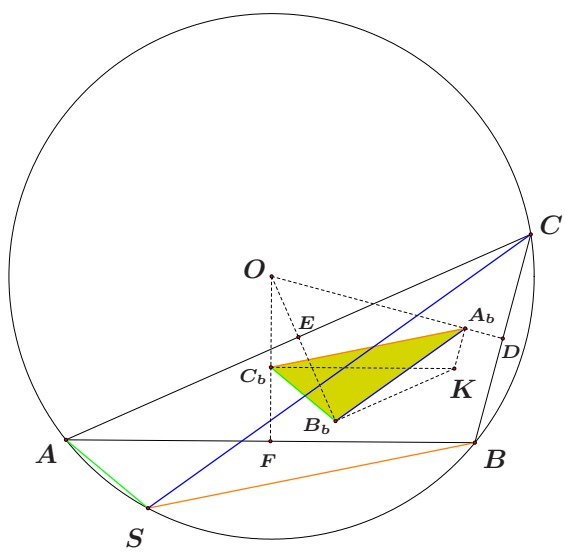

Figure 7: The Steiner point S is the intersection of parallels through vertices to sides of the first Brocard triangle $A_bB_bC_b$.

4. Verification of results

We shall now show how to prove most of the claims from the introduction with the help of computers in Maple V.

We first define the points A, B, C, D, O and P. They will be denoted with small letters because `D` is reserved in Maple V.

```
a:=[-1,-1]:b:=[1,-1]:c:=[1,1]:d:=[-1,1]:o:=[0,0]:p:=[x,y]:
```

Then we give the points U_t for $U = H$, J, O, N, S and $t = a$, b, c, d (all points mentioned in the introduction).

```
Ha:=H_(a,b,p):Hb:=H_(b,c,p):Hc:=H_(c,d,p):Hd:=H_(d,a,p):
Ja:=pro(Ha,li2(a,p)):Jb:=pro(Hb,li2(b,p)):
Jc:=pro(Hc,li2(c,p)):Jd:=pro(Hd,li2(d,p)):
Oa:=O_(a,b,p):Ob:=O_(b,c,p):Oc:=O_(c,d,p):Od:=O_(d,a,p):
Na:=pro(Oa,li2(a,p)):Nb:=pro(Ob,li2(b,p)):
Nc:=pro(Oc,li2(c,p)):Nd:=pro(Od,li2(d,p)):
Sa:=S_(a,b,p):Sb:=S_(b,c,p):Sc:=S_(c,d,p):Sd:=S_(d,a,p):
```

The proof of the Property 2 is accomplished with the following input.

```
s:=O_(o,p,Ja):FS(dis(s,o)-dis(s,Jb));
FS(dis(s,o)-dis(s,Jc));FS(dis(s,o)-dis(s,Jc));
```

Since the output is three times number 0 (zero), we conclude that the points O, P, J_a, J_b, J_c and J_d are on a circle. Its center S gives for the input

```
peQ(li2(o,s),li2(p,s));
```

the output 0, so that S is the center of the negatively oriented square $PONM$. Notice that the points N and M are `n:=rat(p,s,-2):` and `m:=rat(o,s,-2):`.

The proof of the Property 3 requires to see that we get zero as the output of each of the following eight commands.

```
clQ(n,Ha,Ja); clQ(n,Hb,Jb); clQ(n,Hc,Jc); clQ(n,Hd,Jd);
clQ(b,Ha,Ja); clQ(c,Hb,Jb); clQ(d,Hc,Jc); clQ(a,Hd,Jd);
```

Proof of Theorem 1.1. The quadrangle $H_aH_bH_cH_d$ is cyclic if and only if the distance between the circumcenters of the triangles $H_aH_bH_c$ and $H_aH_bH_d$ is equal to zero.

```
dis(O_(Ha,Hb,Hc),O_(Ha,Hb,Hd))^2;
```

This square of distance is equal to

$$\frac{T\,(x-y)^2\,(x+y)^2\left(x^2+y^2-2\right)^2\left(x^2+y^2+2\right)^2}{4\,\left(x^2-2\,x+y^2\right)^2\,(y-1)^2\,(y+1)^2\,(x-1)^2\,(x+1)^2\,\left(x^2+2\,y+y^2\right)^2},$$

where T is the polynomial

$$x^6+3\,x^4y^2+3\,x^2y^4+y^6-2\,x^5+6\,x^4y-8\,x^3y^2+8\,x^2y^3-6\,xy^4+2\,y^5+2\,x^4$$
$$-16\,x^3y+12\,x^2y^2-16\,xy^3+2\,y^4-4\,x^3+12\,x^2y-12\,xy^2+4\,y^3+4\,x^2+4\,y^2.$$

Since $x^2+y^2+2>0$ for all real numbers x and y and $x-y=0$, $x+y=0$ and $x^2+y^2-2=0$ are equations of the lines AC and BD and of the circumcircle k of the square $ABCD$, the claim of Theorem 1.1 follows provided we prove that the polynomial T is equal to zero only for P from the subset of the union $W=AC\cup BD\cup k$.

Let $x=r\cos\theta$ and $y=r\sin\theta$ for $r\geqslant 0$ and $0\leqslant\theta<2\pi$. Let u and v denote $3(\sin\theta-\cos\theta)+\sin 3\theta+\cos 3\theta$ and $3-8\sin 2\theta-\cos 4\theta$. Then $T=r^2\,U$ with $U=r^4+u\,r^3+v\,r^2+2\,u\,r+4$. Hence, it remains to show that the real roots of the equation **(*)** $U=0$ give points from the set W.

Note that $u^2-4v+16=4(1+\sin 2\theta)(4-(\sin 2\theta-1)^2)$ is always positive except for $\theta=\frac{3\pi}{4}$, $\frac{7\pi}{4}$ when it is equal to zero. Let w denote $\sqrt{u^2-4v+16}$.

Applying the basic command `solve` to (*) we see that its roots are

$$r_{1,2}=\frac{-u+w\pm\sqrt{H}}{4},\qquad r_{3,4}=\frac{-u-w\pm\sqrt{K}}{4},$$

where $H=L-2uw$ and $K=L+2uw$ with $L=u^2+w^2-32$. Note that L can be written as $16\left(2\,(\cos\theta)^2-1\right)^2(\sin\theta\,\cos\theta-1)$. Hence, L is always negative except for $\theta=\frac{\pi}{4}$, $\frac{3\pi}{4}$, $\frac{5\pi}{4}$, $\frac{7\pi}{4}$ when it is equal to zero and (*) has roots $\pm\sqrt{2}$.

Since $L^2-(2uw)^2=64(2(\cos\theta)^2-1)^4$ is always positive except for the values $\theta=\frac{\pi}{4}$, $\frac{3\pi}{4}$, $\frac{5\pi}{4}$, $\frac{7\pi}{4}$ and L is there always negative it follows that both H and K are negative so that the four roots above are not real unless they are $\pm\sqrt{2}$. □

Proof of Theorem 1.2. The output of the command

```
peQ(li2(Ja,Jc),li2(Jb,Jd));
```

is $\frac{8(x-y)\left(x^2+y^2+2\right)(x+y)\left(x^2+y^2-2\right)\left(x^2+y^2\right)}{(2+2\,y+y^2+2\,x+x^2)(2-2\,y+y^2-2\,x+x^2)(2+2\,y+y^2-2\,x+x^2)(2-2\,y+y^2+2\,x+x^2)}$. This is equal to zero (i.e., the quadrangle $J_aJ_bJ_cJ_d$ has perpendicular diagonals) if and only if the point P is in the set W. □

Proof of Theorem 1.3. The area Ω of the square $ABCD$ is 4. The output of `FS(ar(Ha,Hb,Hc)+ar(Ha,Hc,Hd)-4*m);` is $\frac{-2\,T}{(1+y)(1-x)(1-y)(1+x)}$, where T denotes the polynomial

$$2\,(1+y)\,(1-x)\,(1-y)\,(1+x)\,m+(x+y)^2\,(x-y)^2\,.$$

For $m\neq 0$, let $\alpha=-\frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{2}{m}}$ and $\beta=\frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{2}{m}}$. Then

$$T=-2m(\alpha\,x^2-\beta\,y^2+1)(\beta\,x^2-\alpha\,y^2-1).$$

For $m<0$, $\alpha>0$ and $\beta>1$ so that P_m is the union of two hyperbolas. For $m=0$, $T=(x-y)^2(x+y)^2$ and P_m is the union of the lines AC and BD. For $0<m<2$, the discriminant $4\,m\,(y-1)^2\,(y+1)^2\,(m-2)$ of T considered as a quadratic trinomial in x^2 is negative so that $T>0$ and P_m is empty. For $m=2$, $T=\left(x^2+y^2-2\right)^2$ and P_m is the circumcircle k. Finally, for $m>2$, $\alpha<0$ and $\beta>0$ so that P_m is the union of two ellipsis. □

Verification of Property 4. It suffices to note that the output for each of the commands

```
dis(Na,rat(p,a,1)); dis(Nb,rat(p,b,1));
dis(Nc,rat(p,c,1)); dis(Nd,rat(p,d,1));
```

is equal to zero. □

Verification of Property 5. It suffices to note that the output for the last three commands

```
t:=O_(Sa,Sb,Sc):  FS(dis(t,Sa)-dis(t,p));
FS(dis(t,Sa)-dis(t,Sd)); clQ(o,p,t);
```

is equal to zero. □

References

[1] Ross Honsberger, Episodes in nineteenth and twentieth century Euclidean geometry, The Mathematical Association of America, New Mathematical Library no. 37, Washington, 1995.

[2] R. A. Johnson, Advanced Euclidean Geometry, Dover Publications, New York, 1960.

[3] Clark Kimberling, Triangle Centers and Central Triangles, volume 129 of Congressus Numerantium, Utilitas Mathematica Publishing, Winnipeg, 1999.

[4] Clark Kimberling, Encyclopedia of Triangle Centers, 2000. (Internet address: `http://cedar.evansville.edu/~ck6/encyclopedia/`).

Zvonko Čerin
Kopernikova 7
10020 Zagreb
Croatia

Gian Mario Gianella
Dipartimento di Matematica
Universita di Torino
Torino
Italy

Annales Mathematicae et Informaticae
***33** (2006) pp. 35–43*
http://www.ektf.hu/tanszek/matematika/ami

Remarks on arithmetical functions $a_p(n)$, $\gamma(n)$, $\tau(n)$

Zoltán Fehér[a], Béla László[a], Martin Mačaj[b], Tibor Šalát[b]

[a]Department of Mathematics and Informatics, Faculty of Central European Studies
Constantine the Philosopher University
e-mail: zfeher@ukf.sk, blaszlo@ukf.sk

[b]Department of Algebra, Geometry and Mathematics Education
Faculty of Mathematics, Physics and Informatics
Comenius University
e-mail: martin.macaj@fmph.uniba.sk

Submitted 31 July 2005; Accepted 1 September 2006

Abstract

In this paper some properties of the arithmetical functions $a_p(n)$, $\gamma(n)$, $\tau(n)$ defined by Šalát in 1994 and Mycielski in 1951, respectively are investigated from the point of view of $\mathcal{I}$-convergence of sequences ($\mathcal{I}$-convergence was defined by Kostyrko, Šalát and Wilczynski in 2000).

1. Introduction

We shall study some properties of the $\mathcal{I}$–convergence of sequences of arithmetical functions $f\colon \mathbb{N} \to \mathbb{N}$, $a_p(n)$, $\gamma(n)$, $\tau(n)$. Elementary properties of the function $a_p(n)$ were studied in [6]. We shall extend these results with properties of $\mathcal{I}$–convergence of the sequence $(a_p(n))_{n=1}^{\infty}$.

We also want to investigate the asymptotic density of the sets $M_f = \{n : f(n) \mid n\}$ and the $\mathcal{I}$–convergence of arithmetical functions $\gamma(n)$, $\tau(n)$ defined by Mycielski in [4].

As usual we put for $A \subset \mathbb{N}$: $A(n) = |\{1, 2, \dots n\} \cap A|$,

$$\underline{d}(A) = \liminf \frac{A(n)}{n}, \overline{d}(A) = \limsup \frac{A(n)}{n}$$

the lower and upper density of A. If $\underline{d}(A) = \overline{d}(A)$, then we set

$$d(A) = \underline{d}(A) = \overline{d}(A), d(A) = \lim_{n\to\infty} \frac{A(n)}{n}.$$

The system $\mathcal{I} \subseteq 2^{\mathbb{N}}$ is called an admissible ideal if $\mathcal{I}$ is additive ($A, B \in \mathcal{I} \Rightarrow A \cup B \in \mathcal{I}$), hereditary ($A \in \mathcal{I}$, $B \subseteq A \Rightarrow B \in \mathcal{I}$) and contains all finite sets. In this paper we are interested in ideals $\mathcal{I}_f = \{A \subseteq \mathbb{N}, |A| < +\infty\}$, $\mathcal{I}_d = \{A \subseteq \mathbb{N} : d(A) = 0\}$, $\mathcal{I}_c = \{A \subseteq \mathbb{N} : \sum_{a\in A} a^{-1} < +\infty\}$ and $\mathcal{I}_c^q = \{A \subseteq \mathbb{N} : \sum_{a\in A} a^{-q} < +\infty\}$ for $q \in (0, 1)$. It is easy to see that for $q \leqslant q' \in (0, 1)$ the following inclusions hold:

$$\mathcal{I}_f \subseteq \mathcal{I}_c^q \subseteq \mathcal{I}_c^{q'} \subseteq \mathcal{I}_c \subseteq \mathcal{I}_d.$$

A given sequence $x = (x_n)_{n=1}^{\infty}$ of real numbers is said to be $\mathcal{I}$–convergent to $L \in R$, if for each $\varepsilon > 0$ we have $A_\varepsilon = \{n : \left|x_n - L\right| \geqslant \varepsilon\} \subseteq \mathcal{I}$ (shortly $\mathcal{I}$–$\lim x_n = L$). The cases of $\mathcal{I}_f$-convergence and $\mathcal{I}_d$-convergence coincide with the usual convergence and the statistical convergence (see [3], [7]), respectively. Therefore we will write $\lim x_n = L$ and $\lim \mathrm{stat}\, x_n = L$ instead of $\mathcal{I}_f$–$\lim x_n = L$ and $\mathcal{I}_d$–$\lim x_n = L$, respectively.

In [7, Lemma 2.2] it is shown that

$$\mathcal{I} \subseteq \mathcal{I}' \Rightarrow \mathcal{I} - \lim x_n = L \Rightarrow \mathcal{I}' - \lim x_n = L.$$

Using this result we completely determine for which q the sequences $a_p(n)$, $\gamma(n)$ and $\tau(n)$ are $\mathcal{I}_c^q$-convergent.

2. $\mathcal{I}$-convergence of $(a_p(n))_{n=1}^{\infty}$

Let p be a prime number. The function $a_p(n)$ is defined in the following way: $a_p(1) = 0$ and if $n > 1$, then $a_p(n)$ is the unique integer $j \geqslant 0$ satisfying $p^j | n$ but $p^{j+1} \nmid n$, i.e., $p^{a_p(n)} \parallel n$. At first we are going to generalize the result that the sequence $\left((\log p)\frac{a_p(n)}{\log n}\right)_{n=2}^{\infty}$ is statistically convergent to 0 [6, Th. 4.2].

Proposition 2.1. *Let $g(n) > 0\,(n = 1, 2\dots)$ and $\lim_{n\to\infty} g(n) = +\infty$. We have*

$$\lim \mathrm{stat}(\log p)\frac{a_p(n)}{g(n)} = 0.$$

Proof. Let $\varepsilon > 0$. Put $A_\varepsilon = \{n > 1 : (\log p)\frac{a_p(n)}{g(n)} \geqslant \varepsilon\}$. We will show that $d(A_\varepsilon) = 0$. Let $\eta > 0$. Choose $m \in N$ such that

$$p^{-m} < \eta. \tag{2.1}$$

By the conditions of the proposition there exists an n_0, such that for any $n > n_0$ we have

$$\frac{\varepsilon g(n)}{\log p} > m. \tag{2.2}$$

Let $n > n_0$ and $n \in A_\varepsilon$. It follows from (2.2) and the definition of A_ε that

$$(\log p)\frac{a_p(n)}{g(n)} \geqslant \varepsilon,$$

$$a_p(n) \geqslant \frac{\varepsilon g(n)}{\log p} > m.$$

Hence for the numbers $n > n_0, n \in A_\varepsilon$ implies $p^m | n$. This leads to the conclusion that $A_\varepsilon \subseteq \{1, 2, \ldots, n_0\} \cup \{n > n_0 : p^m \mid n\}$ and considering (2.1) we get $\overline{d}(A_\varepsilon) \leqslant p^{-m} < \eta$. Since $\eta > 0$ is an arbitrary positive number, $d(A_\varepsilon) = 0$. □

Remark 2.2. It is proved [6, Th. 4.1] that the sequence $\left((\log p)\frac{a_p(n)}{\log n}\right)_{n=2}^{\infty}$ is dense in interval $(0, 1)$. But $\left((\log p)\frac{a_p(n)}{g(n)}\right)_{n=2}^{\infty}$ which is statistically convergent to zero if $g(n) \to +\infty$, is not always dense in $(0, 1)$: For example if we define the function $g(n) = max\{1, \log^2 n\}$, then we have

$$\lim_{n\to\infty} (\log p)\frac{a_p(n)}{\log^2 n} = 0$$

and also

$$\lim \operatorname{stat} \frac{a_p(n)}{\log^2 n} = 0,$$

but this sequence is not dense in $(0, 1)$.

Theorem 2.3. *The sequence $(a_p(n))_{n=1}^{\infty}$ is I_c–convergent to 0 and $\mathcal{I}_c^q$–divergent for $q \in (0, 1)$.*

Proof. Let $\varepsilon > 0$ and denote

$$A_\varepsilon = \{n \in \mathbb{N} : (\log p)\frac{a_p(n)}{\log n} \geqslant \varepsilon\}.$$

Let $q \in (0, 1)$. We want to show that

$$\sum_{n\in A_\varepsilon} \frac{1}{n} < +\infty \tag{2.3}$$

and for $0 < \varepsilon < 1 - q$

$$\sum_{n\in A_\varepsilon} \frac{1}{n^q} = +\infty. \tag{2.4}$$

For nonnegative integer i denote $A_\varepsilon^i = \{n \in A_\varepsilon; n = p^i u, (u, p) = 1\}$. We have $A_\varepsilon^i \cap A_\varepsilon^j = \emptyset$ for $i \neq j$ and for any $t > 0$

$$\sum_{n\in A_\varepsilon} \frac{1}{n^t} = \sum_{i=0}^{\infty} \sum_{n\in A_\varepsilon^i} \frac{1}{n^t}. \tag{2.5}$$

a) Consider that $n \in A_\varepsilon^i$ if and only if $n = p^i u$ where $(u,p) = 1$ and also

$$(\log p)\frac{a_p(n)}{\log n} \geqslant \varepsilon.$$

Then

$$(\log p)\frac{i}{i\log p + \log u} \geqslant \varepsilon$$

from which we obtain $u \leqslant p^{i\delta}$, where $\delta = (1-\varepsilon)/\varepsilon$. Hence

$$\sum_{n\in A_\varepsilon^i}\frac{1}{n} \leqslant \frac{1}{p^i}\sum_{u\leqslant p^{i\delta}}\frac{1}{u} \leqslant \frac{1}{p^i}\left(1+\int_1^{p^{i\delta}} dt/t\right) = \frac{1}{p^i}(1+i\delta\log p) \leqslant A\delta\frac{i}{p^i}\log p$$

where $A > 0$ is only dependent on ε, p and not on i. The series $\sum\limits_{i=0}^{\infty}\frac{i}{p^i}$ converges, this proves (2.3).

b) We write

$$\sum_{n\in A_\varepsilon^i}\frac{1}{n^q} = \frac{1}{p^{iq}}\sum_{\substack{u\leqslant p^{i\delta}\\(u,p)=1}}\frac{1}{u^q}.$$

Then we have

$$\begin{aligned}\sum_{\substack{u\leqslant p^{i\delta}\\(u,p)=1}}\frac{1}{u^q} &= \sum_{u\leqslant p^{i\delta}}\frac{1}{u^q} - \sum_{k\leqslant p^{i\delta-1}}\frac{1}{(kp)^q} = \sum_{u\leqslant p^{i\delta}}\frac{1}{u^q} - \frac{1}{p^q}\sum_{k\leqslant p^{i\delta-1}}\frac{1}{k^q}\\ &= \left(1-\frac{1}{p^q}\right)\sum_{v\leqslant p^{i\delta-1}}\frac{1}{v^q} + \sum_{p^{i\delta-1}<v\leqslant p^{i\delta}}\frac{1}{v^q}\\ &\geqslant \sum_{p^{i\delta-1}<v\leqslant p^{i\delta}}\frac{1}{v^q} \geqslant (p^{i\delta}-p^{i\delta-1})\frac{1}{p^{i\delta q}}\\ &= p^{i\delta}(1-\frac{1}{p})\frac{1}{p^{i\delta q}} = (1-\frac{1}{p})p^{i\delta(1-q)}.\end{aligned}$$

Finally we obtain

$$\sum_{n\in A_\varepsilon}\frac{1}{n^q} = \sum_{i=0}^{\infty}\sum_{v\in A_\varepsilon^i}\frac{i}{v^q} \geqslant (1-\frac{1}{p})\sum_{i=0}^{\infty}\frac{1}{p^{i[q+(q-1)\delta]}}.$$

The series on the right-hand side diverges if $q+(q-1)\delta < 0$, i.e. $\varepsilon < 1-q$. This proves the I_c^q–divergence of $(a_p(n))_{n=1}^{\infty}$. □

3. On the functions $\gamma(n)$ and $\tau(n)$

In [4] there were new arithmetical functions defined and investigated in connection with the representation of natural numbers of the form $n = a^b$, where a, b are positive integers. Let

$$n = a_1^{b_1} = a_2^{b_2} = \cdots = a_{\gamma(n)}^{b_{\gamma(n)}} \tag{3.1}$$

be all such representations of a given natural number n, where $a_i, b_i \in N$.

Denote by

$$\tau(n) = b_1 + \cdots + b_{\gamma(n)}, (n > 1).$$

It is clear that $\gamma(n) \geqslant 1$, because for any $n > 1$ there exists a representation in the form n^1.

We are going to study some new properties of the functions $\gamma(n)$ and $\tau(n)$.

Put $T(n) = \gamma(2) + \cdots + \gamma(n), (n \geqslant 2)$. It is proved in [4], that

$$T(n) = \sum_{s=1}^{[\log_2 n]} [\sqrt[s]{n}] - [\log_2 n] = n + \sum_{s=2}^{[\log_2 n]} [\sqrt[s]{n}] - [\log_2 n]. \tag{3.2}$$

Remark 3.1. It is easy to show that the average order of the function $\gamma(n)$ is 1, i.e.,

$$\lim_{n\to\infty} \frac{T(n)}{n} = 1.$$

It follows from (3.2) that

$$T(n) = n + T_1(n) - [\log_2 n],$$

where $T_1(n) = n + \sum\limits_{s=2}^{[\log_2 n]} [\sqrt[s]{n}]$. Then simple estimations give

$$([\log_2 n] - 1)[\sqrt[[\log_2 n]]{n}] \leqslant T_1(n) \leqslant ([\log_2 n] - 1)\sqrt{n}$$

from which we get $\lim\limits_{n\to\infty} \frac{T_1(n)}{n} = 0$.

In papers [1, 2] sets of the form $M_f = \{n \in \mathbb{N} : f(n) \mid n\}$, $f : \mathbb{N} \to \mathbb{N}$ are investigated. For some of the known arithmetical functions the sets M_f have zero asymptotic density: e.g. the functions $\omega(n)$ (the number of prime divisors of n), $s_g(n)$ (the digital sum of n in the representation with base g), $\pi(n)$ (the number of primes not exceeding n).

Proposition 3.2. *Put* $A_k = \{n > 1 : n = p_1^{\alpha_1} \dots p_n^{\alpha_n}, (\alpha_1, \dots, \alpha_n) = k\}$ $(k = 1, 2, \dots)$. *Then*

$$d(A_1) = 1. \tag{3.3}$$

Proof. Denote by $B = \cup_{k=2}^{\infty} A_k$, then $\mathbb{N}\backslash\{1\} = A_1 \cup B$, where $A_1 \cap B = \emptyset$. It can be easily shown that $d(B) = 0$, from which (3.3) follows immediately. The elements of the set B are only numbers of the form $t^s (t > 1,\ s > 1)$. Denote by H the set of all numbers $t^s (t > 1,\ s > 1)$. The series of reciprocal values of these numbers is equal to $\sum\limits_{t=2}^{\infty} \sum\limits_{s=2}^{\infty} \frac{1}{t^s}$ which is convergent to 1 (cf. [4]). Then we have $d(H) = 0$ and it implies that also $d(B) = 0$. □

Let us investigate the asymptotic density of $M_\gamma = \{n : \gamma(n) \mid n\}$ and $M_\tau = \{n : \tau(n) \mid n\}$.

Proposition 3.3. *We have*
(i) $d(M_\gamma) = 1$,
(ii) $d(M_\tau) = 1$.

Proof. (i) If $n \in A_1$, then evidently $\gamma(n) = 1$ and $n \in M_\gamma$. Thus $A_1 \subseteq M_\gamma$ and considering (3.3) we get $d(M_\gamma) = 1$.
(ii) Similarly. □

In [4, Th. 3, Th. 5] there are proofs of the following results:

$$\sum_{n=2}^{\infty} \frac{\gamma(n) - 1}{n} = 1, \sum_{n=2}^{\infty} \frac{\tau(n) - 1}{n} = 1 + \frac{\pi^2}{6}.$$

In connection with these results we have investigated the convergence of series for any $\alpha \in (0, 1)$

$$\sum_{n=2}^{\infty} \frac{\gamma(n) - 1}{n^\alpha}, \sum_{n=2}^{\infty} \frac{\tau(n) - 1}{n^\alpha}.$$

Theorem 3.4. *The series*

$$\sum_{n=2}^{\infty} \frac{\gamma(n) - 1}{n^\alpha}$$

diverges for $0 < \alpha \leqslant \frac{1}{2}$ *and converges for* $\alpha > \frac{1}{2}$.

Proof. a) Let $0 < \alpha \leqslant \frac{1}{2}$. Put $K = \{k^2 : k > 1\}$. A simple estimation gives

$$\sum_{n=2}^{\infty} \frac{\gamma(n) - 1}{n^\alpha} \geqslant \sum_{n \in K} \frac{\gamma(n) - 1}{n^\alpha}.$$

Clearly $\gamma(n) \geqslant 2$ for $n \in K$. Therefore

$$\sum_{n=2}^{\infty} \frac{\gamma(n) - 1}{n^\alpha} \geqslant \sum_{n \in K} \frac{1}{n^\alpha} = \sum_{k=2}^{\infty} \frac{1}{k^{2\alpha}} \geqslant \sum_{k=2}^{\infty} \frac{1}{k} = +\infty. \tag{3.4}$$

b) Let $\alpha > \frac{1}{2}$. We will use the formula

$$\sum_{n=2}^{\infty} \frac{\gamma(n)-1}{n^{\alpha}} = \sum_{k=2}^{\infty} \sum_{s=2}^{\infty} \frac{1}{k^{\alpha s}} = \sum_{k=2}^{\infty} \frac{1}{k^{\alpha}(k^{\alpha}-1)}. \tag{3.5}$$

For a sufficiently large number k ($k > k_0$) we have $\frac{k^{\alpha}}{k^{\alpha}-1} < 2$. We can estimate the series on the right-hand side of (3.5) with

$$\sum_{k=2}^{\infty} \frac{1}{k^{\alpha}(k^{\alpha}-1)} < \sum_{k=2}^{k_0} \frac{1}{k^{\alpha}(k^{\alpha}-1)} + 2 \sum_{k>k_0} \frac{1}{k^{2\alpha}}.$$

Since $2\alpha > 1$ we get

$$\sum_{n=2}^{\infty} \frac{\gamma(n)-1}{n^{\alpha}} < +\infty.$$

□

Corollary 3.5. *The sequence $\gamma(n)$ is*
(i) *$\mathcal{I}_c$-convergent to 1,*
(ii) *$\mathcal{I}_c^q$–divergent for $q \in (0, \frac{1}{2}]$ and $\mathcal{I}_c$–convergent to 1 for $q \in \left(\frac{1}{2}, 1\right)$.*

Proof. (i) Let $\varepsilon > 0$. The set of numbers $\{n > 1 : \left|\gamma(n) - 1\right| \geqslant \varepsilon\}$ is a subset of $H = \{t^s, t > 1, s > 1\}$ and $\sum\limits_{a \in H} \frac{1}{a} < +\infty$. From the definition of I_c–convergence (i) follows.

(ii) Let $\varepsilon > 0$ and denote $A_{\varepsilon} = \{n \in \mathbb{N} : \left|\gamma_n - 1\right| \geqslant \varepsilon\}$. When $0 < q \leqslant \frac{1}{2}$ then for the numbers $n \in K$, $K = \{k^2 : k > 1\}$ considering (3.4) holds

$$\sum_{n \in A_{\varepsilon}} \frac{1}{n^{\alpha}} \geqslant \sum_{n \in K} \frac{1}{n^{\alpha}} \geqslant +\infty.$$

Therefore $\gamma(n)$ is $\mathcal{I}_c^q$–divergent. When $\frac{1}{2} < q < 1$, then $A_{\varepsilon} \subset H$ and

$$\sum_{n=2}^{\infty} \frac{1}{n^{\alpha}} \leqslant \sum_{k=2}^{\infty} \sum_{s=2}^{\infty} \frac{1}{k^{\alpha s}}.$$

The convergence of the series on the right-hand side we proved previously in Theorem 3.4. Therefore $\gamma(n)$ is $\mathcal{I}_c$–convergent to 1 if $q \in (\frac{1}{2}, 1)$. □

Remark 3.6. We have $\lim \operatorname{stat} \gamma(n) = 1$.

Theorem 3.7. *The series*

$$\sum_{n=2}^{\infty} \frac{\tau(n)-1}{n^{\alpha}}$$

diverges for $0 < \alpha \leqslant \frac{1}{2}$ and converges for $\alpha > \frac{1}{2}$.

Proof. Let $0 < \alpha < 1$. We write the given series in the form

$$\sum_{n=2}^{\infty} \frac{\tau(n)-1}{n^{\alpha}} = \sum_{k=2}^{\infty} \sum_{s=2}^{\infty} \frac{s}{k^{\alpha s}}, \tag{3.6}$$

We shall try to use a similar method to Mycielski's proof of the convergence of $\sum\limits_{n=2}^{\infty} \frac{\tau(n)-1}{n^{\alpha}}$ to explain the equality (3.6). Since $\frac{s}{k^{\alpha s}} = -\frac{k}{\alpha}\frac{d}{dt}(\frac{1}{t^{\alpha s}})_{t=k}$ and $\sum\limits_{s=2}^{\infty} \frac{1}{t^{\alpha s}} = \frac{1}{t^{\alpha}(t^{\alpha}-1)}$ the right-hand side of (3.6) is equal to

$$\sum_{s=2}^{\infty} \frac{2k^{\alpha}-1}{k^{\alpha}(k^{\alpha}-1)^2} = \sum_{s=2}^{\infty} a_k.$$

For the k-th term of $\sum a_k$ we have

$$a_k = \frac{2-\frac{1}{k^{\alpha}}}{(1-\frac{1}{k^{\alpha}})^2} \cdot \frac{1}{k^{2\alpha}}.$$

Denote by $b_k = \frac{1}{k^{2\alpha}}$ and consider that $\lim\limits_{k\to\infty} \frac{a_k}{b_k} = 2$. Hence the series $\sum\limits_{s=2}^{\infty} a_k$ converges (diverges) if and only if the series $\sum\limits_{s=2}^{\infty} b_k$ converges (diverges). Since $\sum b_k$ is convergent (divergent) for any $\alpha > \frac{1}{2}$ $(0 < \alpha \leqslant \frac{1}{2})$ so does the series $\sum a_k$ and therefore the series $\sum \frac{\tau(n)-1}{n^{\alpha}}$. □

Corollary 3.8. *The sequence $\tau(n)$ is*
(i) *$\mathcal{I}_c$–convergent to 1,*
(ii) *$\mathcal{I}_c^q$–divergent for $q \in (0, \frac{1}{2}]$ and $\mathcal{I}_c$–convergent to 1 for $q \in \left(\frac{1}{2}, 1\right)$.*

Proof. Similar to the proof of Corollary 3.5. □

Remark 3.9. We have $\lim \operatorname{stat} \tau(n) = 1$.

References

[1] Cooper, C. N., Kennedy, R. E., *Chebyshev's inequality and natural density*, AMM 96 (1998) 118–124.

[2] Erdős, P., *Pomerance, C., On a theorem of Besicovitch: values of arithmetical functions that divide their arguments*, Indian J. Math. 32 (1990) 279–287.

[3] Kostyrko, P., Šalát, T., Wilczynski, W., *I–convergence*, Real Anal. Exchange 26 (2000–2001), 669–686.

[4] Mycielski, J., *Sur les reprĆsentations des nombres naturels par des puissances a base et exposant naturels*, Coll. Math. II (1951) 254–260.

[5] Powel, B. J., Šalát, T., *Convergence of subseries of the harmonic series and asymptotic densities of sets of positive integers*, Publ. de L'institut math., vol. 50. (64) (1991) 60–70.

[6] Šalát, T., *On the function a_p, $p^{a_p(n)} \parallel n\,(n > 1)$*, Math. Slov. 44 (1994) No. 2, 143–151.

[7] Šalát, T., Toma, V., *A classical Olivier's theorem and statistical convergence*, Annales Math. B. Pascal 10 (2003) 305–313.

[8] Schinzel, A., Šalát, T., *Remarks on maximum and minimum exponents in factoring*, Math. Slov. 44 (1994) 505–514.

Zoltán Fehér, Béla László
Department of Mathematics and Informatics
Faculty of Central European Studies
Constantine the Philosopher University
Tr. A. Hlinku 1
949 74 Nitra
Slovak Rep.

Martin Mačaj, Tibor Šalát
Department of Algebra, Geometry and Mathematics Education
Faculty of Mathematics, Physics and Informatics
Comenius University
Mlynska Dolina
842 48 Bratislava
Slovak Rep.

Annales Mathematicae et Informaticae
***33** (2006) pp. 45–56*
http://www.ektf.hu/tanszek/matematika/ami

On prime divisors of remarkable sequences

Ferdinánd Filip[a], Kálmán Liptai[b1], János T. Tóth[c2]

[a]Department of Mathematics University of J. Selye
e-mail: filip.ferdinand@seznam.cz

[b]Institute of Mathematics and Informatics Eszterházy Károly College
e-mail: liptaik@ektf.hu

[c]Department of Mathematics University of Ostrava
e-mail: toth@osu.cz

Submitted 10 November 2006; Accepted 18 December 2006

Abstract

In this paper we study sequences of the form $(a^n+b)_{n=1}^{\infty}$, where $a, b \in \mathbb{N}$. We prove many interesting results connection with sequences with infinitely many prime divisors.

Keywords: prime divisors, Dirichlet's theorem

MSC: 11N13

1. Introduction

There are many mathematical problems when we investigate the divisibility of sequences by a prime. We usually find this kind of interesting examples in national mathematical competitions and in the International Math Olympiad. In this paper we study sequences of the form $(a^n + b)_{n=1}^{\infty}$, where $a, b \in \mathbb{N}$. We prove some results concerning with sequences with infinitely many prime divisors. Moreover we characterize these sequences. Some of our theorems assert that there are infinitely many prime divisors of a sequence. These statements come from easily from the theory of S-units, but in this paper we use only elementary methods to get our results. We mention that our results help to generalize problems which can be found in some exercise books for students.

[1]Research supported by the Hungarian National Foundation for Scientific Research Grant. No. T 048945 MAT

[2]Research supported by Grant ČR 201/04/0381/2

Let $A = \{a_1 < a_2 < \cdots < a_n < \cdots\} \subseteq \mathbb{N}$ be a given set and let us denote by $A(x)$ the number of the elements of A not exceeding x. Let us suppose for any natural number k there is a positive real number x_k such that for all $x > x_k$ the inequality $A(x) > (\log x)^k$ holds. In this case there are infinitely many different prime divisors of the elements of A (see [3], p. 102).

Further we shall study the sequences of positive integers where the previous condition is not true. Let a, b be natural numbers with $a > 1$ and $(a, b) = 1$. Obviously the sequences

$$(a^n + b)_{n=1}^{\infty} \tag{1.1}$$

do not fulfill the above condition, since

$$A(x) = \left[\frac{\log(x-b)}{\log a}\right] \quad \text{if} \quad x > b + 1.$$

In what follows we show that sequences (1.1) have infinitely many different prime divisors. In the special case, when $a = 10$ and $b = 3$ we proved (in [6]) that the sequence $(10^n + 3)_{n=1}^{\infty}$ has infinitely many prime divisors, moreover for infinitely many primes p there are infinitely many $n \in \mathbb{N}$ such that $p \mid 10^n + 3$.

2. Results

First we prove that there are subsequences of sequences (1.1) which have infinitely many prime divisors.

Theorem 2.1. *Let a, b, c, d be natural numbers, $(a, b) = 1$ and $a > 1$. Then there are infinitely many prime divisors of the sequences*

$$(a^{c+(n-1)d} + b)_{n=1}^{\infty}. \tag{2.1}$$

Proof. First we suppose that sequence (2.1) has only finitely many prime divisors. Let us denote these primes by $q_1 < q_2 < \cdots < q_k$. Let us denote by $q_1 < q_2 < \cdots < q_l$ the prime divisors of sequence (2.1) which are divisors of $a^c + b$ as well and $q_{l+1} < q_{l+2} < \cdots < q_k$ which are not divisors of $a^c + b$. Let us denote by α_s for all $1 \leqslant s \leqslant l$ and $s \in \mathbb{N}$ the least natural number such that

$$q_s^{\alpha_s} > a^c + b.$$

Let

$$M = q_1^{\alpha_1} q_2^{\alpha_2} \cdots q_l^{\alpha_l} q_{l+1} q_{l+2} \cdots q_k$$

be a product of prime powers. In this case $(a, M) = 1$ since $(a, b) = 1$. By the theorem of Euler we have

$$M \mid a^{n\varphi(M)} - 1 \tag{2.2}$$

for all $n \in \mathbb{N}$.

Now we investigate the sequence

$$(a^{c+m\varphi(M)d}+b)_{m=1}^{\infty} \tag{2.3}$$

which is obviously a subsequence of sequence (2.1).

Let q be a prime divisor of sequence (2.3) that is

$$a^{m\varphi(M)d+c}+b\equiv 0 \pmod{q} \tag{2.4}$$

for some $m\in\mathbb{N}$. It follows from (2.2) that

$$a^{m\varphi(M)d}-1\equiv 0 \pmod{q}. \tag{2.5}$$

Using (2.4) and (2.5) we have

$$a^{m\varphi(M)d+c}+b=a^{m\varphi(M)d}(a^c-1)+a^{m\varphi(M)d}-1+b+1\equiv a^c+b \pmod{q}.$$

It is clear that $q\mid a^c+b$, it follows that $q\in\{q_1,q_2,\dots,q_l\}$, that is

$$a^{m\varphi(M)d+c}+b=q_1^{\beta_{m_1}}q_2^{\beta_{m_2}}\cdots q_l^{\beta_{m_l}}$$

where $\beta_{m_j}\geqslant 0$ for all $m\in\mathbb{N}$ and $1\leqslant j\leqslant l$.

We show that for all $m\in\mathbb{N}$ and $1\leqslant j\leqslant l$ we have $\beta_{m_j}<\alpha_j$. Let $1\leqslant j\leqslant l$, m be arbitrary natural numbers and $\beta_{m_j}\geqslant\alpha_j$ then

$$q_j^{\alpha_j}\mid a^{m\varphi(M)d+c}+b,$$

that is

$$a^{m\varphi(M)d+c}+b\equiv 0 \pmod{q_j^{\alpha_j}}.$$

Since $q_j^{\alpha_j}\mid M$, it follows from (2.2) that $a^{m\varphi(M)d}-1\equiv 0 \pmod{q_j^{\alpha_j}}$ and

$$a^{m\varphi(M)d+c}+b=a^{m\varphi(M)d}(a^c-1)+a^{m\varphi(M)d}-1+b+1\equiv a^c+b \pmod{q_j^{\alpha_j}},$$

that is $q_j^{\alpha_j}\mid a^c+b$, which is contradiction since $q_j^{\alpha_j}>a^c+b$. It follows that for all terms of (2.3) we have

$$a^{m\varphi(M)d+c}+b<q_1^{\alpha_1}q_2^{\alpha_2}\cdots q_l^{\alpha_l}\leqslant M\,.$$

In this way we obtained a contradiction since sequence (2.3) is not bounded. □

In the sequel we prove an interesting property of the prime divisors of sequence (2.1).

Theorem 2.2. *If $m\in\mathbb{N}$ is a divisor of a term of sequence* (2.1) *then m divides infinitely many terms of sequence* (2.1).

Proof. Let $m \in \mathbb{N}$ be a divisor of a term of sequence (2.1). Let us denote by n_0 the least non-negative number which

$$m \mid a^{c+n_0 d} + b\,. \tag{2.6}$$

Since $(a, m) = 1$, there exists a power h_m of $a \pmod m$. The number m divides $a^n - 1$ if and only if $h_m \mid n$.

Let us consider the sequence

$$(a^{n_k d+c} + b)_{n=1}^{\infty} \tag{2.7}$$

where

$$n_k = (k-1)\frac{h_m}{(h_m, d)} + n_0\,.$$

Obviously sequence (2.7) is a subsequence of sequence (2.1).

We show that m divides only those terms of sequence (2.1) which are the terms of (2.7) as well.

a) First we prove that m divides all terms of sequence (2.7). Obviously we have

$$\begin{aligned} a^{n_k d+c} + b &= a^{n_k d+c} + b - a^{n_0 d+c} + a^{n_0 d+c} = \\ &= a^{n_0 d+c}(a^{(n_k-n_0)d} - 1) + a^{n_k 0+c} + b = \\ &= a^{n_0 d+c}\big(a^{(k-1)\frac{h_m d}{(h_m,d)}} - 1\big) + a^{n_0 d+c} + b. \end{aligned} \tag{2.8}$$

Using that $\frac{d}{(h_m,d)}$ is an integer number and the definition of h_m we have

$$a^{(k-1)\frac{d}{(h_m,d)}h_m} - 1 \equiv 0 \pmod m.$$

It follows that

$$a^{n_k d+c} + b \equiv a^{n_0 d+c} + b \pmod m,$$

that is m divides all terms of (2.7).

b) Secondly we prove that if m divides a term of sequence (2.1) then this term is a term of sequence (2.7).

Let us choose $n \in \mathbb{N}$ such that $m \mid a^{nd+c_1} + b$. Obviously $n \geqslant n_0$. Then we have

$$m \mid a^{nd+c_1} + b - \big(a^{n_0 d+c_1} + b\big) = a^{n_0 d+c_1}\big(a^{d(n-n_0)} - 1\big)\,.$$

Since $(a, m) = 1$, therefore $m \mid a^{d(n-n_0)} - 1$. Using the definition of h_m we have $h_m \mid d(n - n_0)$, and

$$n = (k-1)\frac{h_m}{d} + n_0 \tag{2.9}$$

for some $k \in \mathbb{N}$. From equation (2.9) we deduce

$$n = \left((k-1)\frac{\frac{h_m}{(h_m,d)}}{\frac{d}{(h_m,d)}}\right) + n_0\,.$$

Using that $\left(\frac{h_m}{(h_m,d)}, \frac{d}{(h_m,d)}\right) = 1$, we have that n is an integer if and only if the fraction $\frac{k-1}{\frac{d}{(h_m,d)}}$ is also an integer. Consequently

$$k - 1 = (l-1)\frac{d}{(h_m, d)}\,,$$

and

$$n = (l-1)\frac{h_m}{(h_m, d)} + n_0.$$

Now the theorem is proved. □

In the previous theorems we investigated such subsequences of sequences (1.1) where the powers formed arithmetic progressions. It is known that the asymptotic density of sets of terms of arithmetic progressions are greater than zero, more exactly it equals the reciprocal of the difference. This means that sequence (2.1) is such a subsequence of (1.1) which contains relatively "many" terms of sequence (1.1). In what follows we are looking for subsequences of (1.1) where the density of the set of powers is zero, but they have infinitely many prime divisors. We give two sequences possessing the above conditions. In one of them the powers run through the set of primes and in the other the powers equal the values of Euler's function φ. It is known fact that the asymptotic density of the set of primes and the set of values of Euler's function are zero.

Theorem 2.3. *Let a, b be natural numbers with $(a, b) = 1$ and $a > 1$. Let us denote by p_n the n-th prime number. Then the sequence*

$$(a^{p_n} + b)_{n=1}^{\infty} \tag{2.10}$$

has infinitely many prime divisors.

Proof. Let us suppose that sequence (2.10) has only finitely many prime divisors, namely $q_1, q_2, \ldots, q_k$. We discuss two cases.

We consider first that there are prime divisors of the terms of sequence (2.10) which divide $a + b$. Let us denote by $q_1 < \cdots < q_l$ the divisors of $a + b$ and $q_{l+1} < \cdots < q_k$ which are not divisors of $a + b$. Let us denote by α_s for all $1 \leqslant s \leqslant l$ the least natural number which

$$q_s^{\alpha_s} > a + b\,.$$

Put

$$M = q_1^{\alpha_1} q_2^{\alpha_2} \cdots q_l^{\alpha_l} q_{l+1} \cdots q_k.$$

In this case $(a, M) = 1$ since $(a, b) = 1$. It follows from Euler's theorem that

$$a^{n\varphi(M)} - 1 \equiv 0 \pmod{M} \tag{2.11}$$

for all $n \in \mathbb{N}$. Using the theorem of Dirichlet we get that there are infinitely many prime numbers in the sequence $(n\varphi(M)+1)_{n=1}^{\infty}$. Let us denote these prime numbers by $p_1' < p_2' < \cdots p_n' < \cdots$. Obviously the sequence

$$(a^{p_n'} + b)_{n=1}^{\infty} \tag{2.12}$$

is a subsequence of sequence of (2.10). Let q be a prime divisor of sequence of (2.12). Obviously $q \in \{q_1, q_2, \ldots, q_k\}$, moreover

$$a^{p_i'} + b \equiv 0 \pmod{q} \tag{2.13}$$

for some $i \in \mathbb{N}$. It follows from (2.11) and (2.13) that

$$0 \equiv a^{p_i'} + b \equiv a^{p_i'-1}(a-1) + a^{p_i'-1} + b \equiv a + b \pmod{q}.$$

Thus $q \mid a+b$ and $q \in \{q_1, q_2, \ldots, q_l\}$. In other words $a^{p_i'} + b$ can be written in the form

$$a^{p_i'} + b = q_1^{\beta_{i,1}} q_2^{\beta_{i,2}} \cdots q_l^{\beta_{i,l}}$$

where $\beta_{i,j} \geqslant 0$ for all $1 \leqslant j \leqslant l$ natural numbers.

Now we show that $\beta_{i,j} < \alpha_j$ for all $1 \leqslant j \leqslant l$. If $\beta_{i,j} \geqslant \alpha_j$ for some $1 \leqslant j \leqslant l$ then

$$a^{p_i'} + b \equiv 0 \pmod{q_j^{\alpha_j}}$$

moreover using (2.11) and $q_j^{\alpha_j} \mid M$ we have

$$a^{p_i'-1} \equiv 1 \pmod{q_j^{\alpha_j}}.$$

It follows from the previous congruence that

$$0 \equiv a^{p_i'} + b \equiv a^{p_i'-1}(a-1) + a^{p_i'-1} + b \equiv a + b \pmod{q_j^{\alpha_j}}$$

which contradicts the fact that $q_j^{\alpha_j} > a + b$. In this way we get

$$a^{p_i'} + b < q_1^{\alpha_1} q_2^{\alpha_2} \cdots q_l^{\alpha_l} \leqslant M$$

for all $i \in \mathbb{N}$. Here we have obtained a contradiction since sequence (2.12) is not bounded.

In the second case we study when the terms of sequence (2.10) do not have such prime divisors which divide $a + b$. Put

$$L = q_1 q_2 \cdots q_k.$$

Since $(a, L) = 1$, therefore

$$a^{n\varphi(L)} - 1 \equiv 0 \pmod{L} \tag{2.14}$$

for all $n \in \mathbb{N}$. Let

$$Q = l\varphi(L) + 1$$

be a prime and q be a prime divisor of $a^Q + b$.

It follows from the definition of Q and from (2.14) that

$$a^{Q-1} \equiv 1 \pmod{q}$$

where $q \in \{q_1, q_2, \ldots, q_k\}$. Obviously

$$0 \equiv a^Q + b \equiv a^{Q-1}(a-1) + a^{Q-1} + b \equiv a + b \pmod{q},$$

which contradicts the fact that q is not a divisor of $a + b$. □

It is worth investigating that if a term of sequence (2.10) is divisible by a prime then this prime is a divisor of infinitely many terms of the sequence. The answer is not as obvious as before. First of all we prove a Lemma which help us in this case and other similar cases, too.

Lemma 2.4. *Let a, b be natural numbers with $(a, b) = 1$ and $a > 1$. If q is a prime divisor of sequence* (1.1) *then*

1. *There exists an exponent h_q of a with respect to q.*
2. *If q is a divisor of $a^k + b$ then q is a divisor of those terms of sequence* (1.1) *which can be given of the form*

$$a^{k+zh_p} + b$$

where $z \in \mathbb{Z}$ and $k + zh_p \geqslant 0$.

Proof. 1. The first statement is trivial. If $(a, b) = 1$ and q is a divisor of a term of sequence (1.1) then $(a, q) = 1$.

2. Let q is a prime divisor of $a^k + b$. Let us denote by h_q an exponent of a with respect to q. Let us consider a term in the form $a^m + b$ of sequence (1.1). In this case q is a divisor of $a^m + b$ if and only if

$$(a^k + b) - (a^m + b) \equiv 0 \pmod{q}. \tag{2.15}$$

Using elementary conversions we have

$$(a^k + b) - (a^m + b) = a^{\min\{k,m\}}\big(a^{|m-k|} - 1\big).$$

Since $(a, q) = 1$ and h_q is an exponent of a we get that congruence (2.15) is valid if and only if h_q is a divisor of $|m-k|$. This statement is equivalent to our statement. □

Conclusion 2.5. If a prime q is a divisor of two different terms of sequence (2.10) then it is a divisor of infinitely many terms of the sequence.

Proof. Let q be a prime divisor of at least two different terms of sequence (2.10). Let us denote these terms by $a^{p_1}+b$ and $a^{p_2}+b$ where $p_1 < p_2$. It follows from Lemma 1 that

$$p_2 = p_1 + nh_q$$

for some natural number n. Since p_1 and p_2 are primes therefore $(p_1, h_q) = 1$. Using Dirichlet's theorem we have that there is a subsequence $(p'_n)_{n=1}^{\infty}$ with prime terms of the sequence $(p_1 + nh_q)_{n=1}^{\infty}$. It follows from Lemma 1 that q is a divisor of all terms of the sequence

$$(a^{p'_n} + b)_{n=1}^{\infty}.$$

□

Further we study a subsequence of (1.1) where the powers are the values of Euler's function φ. Similarly to the previous sequence the asymptotic density of the set of values of Euler's function φ equals zero. First we prove that there are infinitely many prime divisors of this sequence.

Theorem 2.6. *Let a, b be natural numbers where $(a, b) = 1$ and $a > 1$. Then there are infinitely prime divisors of the sequence*

$$(a^{\varphi(n)} + b)_{n=1}^{\infty}. \tag{2.16}$$

Proof. Let us suppose that there are only finitely many prime divisors of sequence (2.16) namely $q_1, q_2, \ldots, q_k$. We distinguish two cases.

In the first case we suppose that among the prime divisors of sequence (2.16) there are divisors which divide $b+1$. Let us denote these divisors by $q_1 < \cdots < q_l$ and the others by $q_{l+1} < \cdots < q_k$.

Let us denote by α_s for all s $(1 \leqslant s \leqslant l)$ the least natural number which

$$q_s^{\alpha_s} > b + 1\,.$$

Put

$$M = q_1^{\alpha_1} q_2^{\alpha_2} \cdots q_l^{\alpha_l} q_{l+1} \cdots q_k.$$

Obviously $(a, M) = 1$. It follows from Euler's theorem that

$$a^{n\varphi(M)} \equiv 1 \pmod{M} \tag{2.17}$$

for all natural numbers n. Let us consider an increasing sequence of prime numbers $(p_i)_{i=1}^{\infty}$ where $(p_i, M) = 1$ for all $i \in \mathbb{N}$. Since the Euler's function is multiplicative we have that the sequence

$$(a^{\varphi(p_i)\varphi(M)} + b)_{i=1}^{\infty} \tag{2.18}$$

is a subsequence of sequence (2.16).

It is obvious that the prime divisors of sequence (2.18) belong to the set

$$\{q_1, q_2, \ldots, q_k\}.$$

We choose one of them and let us denote it by q. It follows from (2.17) that

$$0 \equiv a^{\varphi(p_i)\varphi(M)} + b \equiv b + 1 \pmod{q},$$

that is q is a divisor of $b+1$ and $q \in \{q_1, q_2, \ldots, q_l\}$. Thus we have

$$a^{\varphi(p_i)\varphi(M)} + b = q_1^{\beta_{i,1}} q_2^{\beta_{i,2}} \cdots q_l^{\beta_{i,l}}$$

where $\beta_{i,j} \geqslant 0$ for all $1 \leqslant j \leqslant l$ and $i \in \mathbb{N}$.

Henceforth we show that $\beta_{i,j} < \alpha_j$ for all $1 \leqslant j \leqslant l$ and $i \in \mathbb{N}$. If $\beta_{i,j} \geqslant \alpha_j$ for any $1 \leqslant j \leqslant l$ and for $i \in \mathbb{N}$, then we have

$$a^{\varphi(p_i)\varphi(M)} + b \equiv 0 \pmod{q_j^{\alpha_j}} \quad \text{and} \quad a^{\varphi(p_i)\varphi(M)} - 1 \equiv 0 \pmod{q_j^{\alpha_j}}.$$

It follows from the previous congruences that $q_j^{\alpha_j}$ is a divisor of $b+1$, this contradicts the fact that $q_j^{\alpha_j} > b+1$. Hence

$$a^{\varphi(p_i)\varphi(M)} + b < q_1^{\alpha_{i,1}} q_2^{\alpha_{i,2}} \cdots q_l^{\alpha_{i,l}} \leqslant M\,,$$

which is a contradiction since sequence (2.18) is not bounded.

In the second case we suppose that the divisors of sequence (2.16) are not divisors of $b+1$. Put

$$L = q_1 q_2 \cdots q_k.$$

Since $(a, L) = 1$, it follows from the Euler's theorem that

$$a^{\varphi(L)} - 1 \equiv 0 \pmod{L}. \tag{2.19}$$

Obviously $a^{\varphi(L)} + b$ is a term of sequence (2.16). Let q be a prime divisor of sequence (2.16). In this case

$$0 \equiv a^{\varphi(L)} + b \equiv a^{\varphi(L)} - 1 + b + 1 \equiv b + 1 \pmod{q},$$

that is q is a divisor of $b+1$ which is contradiction. □

Further we investigate when a prime divisor of sequence (2.16) divides infinitely many terms of sequence (2.16). This problem is more difficult than in case (2.10). We give two sufficient conditions.

Theorem 2.7. *If q is a prime divisor of sequence* (2.16) *and $b + 1 \equiv 0 \pmod{q}$, then q is a divisor of infinitely many terms of sequence* (2.16).

Proof. Let q be an odd prime divisor of sequence (2.16) with the condition $b+1 \equiv 0 \pmod{q}$. Since $(a, q) = 1$, it follows from the Euler's theorem that $a^{\varphi(q)} \equiv 1 \pmod{q}$. Obviously we have

$$a^{\varphi(q)} + b \equiv a^{\varphi(q)} - 1 + b + 1 \equiv 0 \pmod{q}.$$

Let $(p_n)_{n=1}^{\infty}$ be an arbitrary increasing sequence of prime numbers, where q is not a term of this sequence.

We show that q is a divisor of all terms of the sequence

$$(a^{\varphi(qp_n)}+b)_{n=1}^{\infty}.$$

Since φ is a multiplicative function and $(q,p_n)=1$ we have that

$$a^{\varphi(qp_n)}+b \equiv a^{\varphi(q)\varphi(p_n)}+b \equiv (a^{\varphi(q)})^{\varphi(p_n)}-1+b+1 \equiv 0 \pmod{q}$$

for all natural numbers n. □

Theorem 2.8. *Let q be a prime divisor of sequence* (2.16) *and the power of a is an odd number* $\pmod{q}$. *Then q is a divisor of infinitely many terms of sequence* (2.16).

Proof. Let q be such a prime divisor of sequence (2.16) that the power of a is odd $\pmod{q}$. Let us denote by n_0 the least natural number where q is a divisor of $a^{\varphi(n_0)}+b$. Since the power h_q of a is odd $\pmod{q}$ from the Dirichlet's theorem follows that the sequence

$$(kh_q+2)_{k=1}^{\infty} \tag{2.20}$$

contains infinitely many prime numbers. Let us choose a subsequence

$$(p_n')_{n=1}^{\infty}$$

of sequence (2.20) which terms are primes and not divisors of the number n_0. Since φ multiplicative we have that

$$a^{\varphi(n_0p_n')}+b = a^{\varphi(n_0)\varphi(p_n')}+b = a^{\varphi(n_0)(p_n'-1)}+b =$$

$$= a^{\varphi(n_0)(kh_q+1)}+b = a^{\varphi(n_0)+\varphi(n_0)kh_q}+b$$

for all $n \in \mathbb{N}$. Using Lemma 1 we have that q is a divisor of all terms of the sequence

$$(a^{\varphi(n_0p_n')}+b)_{n=1}^{\infty}.$$

□

Finally we show that there are infinitely many primes which do not divide any term of sequence (2.16). First we prove a more general theorem.

Theorem 2.9. *Let $a>1$ and $b>1$ be natural numbers where b is odd and $(a,b)=1$. Then there are infinitely many primes p which do not divide any term of sequence*

$$(a^{2n}+b)_{n=1}^{\infty} \tag{2.21}$$

Proof. Let p be an arbitrary prime. Then p is not a divisor of any term of sequence (2.21) if and only if there is no solution of the quadratic congruence $x^2 \equiv -b \pmod{p}$. Using the Jacobi's symbol we have

$$\left(\frac{-b}{p}\right) = -1.$$

Let p be an odd prime number where $(b, p) = 1$. Applying the law of quadratic reciprocity of Gauss we have

$$\left(\frac{-b}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{p}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{p}{b}\right) \ (-1)^{\frac{p-1}{2}\frac{b-1}{2}} = \left(\frac{p}{b}\right)(-1)^{\frac{p-1}{2}\frac{b+1}{2}}. \qquad (2.22)$$

We distinguish two cases.

First we suppose that $b = 4l + 1$ where l is a natural number. Let us consider primes of the form

$$p = 4bk + 2b + 1, \quad \text{where} \quad k \in \mathbb{N}.$$

It follows from the Dirichlet's theorem that there are infinitely many primes of the form as above since $(4b, 2b + 1) = 1$.

In this case $\left(\frac{p}{b}\right) = \left(\frac{1}{b}\right) = 1$ and $\frac{p-1}{2}\frac{b+1}{2}$ is odd natural number. Using (2.22) we have

$$\left(\frac{-b}{p}\right) = -1.$$

That is p doesn't divide any term of sequence (2.21).

In the second case we suppose that $b = 4l + 3$ where l natural number.
Let us consider primes of the form

$$p = 2bk + 2b - 1.$$

Using the previous method we get that there are infinitely many primes of this form. Obviously $\frac{b+1}{2}$ is even. Moreover

$$\left(\frac{p}{b}\right) = \left(\frac{-1}{b}\right) = (-1)^{\frac{b-1}{2}} = -1.$$

Using (2.22) we have equation

$$\left(\frac{-b}{p}\right) = -1.$$

That is p doesn't divide any term of sequence a (2.21). □

Conclusion 2.10. There are infinitely many primes which do not divide any term of sequence (2.16).

Proof. Using the previous theorem we get this statement since the Euler's function φ is even except those cases when $\varphi(1) = \varphi(2) = 1$. □

References

[1] Hardy, G. H., Wright, E. M., An introduction to the theory of numbers, Oxford, 1954.

[2] Sárközi, A., Számelmélet, Műszaki Könyvkiadó, Budapest, 1976.

[3] Sárközi, A., Surányi, J., Számelmélet feladatgyűjtemény, 13. kiadás, Tankönyvkiadó, Budapest, 1990.

[4] Sierpinsky, W., Elementary theory of numbers, PWN, Warszawa, 1964.

[5] Sierpinsky, W., 200 feladat az elemi számelméletből, Tankönyvkiadó, Budapest, 1964.

[6] Tóth, J., Egy számsorozat prímosztóiról, Polygon, Szeged III (2) (1993), 78–79.

Ferdinánd Filip
Department of Mathematics
University of J. Selye
SK-94501 Komárno
Rožníckej Školy 1514
Slovakia

Kálmán Liptai
Institute of Mathematics and Informatics
Eszterházy Károly College
H-3300 Eger
Leányka út 4.
Hungary

János T. Tóth
Department of Mathematics
University of Ostrava
CZ-701 03 Ostrava
30. dubna 22
Czech Republic

Annales Mathematicae et Informaticae
***33** (2006) pp. 57–68*
http://www.ektf.hu/tanszek/matematika/ami

Solving ordinary differential equation systems by approximation in a graphical way

Gábor Geda[a], Anikó Vágner[b]

[a]Department of Computer Science
Eszterházy Károly College
e-mail: gedag@aries.ektf.hu

[b]Department of Information Technology
Eszterházy Károly College
e-mail: vagnera@aries.ektf.hu

Submitted 11 November 2006; Accepted 18 December 2006

Abstract

Our aim was to find a graphic numeric solution method for higher-order differential equations and differential equation systems. To understand this method the basic mathematical knowledge taught in the secondary school must be enough, we have to complete it with geometric meaning of differential quotient and generalization of knowledge about two-dimensional vector space. We considered it important to make this method easy to algorithm. Such method and its practical experience are shown in this paper.

MSC: 65L05, 65L06, 53A04,97D99

1. Introduction

It is well-known how important the differential equation models are in the mathematical description of different processes and systems.

Our aim is to find approximate methods which are based on the demonstration and there is no need for higher mathematical knowledge to understand and apply them. Moreover, it is easy to algorithmise them even in the possession of the secondary school material.

The problem concerning this topic such as mechanical oscillations can be given as an ordinary n-order differential equation:

$$y^{(n)}(t) = g(t, y(t), y^{(1)}(t), y^{(2)}(t), \ldots, y^{(n-1)}(t))$$

This can be transformed to explicit ordinary differential equation system (abbreviated as ODES in the followings):

$$\dot{x}_i(t) = f_i(t, x_1(t), x_2(t), \ldots, x_n(t)) \quad (i = 1, \ldots, n). \tag{1.1}$$

The solution for these equations, if it exist at all, can be given with $x_i(t)$ $(i = 1, \ldots, n)$ functions. In most cases to produce such solutions is a difficult task which needs the knowledge of serious mathematical devices.

We consider the solutions $n + 1$-dimensional space curve

$$\mathbf{x}(t) = (t, x_1(t), x_2(t), \ldots, x_n(t)).$$

So (1.1) corresponds a vector to any point in $n + 1$-dimensional vector place, and the vector is parallel with the tangent line at a given P point of the solution of ODES. The only problem is that we do not know which points should be considered belonging to the same curve among the points close to one another.

In certain cases there is no need to present all the possible solutions, that is all curves, only the $\mathbf{x}(t)$ curve is necessary of which a given

$$P_0(p_0; p_1; p_2; \ldots; p_n)$$

fits, and on the coordinates of which

$$x_i(p_0) = p_i \quad (i = 1, \ldots, n)$$

is realized. In this case we can say that we solve an initial value problem. By expressing an initial value problem we choose one of the curves which are solutions for ODES. Other times we have to be contented with the approximate solution of the problem.

In a geometrical point of view the solution for an initial value problem by approximation is giving a $P_0, P_1, \ldots, P_k$ point serial the elements of which fit to the chosen curve by desired accuracy. The serial of points $(0 \leqslant i \leqslant k)$ determines a broken line the points of which approximate well the points of the curve.

The accuracy of the approximation is influenced by several factors. The most important ones among them are the approximate algorithm and ODES itself.

This way, when we select the successive elements of the point serial we should take the changes of the curve of the function into consideration.

2. Demonstration of an approximate method

Let (1.1),

$$P_0(p_0; p_1; p_2; \ldots; p_n)$$

point on coordinates of which

$$x_i(p_0) = p_i \quad (i = 1, \ldots, n)$$

is realized and a suitable minor d distance. We would like to determine the broken line running through P_0 point and approaching the

$$\mathbf{x}(t) = (t, x_1(t), x_2(t), \ldots, x_n(t))$$

function curve meaning the solution in the surroundings of P_0 given point.

Let $\mathbf{m^p}$ vector be parallel with tangent line to curves at P_0 point. The coordinates of $\mathbf{m^p}$ are:

$$m_0^p = 1$$

$$m_i^p = \dot{x}_i(p_0) \quad (i = 1, \ldots, n).$$

Define $\mathbf{p}$ vector, where $\mathbf{p}$ is parallel with $\mathbf{m^p}$ vector and $\|\mathbf{p}\|$=d, that is

$$\mathbf{p} = \frac{\mathbf{m^p}}{\|\mathbf{m^p}\|} d. \tag{2.1}$$

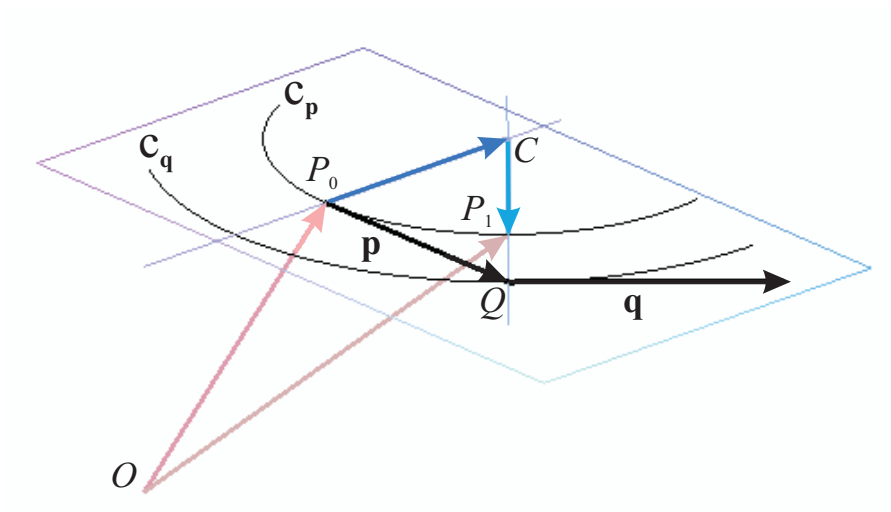

Figure 1: The c_p and the c_q osculation circles in case $n = 2$ (in 3 dimension)

Define Q point, where $\overrightarrow{P_0Q} = \mathbf{p}$. Then coordinates of Q point can be calculated (see figure 1). Let coordinates of Q point be $Q(q_0; q_1; \ldots; q_n)$.

Coordinates of $\mathbf{m^q}$ vector, which is parallel with tangent line to curves at Q point is:

$$m_0^q = 1$$

$$m_i^q = \dot{x}_i(q_0) \quad (i = 1, \ldots, n).$$

If d is minor enough, then Q is close enough to the curve which is the solution for the initial value problem. This way, $\mathbf{m^q}$ well approximates the steepness of the curve in one of its points near to Q.

Define $\mathbf{q}$ vector, where $\mathbf{q}$ is parallel $\mathbf{m^q}$ vector and $\|\mathbf{q}\| = d$, in other words

$$\mathbf{q} = \frac{\mathbf{m^q}}{\|\mathbf{m^q}\|} d. \tag{2.2}$$

If $\mathbf{q}$ vector is parallel with $\mathbf{p}$ vector, then we accept Q point as the next element of serial of points, and we continue the approaching from this point.

Otherwise in the narrow surroundings of P_0 the curve can be well approximated in the plane, which $\mathbf{p}$ and $\mathbf{q}$ vectors define with a proper arc (c_p), which is the osculating circle of the curve in P_0. Similarly, we can fit an arch (c_q) in $(\mathbf{p}, \mathbf{q})$ plane in the narrow surroundings of Q to the curve on which Q fits (see figure 2.a). The lines which are perpendicular tangent lines in P_0 and Q points intersect at point C. This point can be considered to be the common central point of the two circles (c_p and c_q) if the d is minor enough.

Define $\mathbf{a}$ and $\mathbf{b}$ vectors for coordinates of C point: $\mathbf{a} = \mathbf{p} + \lambda\mathbf{q}$, and let $\mathbf{a}$ be perpendicular to $\mathbf{p}$ vector, and $\mathbf{b} = \mathbf{q} + \omega\mathbf{p}$ and let $\mathbf{b}$ be perpendicular to $\mathbf{q}$ vector (see figure 2.b; 2.c).

As $\mathbf{p}$ and $\mathbf{a}$ are perpendicular to each other, their scalar product is null, from which λ can be calculated:

$$\lambda = -\frac{\sum_{i=0}^{n} p_i^2}{\sum_{i=0}^{n} p_i q_i}.$$

Similar way, can we get value of ω from scalar product of $\mathbf{q}$ and $\mathbf{b}$:

$$\omega = -\frac{\sum_{i=0}^{n} q_i^2}{\sum_{i=0}^{n} p_i q_i}.$$

On the one hand, $\overrightarrow{OC}$ local vector can be written with $\overrightarrow{OP_0}$ local vector and $\mathbf{a}$ vector multiplies by a constant, on the other hand, with $\overrightarrow{OQ}$ local vector and $\mathbf{b}$ vector multiplied by an other constant. That is:

$$\overrightarrow{OC} = \overrightarrow{OP_0} + \phi\mathbf{a} = \overrightarrow{OP_0} + \phi\mathbf{p} + \phi\lambda\mathbf{q}, \tag{2.3}$$

$$\overrightarrow{OC} = \overrightarrow{OQ} + \psi\mathbf{b} = \overrightarrow{OQ} + \psi\mathbf{q} + \psi\omega\mathbf{p}. \tag{2.4}$$

We know, that

$$\overrightarrow{OQ} - \overrightarrow{OP_0} = \mathbf{p}.$$

Then in the $(\mathbf{p}, \mathbf{q})$ base are the value of ϕ and ψ can be calculated:

$$\phi = \frac{1}{\lambda\omega - 1}; \quad \psi = \frac{\lambda}{\lambda\omega - 1}. \tag{2.5}$$

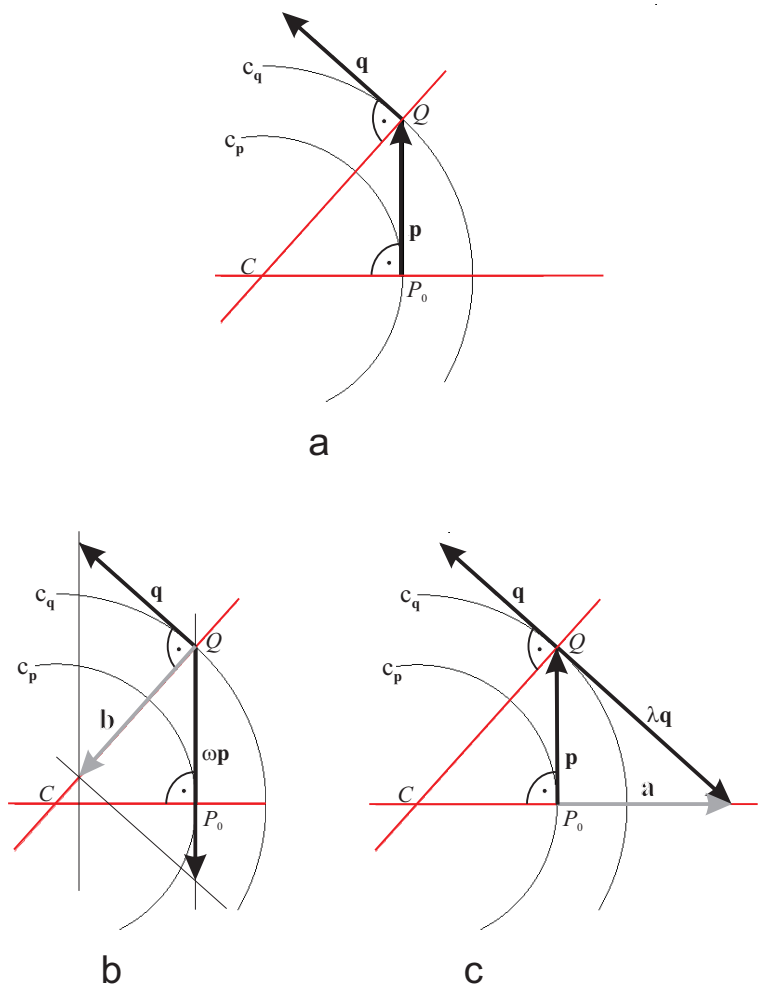

Figure 2: The c_p and c_q osculation circles in the plane of $\mathbf{p}$ and $\mathbf{q}$

Then coordinates of C point can also be calculated in both $(\mathbf{p}, \mathbf{q})$ base and $n+1$-dimensional vector place.

Knowing coordinates of C point we can define P_1 point, as a point of the line defined by C and Q points and of c_p arc, namely $\overrightarrow{CP_1}$ vector is parallel with $\overrightarrow{CQ}$ vector, $\left\|\overrightarrow{CP_1}\right\| = \left\|\overrightarrow{CP}\right\|$ and P_1 point is on the $\overline{CQ}$ half-line. So

$$\overrightarrow{OP_1} = \overrightarrow{OP_0} + \overrightarrow{P_0C} + \overrightarrow{CP_1},$$

where $\overrightarrow{OP_0}$ and $\overrightarrow{OP_1}$ are local vectors (see figure 1).

To determine the following approximate point, the starting point will be P_1 as it was P_0 earlier.

The promptness of the approximation depends on the selection of the value d. If it is too big, C will not be a good approximation of the common centre of the two osculating circles (c_p and c_q).

At the same time, if we find an appropriate C point then the distance of C and P_0 approximate the radius of the circle of curvature at P_1. This can be used to get a better defining of the value of d. If we can choose the value of d according to the characteristics of the curve we can approximate the function more precisely, and the algorithm will be faster.

To understand the operation of this method we only need the knowledge of graphic meaning of the differential quotient as the exact definition is not used in this case.

If we regard an ODES as a function which orders vector to the point of $n+1$-dimensional place, where the vector is parallel with the tangent line at the point then the point serial giving the solution can be written by the use of vector operation based on the method mentioned above (in the followings OCM–osculating circle method) which approximates the solution of initial value problem.

To give the algorithm we need the knowledge of the equation system and vector operations (such as scalar product, vector addition).

3. Look at the problem in case n = 2

Let the next initial value problem be given

$$\dot{x}_1(t) = f_1(t, x_1(t), x_2(t)),$$

$$\dot{x}_2(t) = f_i(t, x_1(t), x_2(t)),$$

$$P_0(p_0; x_1(p_0); x_2(p_0)).$$

The solution of initial value problem is the $\mathbf{x}(t) = (t, x_1(t), x_2(t))$ curve, which can be approximated with $P_0, P_1, \ldots, P_k$ serial of points. The $\mathbf{m^p}(1, \dot{x}_1(p_0), \dot{x}_2(p_0))$ vector is parallel with tangent line of the curve at P_0 point. Then coordinates of $\mathbf{p}$ vector can be calculated on grounds (2.1).

If coordinates of Q point are $Q(q_0; q_1; \ldots; q_n)$ then $\mathbf{m^q}(1, \dot{x}_1(q_0), \dot{x}_2(q_0))$ is the vector belonging to Q. From $\mathbf{m^q}$ coordinates of $\mathbf{q}$ vector are calculatable ground of (2.2). Value of λ and ω can be defined:

$$\lambda = -\frac{p_0^2 + p_1^2 + p_2^2}{p_1 q_1 + p_1 q_1 + p_2 q_2},$$

$$\omega = -\frac{q_0^2 + q_1^2 + q_2^2}{p_1 q_1 + p_1 q_1 + p_2 q_2}.$$

Ground of these considering (2.3), (2.4) and (2.5) coordinates of C point can be calculatable, from which coordinates of P_1 are also calculatable grounding of $\overrightarrow{CP_0}$, $\overrightarrow{CQ}$ vectors and

$$\overrightarrow{OP_1} = \overrightarrow{OP_0} + \overrightarrow{P_0C} + \overrightarrow{CP_1}$$

vector.

4. Examples

To illustrate the usefulness of algorithm we show two examples. Data for the figures of the examples were provided by a program, which was made on the base of above demonstrated algorithm. The initial values and parameters can be chosen randomly, the values in the examples provide the demonstration of working of algorithm. The aim was not to show the mathematical model.

4.1. Equation of damped oscillation

Generally:

$$x^{(2)}(t) + \frac{c}{m}x^{(1)}(t) + \omega^2 x(t) = 0,$$

where c, m and ω are constants characteristic of the system. We get the next equation system after transforming:

$$\begin{aligned} \dot{x}_1(t) &= x_2(t), \\ \dot{x}_2(t) &= -\frac{c}{m}x_2(t) - \omega^2 x_1(t). \end{aligned}$$

Choose $c = 1$, $m = 2$ and $\omega = \pi$ in the example. Then

$$\begin{aligned} \dot{x}_1(t) &= x_2(t), \\ \dot{x}_2(t) &= -\frac{1}{2}x_2(t) - \pi^2 x_1(t). \end{aligned} \tag{4.1}$$

Define the approximated solution of equation system where the initial conditions are

$$\begin{aligned} x_1(0) &= 0.28 \\ x_2(0) &= 0.28 \end{aligned}$$

values by using OCM algorithm.

Compare our solution to numeric solution produced by Runge-Kutta4 method of Maple program. In both methods we have approximated the solution ($h = 350$).

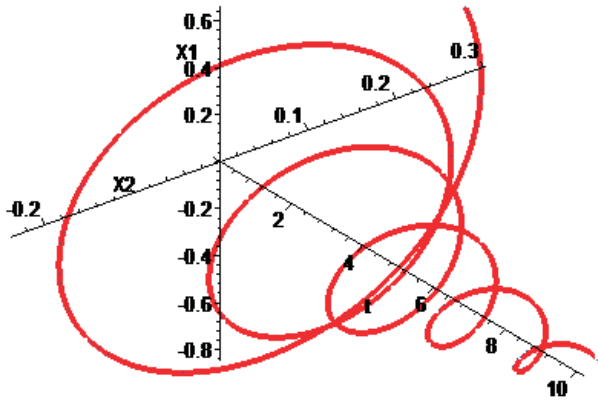

Figure 3: Curve of meaning the solution of (4.1) equation system.

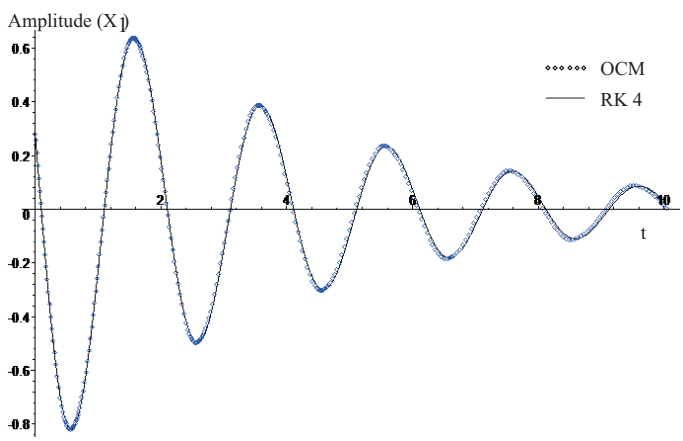

Figure 4: Deflection-time function.

4.2. Lotka-Volterra model

Lotka-Volterra equations are suitable for modelling various occurrences, systems for examples ecological systems, chemical processes. The model can be defined with next equations:

$$\dot{x}_1(t) = -ax_1(t) + bx_1(t)x_2(t) - mx_1^2(t),$$
$$\dot{x}_2(t) = cx_1(t) - dx_1(t)x_2(t) - lx_2^2(t).$$

The actual entity number of predatory is X_1, prey is X_2. a, b, c, d, m, l are constant characteristic of the system. In our example we examine the system $a = 2$; $b = 0.015$; $c = 1$; $d = 0.03$; $m = 0$; $l = 0.0005$:

$$\begin{aligned} \dot{x}_1(t) &= -2x_1(t) + 0.015x_1(t)x_2(t), \\ \dot{x}_2(t) &= x_1(t) - 0.03x_1(t)x_2(t) - 0.0005x_2^2(t). \end{aligned} \tag{4.2}$$

Initial condition is:

$$\begin{aligned} x_1(0) &= 150, \\ x_2(0) &= 50. \end{aligned}$$

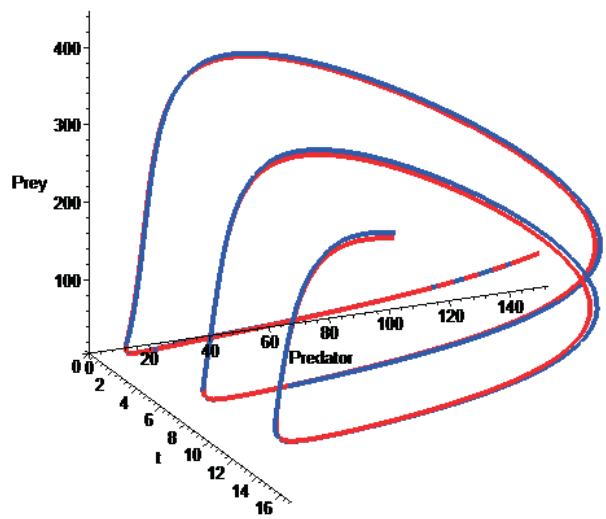

Figure 5: Curve meaning the solution of (4.2) equation system.

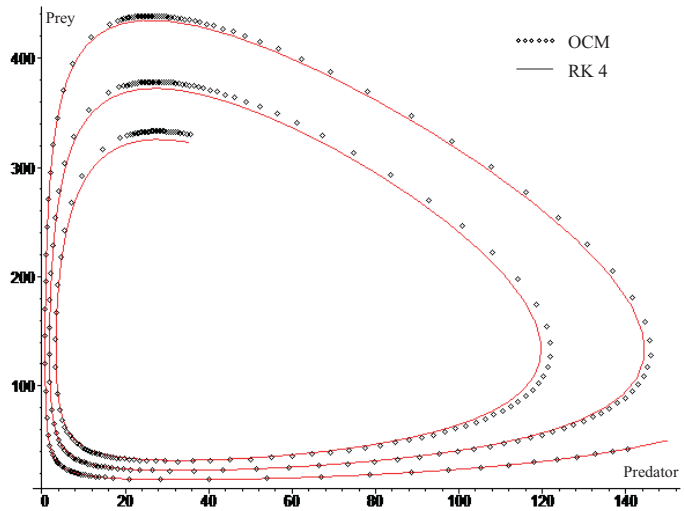

Figure 6: Trajectory of (4.2).

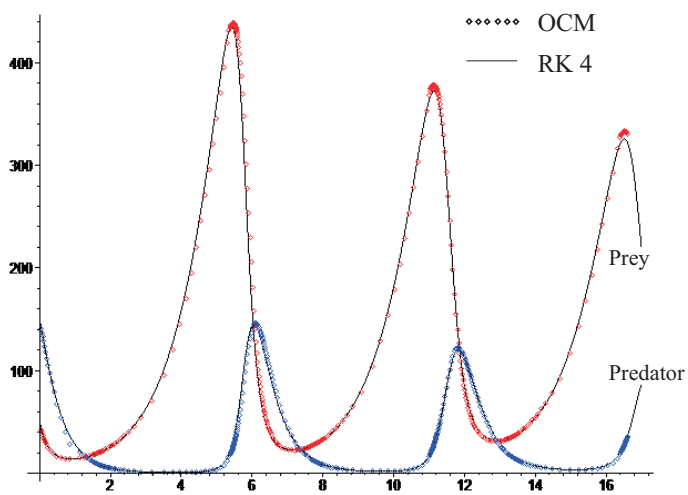

Figure 7: Change of entity number of predators and preys in time.

5. Conclusion

In case of various approximating method the solution can be approximated more precisely by increasing of the applied basis points. But this makes the method more difficult and needs more counting. Despite OCM, considers two basis points, the method provides comparatively great precision. It can be explained by approximating the curve on short periods with arcs. The curve piece, which is between any definite two basis points, can be approximated with an arc with suitable radius, in other words, any point on the curve piece between P_i, P_{i+1} points can be approximated with a suitable point of arch. Giving the solution by vector equation, despite we give in $n+1$ dimensional vector space, can be easy. As we reduce the solution to a 2-dimensional case, we can avoid the solution of equation system, which has more equation than two.

References

[1] Geda, G.,Solving initial value problem by different numerical methods: Practical investigation, *Annales Mathematicae et Informaticae* (2005).

[2] Geda, G., Solving Initial Value Problem by Approximation in Different Graphic Ways, 5^{th} *Intrnational Conference of PhD Students, Miskolc* (2005).

[3] Arató, M. A Famous Nonlinear Stocshastic Equation (Lotka-Volterra Model with Diffusion), *Mathematical and Computer Modelling*, (2003).

[4] Geda, G. Investigation of Stability of Nonlinear Differential Equations with Stochastic Methods, *XXVI. Seminar on Stability Problems for Stochastic Models*, (2005).

[5] Stoyan, G., Tako, G., Numerikus módszerek, *ELTE-TypoTEX, Budapest*, (1993).

[6] Hatvani, L., Pintér, L., Differenciálegyenletes modellek a középiskolában, *POLYGON, Szeged*, (1997).

[7] Rontó, M., Raisz, Péterné., Differenciálegyenletek műszakiaknak, *Miskolci Egyetemi Kiadó, Miskolc*, (2004).

[8] Szőkefalvi-Nagy Gy., Gehér L., Nagy P., Differenciálgeometria, *Műszaki Könyvkiadó, Budapest,*, (1979).

[9] Geda, G., Kezdetiérték-probléma közelítő megoldásának egy geometriai szemléltetése, *Tavaszi Szél, Debrecen*, (2005).

[10] Pólya, Gy., Matematikai módszerek a természettudományban, *Gondolat, Budapest*, (1984).

Gábor Geda
Department of Computer Science
Eszterházy Károly College
Leányka str. 6.
H-3300 Eger, Hungary

Anikó Vágner
Department of Information Technology
Eszterházy Károly College
Leányka str. 6.
H-3300 Eger, Hungary

Annales Mathematicae et Informaticae
***33** (2006) pp. 69–75*
http://www.ektf.hu/tanszek/matematika/ami

Ljunggren's Diophantine problem connected with virus structure

Aleksander Grytczuk

Faculty of Mathematics, Computer Science and Econometrics
University of Zielona Góra
e.mail: A.Grytczuk@wmie.uz.zgora.pl

Submitted 5 March 2006; Accepted 22 December 2006

Abstract

In this paper we give an effective method for determination of all solutions of the Ljunggren's Diophantine equation

$$x^2 + 3y^2 + 12z = 4M, \tag{L}$$

in odd positive integers x, y and non-negative integers z, where $M = a^2 + ab + b^2$, $N = 10M + 2$ and a, b are given non-negative integers. Equation (L) is strictly connected with virus structure.

1. Introduction

In virology are known (see [3, pp. 171–200]) different groups of viruses. One of such groups has been found by Stoltz [5], [6] and by Wrigley [7], [8] and is called as symmetrons. Virus particles are invariably enclosed by shells of protein subunits and these are packed geometrically according to symmetry rules. More of known examples are close packed with each subunit surrounded by six neighbours, except the twelve vertices which have five neighbours. In the paper [1], Goldberg indicated that total number of nearly identical subunits which may be regularly packed on the closed icosahedral surface is given by the following formula:

$$N = 10\left(a^2 + ab + b^2\right) + 2, \tag{G}$$

where a, b are given non-negative integers.

Stoltz ([5], [6]) and Wrigley ([7], [8]) discovered that the symmetrons have the construction of linear, triangular and pentagonal and are called: disymmetrons,

trisymmetrons and pentasymmetrons, respectively. Moreover, it is known [7], that an icosahedron has 30 axes of twofold symmetry, 20 of threefold symmetry and 12 of fivefold symmetry. Hence, the subunits on the surface of an icosahedral virus may be divided into 30, 20 or 12 corresponding previously listed groups symmetry. Let the 30 disymmetrons contain d_u subunits, the 20 trisymmetrons contain t_v subunits and the 12 pentasymmetrons contain p_w subunits, then we have

$$N = 30d_u + 20t_v + 12p_w, \tag{S-W}$$

where

$$d_u = u - 1, \quad t_v = \frac{(v-1)\,v}{2}, \quad p_w = \frac{5w\,(w-1)}{2} + 1 \tag{1.1}$$

and u, v, w are positive integers.

Now, we remark that for each value of N given by the equation (G) the number $f(N)$ of the solutions of the equation (S-W) corresponds to the number theoretically possible ways of making a virus with N subunits, but with different combinations of symmetrons.

Putting

$$x = 2v - 1, \quad y = 2w - 1, \quad z = u - 1, \quad N = 10M + 2, \quad M = a^2 + ab + b^2$$

and using (1.1) Ljunggren [2] transformed the equation (S-W) to the following form:

$$x^2 + 3y^2 + 12z = 4M. \tag{L}$$

Moreover, he proved that total number $f(N)$ of solutions of the Diophantine equation (L) is equal to

$$f(N) = \frac{\pi\sqrt{3}}{180}N + k_1\sqrt{N}, \tag{L_1}$$

where k_1 is bounded and is independent of N. From (L_1) immediately follows that

$$\lim_{N\to\infty} \frac{f(N)}{N} = \frac{\pi\sqrt{3}}{180} \approx 0.03.$$

Geometrically, the formula (L_1) denote that the points (x, y) satisfying of the equation (L) all lie in the neighbourhood of the two lines:

$$y = 0.03x, \quad y = 0.015x.$$

On page 54 of the paper [2] Ljunggren remarked (see [2, p. 54]) that the following problem is important for applications in virology:

Ljunggren's Problem. *Find all odd, positive integers x, y and all non-negative integers z satisfying the equation* (L) *for given non-negative integers values of a and b.*

In this paper we give an effective method for the solution of this Ljunggren's Problem.

2. Solution of the Ljunggren's Problem

The Diophantine equation (L) we can present in the following form

$$x^2 + 3y^2 = 4(M - 3z), \tag{2.1}$$

where $M = a^2 + ab + b^2$ and a, b are given non-negative integers. Since $x^2 + 3y^2 \geqslant 0$ and $z \geqslant 0$, then by (2.1) it follows that

$$0 \leqslant z \leqslant \frac{1}{3}M. \tag{2.2}$$

From (2.2) follows that there is only finite number of integers z satisfying (2.2), because for given non-negative a, b the number $M = a^2 + ab + b^2$ is fixed.

Now, let $z = z_0 \in \left[0, \frac{1}{3}M\right]$ and let

$$M_0 = M - 3z_0. \tag{2.3}$$

From (2.1) and (2.3) we have

$$x^2 + 3y^2 = 4M_0. \tag{2.4}$$

Since M_0 is non-negative integer then we can present this number in the form

$$M_0 = 2^{\alpha} p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}, \tag{2.5}$$

where $\alpha \geqslant 0$, $\alpha_j \geqslant 1$ are integers for $j = 1, 2, \ldots, r$ and p_j are odd distinct primes. Substituting (2.5) to (2.4) we obtain

$$x^2 + 3y^2 = 2^{\alpha+2} p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}. \tag{2.6}$$

From (2.6) and well-known properties of the divisibility and congruence relations we get

$$p_j \mid x^2 + 3y^2 \;\Leftrightarrow\; x^2 \equiv -3y^2 \pmod{p_j}, \tag{2.7}$$

for each $j = 1, 2, \ldots, r$.

By (2.7) and the properties of the Legendre's symbol it follows that

$$\left(\frac{-3y^2}{p_j}\right) = \left(\frac{-3}{p_j}\right)\left(\frac{y^2}{p_j}\right) = \left(\frac{-3}{p_j}\right)\left(\frac{y}{p_j}\right)^2 = \left(\frac{-3}{p_j}\right) = +1.$$

It is to observe that the equality $\left(\frac{-3}{p_j}\right) = +1$, imply that for each $j = 1, 2, \ldots, r$ the prime p_j is the form $p_j = 6k_j + 1$.

Indeed, suppose that for some $j = 1, 2, \ldots, r$ the equality $\left(\frac{-3}{p_j}\right) = +1$, imply that $p_j \neq 6k_j + 1$. Since p_j is prime, then $p_j = 6m_j + 5$.

Hence, by well-known properties of Legendre's symbol it follows that

$$\left(\frac{-3}{p_j}\right) = \left(\frac{-1}{p_j}\right)\left(\frac{3}{p_j}\right) = (-1)^{\frac{p_j-1}{2}}\left(\frac{3}{p_j}\right). \tag{2.8}$$

On the other hand from the Gauss reciprocity law we have

$$\left(\frac{3}{p_j}\right)\left(\frac{p_j}{3}\right) = (-1)^{\frac{(3-1)(p_j-1)}{2}} = (-1)^{\frac{p_j-1}{2}}. \tag{2.9}$$

Since $p_j = 6m_j + 5$, then we have

$$\left(\frac{p_j}{3}\right) = \left(\frac{6m_j+5}{3}\right) = \left(\frac{2}{3}\right) = -1. \tag{2.10}$$

By (2.9) and (2.10) it follows that

$$\left(\frac{3}{p_j}\right) = (-1)^{\frac{p_j-1}{2}+1}. \tag{2.11}$$

From (2.11) and (2.8) follows that

$$\left(\frac{-3}{p_j}\right) = (-1)^{p_j} = (-1)^{6m_j+5} = -1, \tag{2.12}$$

so proves our assertion. This fact implies that every odd prime p_j of the right hand of (2.6) is the form

$$p_j = 6k_j + 1, \quad j = 1, 2, \ldots, r.$$

By the Theorem 5 of the monograph [4, p. 349] it follows that every prime p which is of the form $p = 6k + 1$ is of the form $p = m^2 + 3n^2$, where m, n are positive integers. Therefore, we have

$$p_j = x_j^2 + 3y_j^2, \quad \text{for every} \quad j = 1, 2, \ldots, r.$$

Now, we note that if the equation (2.6) has a solution in odd positive integers x, y then we have

$$2^{\alpha+2} \mid x^2 + 3y^2. \tag{2.13}$$

Since $x = 2v - 1$ and $y = 2w - 1$ then

$$x^2 + 3y^2 = (2v-1)^2 + 3(2w-1)^2 = 4[v(v-1) + 3w(w-1) + 1]. \tag{2.14}$$

By (2.13) and (2.14) it follows that

$$2^{\alpha} \mid v\,(v-1) + 3w\,(w-1) + 1. \tag{2.15}$$

It is easy to see that the sum $v\,(v-1) + 3w\,(w-1) + 1$ is odd positive integer for any positive integers v, w and consequently the relation (2.15) is impossible for any

positive integers $\alpha \geqslant 1$. Since $\alpha \geqslant 0$, then we have $\alpha = 0$ and the equation (2.6) reduces to the form:

$$x^2 + 3y^2 = 4p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}, \tag{2.16}$$

where

$$p_j = x_j^2 + 3y_j^2, \quad \alpha_j \geqslant 1, \quad j = 1, 2, \ldots, r. \tag{2.17}$$

The representation p_j in the form (2.17) is unique. This fact follows by the Theorem 10 on page 221 of [4].

Further, we note that the following identity is true,

$$\left(u^2 + 3v^2\right)\left(r^2 + 3s^2\right) = (ur - 3vs)^2 + 3(us + vr)^2. \tag{2.18}$$

From (2.18) and by uniqueness representation the prime number p_j in the form (2.17) it follows that

$$p_j^{\alpha_j} = \left(x_j^2 + 3y_j^2\right)^{\alpha_j} = R_j^2 + 3S_j^2, \quad j = 1, 2, \ldots, r, \tag{2.19}$$

where R_j, S_j are positive integers of different parity and representation (2.19) is also unique.

Moreover, we remark that we can determine R_j and S_j in explicit form. Namely, we have

$$R_j = \frac{\left(x_j + i\sqrt{3}y_j\right)^{\alpha_j} + \left(x_j - i\sqrt{3}y_j\right)^{\alpha_j}}{2}, \quad S_j = \frac{\left(x_j + i\sqrt{3}y_j\right)^{\alpha_j} - \left(x_j - i\sqrt{3}y_j\right)^{\alpha_j}}{i\sqrt{3}}, \tag{2.20}$$

for $j = 1, 2, \ldots, r$. By (2.19), (2.16) and (2.18) it follows that

$$x^2 + 3y^2 = 4\prod_{j=1}^{r}\left(R_j^2 + 3S_j^2\right) = 4\left(R^2 + 3S^2\right), \tag{2.21}$$

where R, S are positive integers of different parity and are effectively determined by (2.18), (2.20) and (2.21).

Now, we observe that

$$4 = 1^2 + 3 \times 1^2. \tag{2.22}$$

From (2.22) and (2.18) we get

$$4\left(R^2 + 3S^2\right) = \left(1^2 + 3 \times 1^2\right)\left(R^2 + 3S^2\right) = (R - 3S)^2 + 3(R + S)^2. \tag{2.23}$$

By (2.21) and (2.23) it follows that odd positive integers satisfy the following equation:

$$x^2 + 3y^2 = (R - 3S)^2 + 3(R + S)^2. \tag{2.24}$$

Immediately, from (2.24) we get that

$$x = |R - 3S|, \quad y = R + S, \tag{2.25}$$

and the positive integers x, y determined by the formula (2.25) are both odd and satisfy the virulogical Ljunggren's Diophantine equation (L).

On the other hand we note that the representation of (2.24) can be nonuniquenes and for determined eventuelle other solutions of (2.24) we can applied the following estimate, whose immediately follows from (2.21);

$$x < 4\max\{R, S\}, \quad y < 3\max\{R, S\}. \tag{2.26}$$

In this way we determine all odd positive integers solutions of the Ljunggren's Diophantine equation (L).

3. Remark and an example

We note that if the equation (2.6) has a solution in odd positive integers x, y then the number $M - 3z$ on the right hand of (2.6) must be odd non-negative integer. Therefore, if M is odd then it suffices consider only even non-negative integers $z \in \left[0, \frac{1}{3}M\right]$.

The following example is illustration of our method for this case:

Let $a = 5, b = 3$. Then $M = a^2 + ab + b^2 = 5^2 + 5 \times 3 + 3^2 = 49$ and consequently the equation (2.6) has the form:

$$x^2 + 3y^2 = 4(49 - 3z), \tag{3.1}$$

where $0 \leqslant z \leqslant 49\frac{1}{3}$. Since z must be even integer then we can consider only the case when $z = 0, 2, 4, 6, 8, 10, 12, 14$ and 16.

If $z = 0$ then the equation (3.1) has the form:

$$x^2 + 3y^2 = 4 \times 7^2.$$

Since $7 = 6 \times 1 + 1 = 2^2 + 3 \times 1^2$, then by (2.18) it follows that $7^2 = 1^2 + 3 \times 4^2$ and we have $R = 1, S = 4$, so

$$x = |R - 3S| = |1 - 12| = 11, \quad y = R + S = 1 + 4 = 5.$$

Moreover, using (2.26) we obtain second solution, $x = y = 7$.

If $z = 2$, then $M - 6 = 49 - 6 = 43 = 6 \times 7 + 1 = 4^2 + 3 \times 3^2$, so $R = 4$, $S = 3$ and $x = 5$, $y = 7$ or $x = 13$, $y = 1$.

If $z = 4$, then $M - 12 = 37 = 6 \times 6 + 1 = 5^2 + 3 \times 2^2$, so $R = 5$, $S = 2$ and we have $x = 1$, $y = 7$ or $x = 11$, $y = 3$.

If $z = 6$, then we obtain $M - 18 = 31 = 6 \times 5 + 1 = 2^2 + 3 \times 3^2$, so $R = 2$, $S = 3$, and $x = 7$, $y = 5$ or $x = 11$, $y = 1$.

If $z = 8$, then $M - 24 = 25 = 5^2$ and $5 \neq 6k + 1$, so the equation (2.6) has no solutions.

If $z = 10$, then $M - 30 = 19 = 6 \times 3 + 1 = 4^2 + 3 \times 1^2$, so $R = 4$, $S = 1$ and $x = 1$, $y = 5$ or $x = 7$, $y = 3$.

If $z = 12$, then $M - 36 = 13 = 6 \times 2 + 1 = 1^2 + 3 \times 2^2$, so $R = 1$, $S = 2$ and $x = 5$, $y = 3$ or $x = 7$, $y = 1$.

If $z = 14$, then $M - 42 = 7 = 6 \times 1 + 1 = 2^2 + 3 \times 1^2$, so $R = 2$, $S = 1$ and $x = 1$, $y = 3$ or $x = 5$, $y = 1$.

If $z = 16$, then $M - 48 = 1$ and we have $x^2 + 3y^2 = 4$, so there is only one trivial solution in odd positive integers, namely $x = y = 1$.

References

[1] Goldberg, M., *A class of multi-symmetric polyhedra*, Tohoku Math. J. 43 (1937) 104–108.

[2] Ljunggren, W., *Diophantine analysis applied to virus structure,* Math. Scand. 34 (1974) 51–57 (presented by N. G. Wrigley and V. Brun).

[3] Schlegel, H. G., *Mikrobiologia ogólna*, PWN, Warszawa, 2005 (Polish Edition-Allgemeine Mikrobiologie, Georg Thieme Verlag, Stuttgart, 1992).

[4] Sierpiński, W., *Elementary Theory of Numbers*, PWN, Warszawa, 1987 (Editor: A. Schinzel).

[5] Stoltz, D. B., *The structure of icosahedral cytoplasmic Deoxyriboviruses,* J. Ultrastruct. Res. 37 (1971) 219–239.

[6] Stoltz, D. B., *The structure of icosahedral cytoplasmic Deoxyriboviruses, II: An alternative model,* J. Ultrastruct. Res. 43 (1973) 58–74.

[7] Wrigley, N. G., *An electron microscope study of the structure of Sericesthis iridescent virus,* J. gen. Virol. (1969) 123–134.

[8] Wrigley, N. G., *An electron microscope study of the structure of Tipula iridescent virus,* J. gen. Virol. 6 (1969) 169–173.

Aleksander Grytczuk
Faculty of Mathematics,
Computer Science and Econometrics
University of Zielona Góra
ul. Prof. Szafrana 4a
65-516 Zielona Góra
Poland

Annales Mathematicae et Informaticae
***33** (2006) pp. 77–91*
http://www.ektf.hu/tanszek/matematika/ami

Positive trigonometric sums and applications

Stamatis Koumandos

University of Cyprus
e-mail: skoumand@ucy.ac.cy

Submitted 12 September 2005; Accepted 1 June 2006

Abstract

Some new positive trigonometric sums that sharpen Vietoris's classical inequalities are presented. These sharp inequalities have remarkable applications in geometric function theory. In particular, we obtain information for the partial sums of certain analytic functions that correspond to starlike functions in the unit disk. We also survey some earlier results with additional remarks and comments.

Keywords: Positive trigonometric sums, starlike functions.

MSC: 42A05, 42A32, 26D05, 30C45.

1. Introduction

The problem of constructing positive trigonometric sums is very old and has been dealt with by many authors. The most familiar examples are the Fejér-Jackson-Gronwall inequality

$$\sum_{k=1}^{n} \frac{\sin k\theta}{k} > 0, \quad \text{for all} \quad n \in \mathbb{N} \ \text{ and } \ 0 < \theta < \pi, \tag{1.1}$$

and Young's inequality

$$1 + \sum_{k=1}^{n} \frac{\cos k\theta}{k} > 0, \quad \text{for all} \quad n \in \mathbb{N} \ \text{ and } \ 0 < \theta < \pi. \tag{1.2}$$

As it is known, inequality (1.1) conjectured by Fejér in 1910, and its first proof published by D. Jackson [20], in 1911 and also was proved independently by T. H. Gronwall in [22] few months later. Inequality (1.2) was established by W. H. Young

[35], in 1913. Since then these inequalities have attracted the attention of several mathematicians who offered new and simpler proofs, various generalizations and extensions of different type. A complete account of related results, historical comments and an extensive bibliography can be found in monograph [10] and also in [26, Ch.4] and in the survey article [12]. It is worth mentioning that these inequalities are naturally incorporated in the context of more general results for classical orthogonal polynomials. (cf. [10, 11, 12]). Several applications have indicated which generalizations of (1.1) and (1.2) are essential and have led to a deeper understanding of these results. Conversely, a variety of problems reduces to positivity results for trigonometric or other orthogonal sums of this type. Indeed, these inequalities have remarkable applications in the theory of Fourier series, summability theory, approximation theory, positive quadrature methods, the theory of univalent functions and many others.

We refer the reader to the recently published research articles [1, 2, 3, 4, 5, 6, 7], [13], [16], [19], [21] and [23] for some new results on positive trigonometric sums including refinements and extensions of (1.1) and (1.2) and various applications. Of course, we cannot survey this whole subject here and we restrict ourselves on some recently found refinements of a far reaching extension of (1.1) and (1.2) due to Vietoris and some applications of them in geometric function theory. We also summarize some earlier closely related results. We note that positivity results for trigonometric sums and geometric function theory have been closely related subjects over the past century. Both areas have taken and given to each other and this paper intends to present few more results of this interplay.

2. Vietoris's inequalities

In 1958 L. Vietoris [34] gave a striking generalization of both (1.1) and (1.2). In particular, he showed that if $a_k,\ k = 0, 1, 2\ldots$ is a decreasing sequence of positive real numbers such that

$$2ka_{2k} \leqslant (2k-1)a_{2k-1}, \quad k \geqslant 1, \tag{2.1}$$

then for all positive integers n, we have

$$\sum_{k=0}^{n} a_k \cos k\theta > 0, \quad 0 < \theta < \pi, \tag{2.2}$$

and

$$\sum_{k=1}^{n} a_k \sin k\theta > 0, \quad 0 < \theta < \pi. \tag{2.3}$$

Vietoris observed that (2.2) and (2.3) follow by a partial summation from the special case $a_k = c_k$, where

$$c_0 = c_1 = 1 \quad \text{and} \quad c_{2k} = c_{2k+1} = \frac{1.3.5\ldots(2k-1)}{2.4.6\ldots 2k}, \quad k \geqslant 1\,. \tag{2.4}$$

Conversely, this c_k is an extreme sequence in (2.1). It is clear that the sequence $a_0 = 1,\ a_k = \frac{1}{k},\ k \geqslant 1$ satisfies (2.1), hence inequalities (1.1) and (1.2) are obtained by Vietoris's result.

The importance of Vietoris's inequalities became widely known after the work of R. Askey and J. Steinig [9], (see also [8, p. 375]), who gave a simplified proof of them and showed that they have some nice applications in estimating the zeros of certain trigonometric polynomials. Askey and Steinig also observed that these inequalities are better viewed in the context of more general inequalities concerning positive sums of Jacobi polynomials and they play a role in problems dealing with quadrature methods. (See the comments and remarks in [10, p. 87]). Vietoris's theorem is nowadays one of the most cited results in the area having received attention from several authors who offered various extensions and generalizations. Several new applications of these inequalities have also been given. For instance, S. Ruscheweyh [31] used them to derive some coefficient conditions for starlike univalent functions. In a recent work, S. Ruscheweyh and L. Salinas [33] gave a beautiful interpretation of Vietoris theorem in geometric function theory and they pointed a new direction of applications for this type of results. For more background information, we refer the reader to the recent paper [23].

It is the aim of this article to present some recently found generalizations and extensions of Vietoris's inequalities which are useful in applications and to provide some additional comments and remarks on these results.

We first observe that Vietoris's sine inequality cannot be much improved if we require all the sums in (2.3) to be positive. Indeed, suppose that $a_0 \geqslant a_1 \geqslant \ldots \geqslant a_n > 0$. The condition

$$\sum_{k=1}^{n} (-1)^{k-1} k\, a_k \geqslant 0, \quad \text{for all } \ n \geqslant 1 \tag{2.5}$$

is necessary for the positivity of all these sine sums in $(0,\, \pi)$. (Divide (2.3) by $\sin\theta$ and take the limit as $\theta \to \pi$ to obtain (2.5)). Obviously, for a non-negative sequence (a_n), the condition (2.5) is equivalent to

$$\sum_{k=1}^{n} ((2k-1)a_{2k-1} - 2k a_{2k}) \geqslant 0 \quad \text{for all } \ n \geqslant 1,$$

which holds as an equality for the extreme sequence (2.4).

In an impressive paper, A. S. Belov [13] proved that the condition (2.5) is also sufficient for the validity of (2.3) and it also implies the positivity of the corresponding cosine sums (2.2). Clearly, Belov's result implies Vietoris's theorem. It should be noted that if either (2.1) or (2.5) is weakened then the sums in (2.3) are not everywhere positive in $(0,\, \pi)$. It is possible, however, to have everywhere positive sine sums in (2.3) under weaker conditions on the coefficients in the case where n is odd. This will be discussed in the next section.

We now turn to Vietoris's cosine inequality (2.2) which has received a substantial improvement and sharpening over the past twenty years.

The first result in this direction is due to G. Brown and E. Hewitt [15] who proved that all the cosine sums in (2.2) remain positive when (2.1) is replaced by the weaker condition $(2k+1)\,a_{2k} \leqslant 2k\,a_{2k-1}$, $k \geqslant 1$. It is interesting to observe that their result follows from the particular case $a_k = p_k$, where $p_{2k} = p_{2k+1} = \dfrac{2.4.6.\ldots.(2k)}{3.5.7\ldots(2k+1)}$.

In [13], A. S. Belov established another sufficient condition for the positivity of cosine sums (2.2). Namely, let $a_{2k} = a_{2k+1} = \gamma_k$, where γ_k is a decreasing sequence of non-negative real numbers satisfying

$$\sum_{k=0}^{m} \gamma_k - \sum_{k=m}^{n} \gamma_k + \frac{2}{3}\,(n-3\,m)\,\gamma_m \geqslant 0 \ \text{ for all } \ n \geqslant 1, \tag{2.6}$$

where $m = \left[\frac{n+1}{3}\right]$, the square brackets denoting the integer part of $(n+1)/3$, then inequality (2.2) holds. Take now as γ_k the sequence $\dfrac{1.3.5\ldots(2k-1)}{2.4.6\ldots 2k}$ and see that this sequence satisfies both (2.5) and (2.6). Then take as γ_k the sequence $\dfrac{2.4.6.\ldots.(2k)}{3.5.7\ldots(2k+1)}$ and observe that this satisfies only (2.6). So, Belov's result implies both the Vietoris and the Brown-Hewitt theorem. Condition (2.6), however, provides no best possible extension of Vietoris's cosine inequality. Consider, for example $a_0 = a_1 = 1$ and $a_{2k} = a_{2k+1} = \gamma_k = \dfrac{3.5\ldots(2k+1)}{4.6\ldots(2k+2)}$, $k = 1, 2, \ldots$. G. Brown and Q. Yin showed in [17] that for this choice of coefficients the cosine sums (2.2) are positive. This is a further extension of both Vietoris and Brown-Hewitt cosine inequalities which is, still, not best possible. We observe that for this sequence γ_k inequality (2.6) fails to hold for some n, therefore this sharpening is not deduced from Belov's result. Other examples of this type will be given in the next section.

In [17] the following direction for a further improvement of (2.2) was suggested. Suppose that $a_0 \geqslant a_1 \geqslant \ldots \geqslant a_n > 0$ such that

$$\frac{a_{2k}}{a_{2k-1}} \leqslant \frac{2k+\beta-1}{2k+\beta}, \quad k \geqslant 1, \tag{2.7}$$

and determine the maximum value of $\beta > 0$, such that condition (2.7) implies (2.2) for all n. The authors observed in [17] that this value of β does not exceed 2.34. Clearly, it is sufficient to consider the extreme sequence $a_k = e_k$ for which we have equality in (2.7). This can be written as $e_0 = e_1 = 1$ and

$$e_{2k} = e_{2k+1} = \delta_k := \frac{\left(\dfrac{1+\beta}{2}\right)_k}{\left(\dfrac{2+\beta}{2}\right)_k}, \quad k = 0, 1, 2, \ldots, \tag{2.8}$$

using the Pochhammer symbol,

$$(a)_0 = 1, \ \text{and} \ \ (a)_k = a(a+1)\ldots(a+k-1) = \frac{\Gamma(k+a)}{\Gamma(a)}, \ \text{for } k = 1, 2, \ldots.$$

Observe that for $\beta = 0, 1, 2$ in (2.8), we obtain the Vietoris, Brown-Hewitt and Brown-Yin results respectively.

The maximum value of β for which condition (2.7) implies the positivity of cosine sums (2.2) is $\beta = 2.33088\ldots$, and it is determined by the case $n = 6$. This has been obtained in [23]. A different extension of Vietoris cosine inequality is used in the proof of this result. This extension and some of its consequences will be presented in the next section.

We complete this section by observing that for the sequence δ_k in (2.8) we have $\delta_k \sim \dfrac{1}{k^{\frac{1}{2}}}$, as $k \to \infty$ for all $\beta > 0$.

If we replace δ_k in (2.8) by $d_k := \dfrac{(1-\alpha)_k}{k!}$, $0 < \alpha < 1$, we see that $d_k \sim \dfrac{1}{k^\alpha}$, as $k \to \infty$, and that Vietoris's sequence (2.4) corresponds to the case $\alpha = \frac{1}{2}$.

3. Extensions of Vietoris cosine inequality

Vietoris's cosine inequality admits the following sharpening.

Theorem 3.1. *Let* $0 < \alpha < 1$ *and*

$$c_0 = c_1 = 1$$
$$c_{2k} = c_{2k+1} = \frac{(1-\alpha)_k}{k!}, \quad k = 1, 2, \ldots .$$

For all positive integers n *and* $0 < \theta < \pi$, *we have*

$$\sum_{k=0}^{n} c_k \cos k\theta > 0,$$

when $\alpha \geqslant \alpha_0$, *where* α_0 *is the unique solution in* $(0, 1)$ *of the equation*

$$\int_0^{\frac{3\pi}{2}} \frac{\cos t}{t^\alpha}\, dt = 0.$$

Also for $\alpha < \alpha_0$

$$\lim_{n\to\infty} \min\left\{\sum_{k=0}^{n} c_k \cos k\theta : \theta \in (0,\, \pi)\right\} = -\infty. \tag{3.1}$$

Numerical methods give $\alpha_0 = 0.3084437\ldots$.

Let us denote

$$d_k = \frac{(1-\alpha)_k}{k!}, \quad k = 0, 1, \ldots .$$

Notice that for the sequence c_k of the theorem above we have

$$\sum_{k=0}^{2n+1} c_k \cos k\theta = 2\cos\frac{\theta}{2}\sum_{k=0}^{n} d_k \cos\big(2k+\frac{1}{2}\big)\theta, \quad 0<\theta<\pi,$$

therefore an immediate consequence of this theorem is

$$\sum_{k=0}^{n} d_k \cos\big(2k+\frac{1}{2}\big)\theta > 0 \quad \text{for all } n \text{ and } \quad 0<\theta<\pi, \tag{3.2}$$

if and only if $\alpha \geqslant \alpha_0$.

We also observe that

$$\sin\frac{\theta}{2}\sum_{k=0}^{2n+1} c_k \cos k\theta = \cos\frac{\theta}{2}\sum_{k=1}^{2n+1} c_k \sin k(\pi-\theta)$$

which is to say that inequalities

$$\sum_{k=0}^{2n+1} c_k \cos k\theta > 0, \quad 0<\theta<\pi \tag{3.3}$$

and

$$\sum_{k=1}^{2n+1} c_k \sin k\theta > 0, \quad 0<\theta<\pi \tag{3.4}$$

are equivalent, hence inequality (3.4) holds for all positive integers n and $0<\theta<\pi$, if and only if $\alpha \geqslant \alpha_0$.

The situation is different when we are looking for a corresponding result for even sine sums, that is,

$$\sum_{k=1}^{2n} c_k \sin k\theta > 0, \quad 0<\theta<\pi. \tag{3.5}$$

Taking into consideration the necessary and sufficient condition (2.5) (or its equivalent version given in Section 2) we infer that (3.5) holds precisely when $\alpha \geqslant 1/2$, hence Vietoris result is, in this case, best possible.

We note that for sine sums having coefficients c_k there is no analogue of (3.1) when $\alpha<\alpha_0$. Indeed, these sums are uniformly bounded below on $(0,\,\pi)$ because the conjugate Dirichlet kernel $\widetilde{D}_n(\theta)=\sum_{k=1}^{n}\sin k\theta$ satisfies $\widetilde{D}_n(\theta)\geqslant -\frac{1}{2}$ for all n and $\theta\in(0,\,\pi)$.

In order to prove Theorem 3.1 we have to consider only the critical case $\alpha=\alpha_0$, then the full result follows by a partial summation. Another interesting observation here is that for the sequence d_k, with $\alpha=\alpha_0$, condition (2.6) fails to hold.

A proof of Theorem 3.1 is given in [23]. A different proof was independently obtained in [18].

In the section that follows, we shall see that the sharpening of Vietoris inequality given in Theorem 3.1 is not artificial and that sharp results of this type are necessary in the resolution of some specific problems in geometric function theory. Inequality (3.2) was an important ingredient in the proof of a conjecture regarding subordination of certain starlike functions, originally presented in [24], then settled in [25]. This, as well as a generalized version of (3.2) will be given next.

4. Applications to geometric function theory

We first, recall some necessary definitions, notations, and background results.

Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ be the unit disk in the complex plane $\mathbb{C}$ and $A(\mathbb{D})$ be the space of analytic functions in $\mathbb{D}$. It is well known that $A(\mathbb{D})$ is a locally convex linear topological space with respect to the topology given by uniform convergence on compact subsets of $\mathbb{D}$. For $\lambda < 1$ let $\mathcal{S}_\lambda$ be the family of functions f *starlike of order* λ, i.e.

$$\mathcal{S}_\lambda = \Big\{ f \in A(\mathbb{D}) : \, f(0) = f'(0) - 1 = 0 \ \text{ and } \ \mathrm{Re}\Big(z\, \frac{f'(z)}{f(z)} \Big) > \lambda, \ \ z \in \mathbb{D} \Big\}.$$

The family $\mathcal{S}_\lambda$ was introduced by M. S. Robertson in [29], and since then it has been the subject of systematic study by several researchers. We note that $\mathcal{S}_\lambda$ is a compact subset of $A(\mathbb{D})$ and that $f_\lambda(z) := \dfrac{z}{(1 - e^{-it}\, z)^{2-2\lambda}}$ belong to $\mathcal{S}_\lambda$, for all $t \in \mathbb{R}$, and they represent the extreme points of the closed convex hull of $\mathcal{S}_\lambda$. We have

$$\overline{\mathrm{conv}}(\mathcal{S}_\lambda) = \Big\{ \int_0^{2\pi} \frac{z}{(1 - e^{-it}\, z)^{2-2\lambda}} \, d\mu(t) \, : \, \mu \in \mathcal{P}(\mathbb{T}) \Big\}, \tag{4.1}$$

where $\mathcal{P}(\mathbb{T})$ denotes the set of all probability measures on the unit circle $\mathbb{T}$. Also

$$\mathrm{ex}(\overline{\mathrm{conv}}(\mathcal{S}_\lambda)) = \Big\{ \frac{z}{(1 - \chi\, z)^{2-2\lambda}} \, : \, \chi \in \mathbb{T} \Big\} \subset \mathcal{S}_\lambda$$

(cf. [14] and [30]). Suppose that

$$f(z) = z \sum_{k=0}^{\infty} a_k \, z^k \in \mathcal{S}_\lambda.$$

Then we have

$$\mathrm{Re}\Big\{ f(z)/z \Big\} = \mathrm{Re} \sum_{k=0}^{\infty} a_k \, z^k > \frac{1}{2}, \quad \text{when} \quad \frac{1}{2} \leqslant \lambda < 1,$$

but this conclusion is not necessarily true for all the partial sums of such a function. For an analytic function $f(z) = \sum_{k=0}^{\infty} a_k z^k$ and $n \in \mathbb{N}$ we set $s_n(f, z) = \sum_{k=0}^{n} a_k z^k$, for the n-th partial sum of f.

It has been shown in [32] that $\operatorname{Re} s_n(f/z, z) > 0$ holds in $\mathbb{D}$ for all $n \in \mathbb{N}$ and for all $f \in \mathcal{S}_{3/4}$ and it has been pointed out that the number $\frac{3}{4}$ can probably be replaced by a smaller one. The smallest value of λ such that $\operatorname{Re} s_n(f/z, z) > 0$ holds in $\mathbb{D}$ for all $n \in \mathbb{N}$ and for all $f \in \mathcal{S}_\lambda$, has been determined in [24]. Indeed, we have the following.

Theorem 4.1. *For all $n \in \mathbb{N}$ and $z \in \mathbb{D}$, we have*

$$Re\, s_n(f/z,\, z) > 0, \quad for\ \ all \quad f \in \mathcal{S}_\lambda\,,$$

if and only if $\lambda_0 \leqslant \lambda < 1$, where λ_0 is the unique solution in $(\frac{1}{2},\, 1)$ of the equation

$$\int_0^{3\pi/2} t^{1-2\lambda} \cos t\, dt = 0\,.$$

In fact,

$$\lambda_0 = \frac{1+\alpha_0}{2} = 0.654222\ldots,$$

where α_0 is as in Theorem 3.1.

We give the idea of the proof of this theorem, in order to see that the sharp versions of positive trigonometric sums are indispensable.

Let

$$f(z) = z \sum_{k=0}^{\infty} a_k\, z^k \in \mathcal{S}_\lambda.$$

It follows from (4.1) that

$$a_k = \hat{\mu}(k)\, \frac{(2-2\lambda)_k}{k!}\,, \quad k = 0, 1, \ldots$$

where $\hat{\mu}(k)$ are the Fourier coefficients of the measure μ. Since

$$s_n(f/z, z) = \sum_{k=0}^{n} a_k\, z^k\,,$$

we deduce from the above that

$$\operatorname{Re}\, s_n(f/z,\, e^{i\theta}) = \int_0^{2\pi} \sum_{k=0}^{n} \frac{(2-2\lambda)_k}{k!}\ \cos k(\theta - t)\, d\mu(t)\,.$$

By the minimum principle for harmonic functions it suffices to prove that

$$\sum_{k=0}^{n} \frac{(2-2\lambda)_k}{k!}\ \cos k\theta > 0, \quad \forall\, n \in \mathbb{N}, \quad \forall\, \theta \in \mathbb{R}\,, \tag{4.2}$$

if and only if $\lambda_0 \leqslant \lambda < 1$. This inequality is different from the one given in Theorem 3.1, in the sense that none implies the other. The proof of both requires

several sharp estimates and a delicate calculus work. To get the flavor of this and the common features of (4.2) with Theorem 3.1, we set $\lambda = \frac{1+\alpha}{2}$ and consider the following limiting case

$$\lim_{n\to\infty} \left(\frac{\theta}{n}\right)^{1-\alpha} \sum_{k=0}^{n} \frac{(1-\alpha)_k}{k!} \cos k\,\frac{\theta}{n} = \frac{1}{\Gamma(1-\alpha)} \int_0^\theta \frac{\cos t}{t^\alpha}\, dt. \tag{4.3}$$

It follows from this that for $\theta = 3\pi/2$ and $\lambda < \lambda_0 = \frac{1+\alpha_0}{2}$, the right hand side of (4.3) will be negative, therefore inequality (4.2) cannot hold for $\lambda < \lambda_0$, appropriate θ and n sufficiently large. See also the discussion in [36, V, 2.29]. There is a simple way of proving (4.3), which reflects the idea of the proof of (4.2). Let

$$\Delta_k := \frac{1}{\Gamma(1-\alpha)\, k^\alpha} - \frac{(1-\alpha)_k}{k!}, \quad k = 1, 2, \ldots, \quad 0 < \alpha < 1.$$

Since

$$\frac{\Gamma(x+1-\alpha)}{\Gamma(x+1)}\, x^\alpha = 1 - \frac{\alpha\,(1-\alpha)}{2}\, \frac{1}{x} + O\Big(\frac{1}{x^2}\Big), \quad \text{as } x \to \infty\,,$$

(See [8]), we have

$$\Delta_k = O\Big(\frac{1}{k^{\alpha+1}}\Big), \quad \text{as } \ k \to \infty.$$

On the other hand,

$$\sum_{k=1}^{n} \frac{1}{k^{\alpha+1}} = \zeta(\alpha+1) + O\Big(\frac{1}{n^\alpha}\Big), \quad \text{as } \ n \to \infty.$$

So that, putting everything together, we arrive at

$$\lim_{n\to\infty} \left(\frac{\theta}{n}\right)^{1-\alpha} \sum_{k=1}^{n} \Delta_k \cos k\,\frac{\theta}{n} = 0.$$

Using this and the results of [27, Part 2, Ch.1, Problems 20–21], the desired asymptotic formula (4.3) follows. The argument given above, reveals that in order to find estimates of the sums on the left hand side of (4.2) it is sufficient to look for appropriate estimates of the sums $\sum_{k=1}^{n} \frac{\cos k\theta}{k^\alpha}$, provided that sharp inequalities for the sequence Δ_k are available. Details of all of this are in [24].

An immediate consequence of (4.2) is the following. Let

$$s_n^\lambda(z) := \sum_{k=0}^{n} \frac{(2-2\lambda)_k}{k!}\, z^k.$$

Then

$$\text{Re } s_n^\lambda(z) > 0, \quad \forall\, n \in \mathbb{N}, \quad \forall\, z \in \mathbb{D}\,, \tag{4.4}$$

if and only if $\lambda_0 \leqslant \lambda < 1$. This is, of course, the particular case of Theorem 4.1 when applied to the extremal function $f_\lambda(z) := \frac{z}{(1-z)^{2-2\lambda}}$ of $\mathcal{S}_\lambda$.

Next we shall give some other ways of extending (4.4). It turns out that inequalities of this type take a very nice and natural form when the notion of complex subordination is employed. We recall the definition of subordination of analytic functions. Let $f(z),\ g(z) \in A(\mathbb{D})$. We say that $f(z)$ is *subordinate to* $g(z)$, if there exists a function $\phi(z) \in A(\mathbb{D})$ satisfying $\phi(0) = 0$ and $|\phi(z)| < 1$ such that $f(z) = g(\phi(z)),\ \forall\, z \in \mathbb{D}$. Subordination is denoted by $f(z) \prec g(z)$. If $f(z) \prec g(z)$ then $f(0) = g(0)$ and $f(\mathbb{D}) \subset g(\mathbb{D})$. Conversely, if $g(z)$ is univalent and $f(0) = g(0)$ and $f(\mathbb{D}) \subset g(\mathbb{D})$ then $f(z) \prec g(z)$. See [28, Ch.2] for proofs and several properties of analytic functions associated with subordination.

It is easily inferred that (4.4) is equivalent to

$$s_n^\lambda(z) \prec \frac{1+z}{1-z}, \quad \forall\, n \in \mathbb{N}, \quad \forall\, z \in \mathbb{D}\,.$$

Consider now the function

$$v(z) := \Big(\frac{1+z}{1-z}\Big)^{\frac{1}{2}} = \sum_{k=0}^{\infty} c_k\, z^k\,, \quad z \in \mathbb{D}\,,$$

where

$$c_0 = c_1 = 1,$$

$$c_{2k} = c_{2k+1} = \frac{\left(\frac{1}{2}\right)_k}{k!} = \frac{1.3.\ldots.(2k-1)}{2.4\ldots 2k}, \quad k = 1, 2, \ldots.$$

This is a univalent function in $\mathbb{D}$ and maps $\mathbb{D}$ onto the sector

$$\Big\{\zeta \in \mathbb{C} \,:\, |\arg\zeta| < \frac{\pi}{4}\Big\}.$$

Observe that these coefficients c_k are exactly the same as in relation (2.4) of Vietoris theorem. We note that the function $v(z)$ plays, indeed, a key role in the proof of Vietoris result as it is given in [9], and its properties inspired the geometric interpretation of this theorem as presented in [33].

Now a strengthening of (4.4) reads as follows.

Theorem 4.2. *For all $n \in \mathbb{N}$ and $z \in \mathbb{D}$ we have*

$$(1-z)^{\frac{1}{2}}\ s_n^\lambda(z) \prec \Big(\frac{1+z}{1-z}\Big)^{\frac{1}{2}} \tag{4.5}$$

if and only if $\lambda_0 \leqslant \lambda < 1$.

This theorem was stated in [24] as a conjecture which was proved in [25]. It has several other consequences for the class of starlike functions $\mathcal{S}_\lambda$. Complete details can be found in [25]. Let us summarize here some of the important facts behind the proof of this result and its relevance to positive trigonometric sums discussed in the previous section.

It is clear that

$$\frac{1}{(1-z)^{1/2}} \prec \Big(\frac{1+z}{1-z}\Big)^{\frac{1}{2}}, \quad z \in \mathbb{D},$$

therefore (4.5) implies (4.4). Accordingly, (4.5) cannot hold for $\lambda < \lambda_0$. But it is not obvious that (4.5) holds, precisely when $\lambda \geqslant \lambda_0$. It is here that the extension of Vietoris's theorem given in Section 3 is applied. We observe that (4.5) is equivalent to

$$\operatorname{Re}\Big\{(1-z)\left[s_n^\lambda(z)\right]^2\Big\} > 0. \tag{4.6}$$

By the minimum principle for harmonic functions, it suffices to establish (4.6) for $z = e^{2i\theta}$, $0 < \theta < \pi$. Let

$$P_n(\theta) := (1-e^{2i\theta})\Big\{\sum_{k=0}^{n} \frac{(2-2\lambda)_k}{k!}\, e^{2ik\theta}\Big\}^2.$$

Then we find that

$$\begin{aligned} &\operatorname{Re} P_n(\theta) \\ = \;& \Big(\sum_{k=0}^{2n+1} c_k \cos k\theta\Big)\Big(\sum_{k=0}^{2n+1} c_k \cos k(\pi-\theta)\Big) + \Big(\sum_{k=1}^{2n+1} c_k \sin k\theta\Big)\Big(\sum_{k=1}^{2n+1} c_k \sin k(\pi-\theta)\Big), \end{aligned}$$

where $c_0 = c_1 = 1$, $c_{2k} = c_{2k+1} = \dfrac{(2-2\lambda)_k}{k!}$, $k = 1, 2, \ldots$. Setting $\lambda = \frac{1+\alpha}{2}$ and using (3.3) and (3.4) we conclude that

$$\operatorname{Re} P_n(\theta) > 0, \quad \text{for } 0 < \theta < \pi,$$

precisely when $\lambda_0 \leqslant \lambda < 1$, which is the desired result.

It is readily shown that (4.5) implies

$$\operatorname{Re}\big\{(1-z)^{\frac{1}{2}}\, s_n^\lambda(z)\big\} > 0, \tag{4.7}$$

for all $n \in \mathbb{N}$ and $\lambda_0 \leqslant \lambda < 1$. It is then natural to ask for the maximum range of λ for which (4.7) is valid. For $z = e^{2i\theta}$, $0 < \theta < \pi$, this is equivalent to

$$\sum_{k=0}^{n} \frac{(2-2\lambda)_k}{k!}\, \cos\Big[\big(2k+\frac{1}{2}\big)\theta - \frac{\pi}{4}\Big] > 0\,. \tag{4.8}$$

On setting $\lambda = \frac{1+\alpha}{2}$, an argument similar to the proof of (4.3) yields the asymptotic formula

$$\begin{aligned} &\lim_{n\to\infty} \Big(\frac{\theta}{n}\Big)^{1-\alpha} \sum_{k=0}^{n} \frac{(1-\alpha)_k}{k!}\, \cos\Big[\big(2k+\frac{1}{2}\big)\,\frac{\theta}{2n} - \frac{\pi}{4}\Big] \\ = \;& \frac{1}{\Gamma(1-\alpha)} \int_0^\theta \frac{\cos\left(t-\frac{\pi}{4}\right)}{t^\alpha}\, dt. \end{aligned} \tag{4.9}$$

The integral in (4.9) is positive for all $\theta > 0$ if and only if $\alpha \geqslant \alpha'$, where α' is the unique solution in $(0, 1)$ of the equation

$$\int_0^{\frac{7}{4}\pi} \frac{\cos\left(t - \frac{\pi}{4}\right)}{t^\alpha}\, dt = 0,$$

whose numerical value is $\alpha' = 0.0923103\ldots$. Then it can be shown that inequality (4.8) holds for all n and $\theta \in (0, \pi)$ if and only if $1 > \lambda \geqslant \lambda' = \frac{1+\alpha'}{2} = 0.546155\ldots$. See [25]. Note that $\lambda' < \lambda_0 = 0.654222\ldots$.

The above results motivated us to consider the following more general problem. Let $p \in [0, 1]$. Determine the maximum range of λ, for which

$$\operatorname{Re}\left[(1-z)^p\, s_n^\lambda(z)\right] > 0, \tag{4.10}$$

for all $n \in \mathbb{N}$ and $z \in \mathbb{D}$. The cases $p = 0$ and $p = 1/2$ have been completely solved, while the case $p = 1$ follows by a partial summation from Fejér's classical inequality

$$\sum_{k=0}^{n} \sin\left(k + \frac{1}{2}\right)\theta = \frac{1 - \cos(n+1)\theta}{2\sin\frac{\theta}{2}} \geqslant 0, \quad 0 < \theta < 2\pi.$$

The conclusion is that for $p = 1$, inequality (4.10) holds for all $1/2 \leqslant \lambda < 1$. This also follows from [32, Theorem 1.1].

The general case of (4.10) reduces to a trigonometric inequality, a limiting case of which requires the positivity of the integral

$$\int_0^{\theta} \frac{\cos\left(t - \frac{p\,\pi}{2}\right)}{t^\alpha}\, dt,$$

for all $\theta > 0$. This holds true if and only if $\alpha \geqslant \alpha(p)$, where $\alpha(p)$ is the unique solution in $(0, 1)$ of the equation

$$\int_0^{(3+p)\,\frac{\pi}{2}} \frac{\cos\left(t - \frac{p\,\pi}{2}\right)}{t^\alpha}\, dt = 0\,.$$

In view of the above, we have led to the conjecture that (4.10) holds if and only if $1 > \lambda \geqslant \lambda(p) = \dfrac{1 + \alpha(p)}{2}$. This conjecture appears to be supported by numerical experimentation. For particular values of $p \in [0, 1]$ this can be proved by the methods we followed in the cases $p = 0, \frac{1}{2}$. It would be interesting, however, to settle this conjecture by a method that comprises as a whole the values of p in $[0, 1]$.

Another interesting problem is the study of the function $\alpha(p)$. Numerical evidence suggests that this is a strictly decreasing function of p for $p \in [0, 1]$.

References

[1] Alzer, H., Koumandos, S., Sharp inequalities for trigonometric sums, *Math. Proc. Camb. Phil. Soc.*, Vol. 134 (2003), 139–152.

[2] Alzer, H., Koumandos, S., A sharp bound for a sine polynomial, *Colloq. Math.*, Vol. 96 (2003), 83–91.

[3] Alzer, H., Koumandos, S., Inequalities of Fejér-Jackson type, *Monatsh. Math.*, Vol. 139 (2003), 89–103.

[4] Alzer, H., Koumandos, S., Sharp inequalities for trigonometric sums in two variables, *Illinois J. Math.*, Vol. 48 (2004), 887–907.

[5] Alzer, H., Koumandos, S., Companions of the inequalities of Fejér-Jackson, *Analysis Math.*, Vol. 31 (2005), 75–84.

[6] Alzer, H., Koumandos, S., A new refinement of Young's inequality, *Proc. Edinburgh Math. Soc.*, to appear.

[7] Alzer, H., Koumandos, S., Sub- and superadditive properties of Fejér's sine polynomial, *Bull. London Math. Soc.*, Vol. 38 (2006), 261–268.

[8] Andrews, G. E., Askey, R., Roy, R., Special functions, Cambridge University Press, Cambridge, 1999.

[9] Askey, R., Steinig, J., Some positive trigonometric sums, *Trans. Amer. Math. Soc.*, Vol. 187 (1974), 295–307.

[10] Askey, R., Orthogonal polynomials and special functions, *Regional Conf. Lect. Appl. Math.*, Vol. 21, Philadelphia, SIAM, 1975.

[11] Askey, R., Gasper, G., Positive Jacobi polynomial sums II, *Amer. J. Math.*, Vol. 98 (1976), 709–737.

[12] Askey, R., Gasper, G., Inequalities for polynomials, in: The Bieberbach Conjecture, A. Baernstein II, D. Drasin, P. Duren, A. Marden (eds.), Math. surveys and monographs (no. 21), Amer. Math. Soc., Providence, RI, 1986, 7–32.

[13] Belov, A. S., Examples of trigonometric series with nonnegative partial sums. (Russian) *Math. Sb.*, Vol. 186 (1995), 21–46; (English translation), Vol. 186 (1995), 485–510.

[14] Brickman, L., Hallenbeck, D. J., MacGregor, T. H., Wilken D. R., Convex hulls and extreme points of families of starlike and convex functions, *Trans. Amer. Math. Soc.*, Vol. 185 (1973), 413–428.

[15] Brown, G., Hewitt, E., A class of positive trigonometric sums, *Math. Ann.*, Vol. 268 (1984), 91–122.

[16] Brown, G., Wang, K., Wilson, D. C., Positivity of some basic cosine sums, *Math. Proc. Camb. Phil. Soc.*, Vol. 114 (1993), 383–391.

[17] Brown, G., Yin, Q., Positivity of a class of cosine sums, *Acta Sci. Math. (Szeged)*, Vol. 67 (2001), 221–247.

[18] Brown, G., Dai, F., Wang, K., Extensions of Vietoris's inequalities, *The Ramanujan Journal*, to appear.

[19] Dimitrov, D. K., Merlo, C. A., Nonnegative trigonometric polynomials, *Construct. Approx.*, Vol. 18 (2002), 117–143.

[20] Jackson, D., Über eine trigonometrische Summe, *Rend. Circ. Mat. Palermo*, Vol. 32 (1911), 257–262.

[21] Gluchoff, A., Hartmann, F., Univalent polynomials and non-negative trigonometric sums, *Amer. Math. Monthly*, Vol. 105 (1998), 508–522.

[22] Gronwall, T. H., Über die Gibbssche Erscheinung und die trigonometrischen Summen $\sin x + \frac{1}{2}\sin 2x + \ldots + \frac{1}{n}\sin nx$, *Math. Ann.,* Vol. 72 (1912), 228–243.

[23] Koumandos, S., An extension of Vietoris's inequalities, *The Ramanujan Journal*, to appear.

[24] Koumandos, S., Ruscheweyh, S., Positive Gegenbauer polynomial sums and applications to starlike functions, *Constr. Approx.*, Vol. 23 (2006), 197–210.

[25] Koumandos, S., Ruscheweyh, S., On a conjecture for trigonometric sums and starlike functions, submitted.

[26] Milovanović, G. V., Mitrinović, D. S., Rassias, Th. M., Topics in Polynomials: Extremal Problems, Inequalities, Zeros, World Scient., Singapore, 1994.

[27] Pólya, G., Szegő, G., Problems and Theorems in Analysis I, Springer-Verlag, Berlin, Heidelberg, New York, 1978.

[28] Pommerenke, C., Univalent functions, Vandenhoeck and Ruprecht, Göttingen, 1975.

[29] Robertson, M. S., On the theory of univalent functions, *Ann. of Math.*, (2) Vol. 37 (1936), 374–408.

[30] Ruscheweyh, S., Convolutions in geometric function theory, *Sem. Math. Sup.*, Vol. 83, Les Presses de l'Université de Montréal (1982).

[31] Ruscheweyh, S., Coefficient conditions for starlike functions, *Glasgow Math. J.*, Vol. 29 (1987), 141–142.

[32] Ruscheweyh, S., Salinas, L., On starlike functions of order $\lambda \in [\frac{1}{2}, 1)$, *Ann. Univ. Mariae Curie-Sklodowska*, Vol. 54 (2000), 117–123.

[33] Ruscheweyh, S., Salinas, L., Stable functions and Vietoris' Theorem, *J. Math. Anal. Appl.*, Vol. 291 (2004), 596–604.

[34] Vietoris, L., Über das Vorzeichen gewisser trigonometrischer summen. S.-B. *Öster. Akad. Wiss.*, Vol. 167 (1958), 125–135; Teil II: *Anzeiger Öster. Akad. Wiss.*, Vol. 167 (1959), 192–193.

[35] Young, W. H., On certain series of Fourier, *Proc. London Math. Soc.*, (2) Vol. 11 (1913), 357–366.

[36] Zygmund, A., Trigonometric Series, 3rd ed., Cambridge Univ. Press, Cambridge, 2002.

Stamatis Koumandos
Department of Mathematics and Statistics
The University of Cyprus
P.O. Box 20537, 1678 Nicosia
Cyprus

Annales Mathematicae et Informaticae
***33** (2006) pp. 93–107*
http://www.ektf.hu/tanszek/matematika/ami

Some properties of solutions of systems of neutral differential equations

Tomáš Mihály

Faculty of Science, University of Žilina, Slovakia
e-mail: tomas.mihaly@fpv.utc.sk

Submitted 26 September 2006; Accepted 28 October 2006

Abstract

The aim of this paper is to present some sufficient conditions for the oscillatory and asymptotic properties of solutions of the system of differential equations of neutral type.

Keywords: neutral equation, oscillatory solution

MSC: 34K11, 34K25, 34K40

1. Introduction

In this paper we consider three-dimensional systems of neutral differential equations of the form:

$$\begin{aligned} \big(y_1(t) - p\, y_1(t-\tau)\big)' &= \quad p_1(t)\, f_1(y_2(h_2(t))), \\ y_2'(t) &= \quad p_2(t)\, f_2(y_3(h_3(t))), \\ y_3'(t) &= \sigma\, p_3(t)\, f_3(y_1(h_1(t))), \end{aligned} \tag{1.1}$$

where $t \in R_+ = [0, \infty)$, $\sigma = 1$ or $\sigma = -1$ and the following conditions are assumed to hold without further mention:

(a) $\tau > 0$, $0 < p < 1$;

(b) $p_i \in C(R_+, R_+)$, $i = 1, 2, 3$ are not identically zero on any subinterval $[T, \infty) \subset R_+$ and

$$\int^{\infty} p_j(t)\, dt = \infty \qquad \text{for } j = 1, 2;$$

(c) $h_i \in C(R_+, R)$ and

$$\lim_{t\to\infty} h_i(t) = \infty \qquad \text{for } i = 1, 2, 3;$$

(d) $f_i(u) = |u|^{\alpha_i} \operatorname{sgn} u$ where $\alpha_i \in R$, $\alpha_i > 0$, $i = 1, 2, 3$.

The assumption (d) implies that

(e) $uf_i(u) > 0$ for $u \neq 0$ and $f_i \in C(R, R)$, $i = 1, 2, 3$ are nondecreasing functions.

Surveying the rapidly expanding literature devoted to the study of oscillatory and asymptotic properties of neutral differential equations, one finds that few papers concern systems of neutral equations (for example [1-9]). The purpose of this paper is to establish some criteria for the oscillation of the system (1.1) for the following cases

I) $\sigma = -1$ and $0 < \alpha_1\,\alpha_2\,\alpha_3 < 1$;

II) $\sigma = -1$ and $\alpha_1\,\alpha_2\,\alpha_3 = 1$;

III) $\sigma = 1$.

Another cases (for example $\sigma = -1$ and $\alpha_1 \geqslant 1$, $0 < \alpha_2 \leqslant 1$, $\alpha_3 > 1$) are studied in [6]. Theorem 1 and Theorem 2 are generalizations of results of V. N. Shevelo, N. V. Varech, A. G. Gritsai in paper [7].

For any $y_1(t)$ we define $z(t)$ by

$$z(t) = y_1(t) - p\,y_1(t - \tau).$$

Let $t_0 \geqslant 0$ be such that

$$t_1 = \min\left\{t_0 - \tau,\ \inf_{t\geqslant t_0} h_i(t),\ i = 1, 2, 3\right\} \geqslant 0.$$

A vector function $y = (y_1,\ y_2,\ y_3)$ is a solution of the system (1.1) if there exists a $t_0 \geqslant 0$ such that y is continuous on $[t_1, \infty)$, $z(t)$, $y_2(t)$, $y_3(t)$ are continuously differentiable on $[t_0, \infty)$ and y satisfies system (1.1) on $[t_0, \infty)$.

Denote by W the set of all solutions $y = (y_1,\ y_2,\ y_3)$ of the system (1.1) which exist on some ray $[T_y, \infty) \subset R_+$ and satisfy

$$\sup\left\{\sum_{i=1}^{3} |\,y_i(t)| \,:\, t \geqslant T\right\} > 0 \qquad \text{for any } T \geqslant T_y.$$

Such a solution is called a proper solution. A proper solution $y \in W$ is defined to be nonoscillatory if there exists a $T_y \geqslant 0$ such that its every component is different from zero for all $t \geqslant T_y$. Otherwise a proper solution $y \in W$ is defined to be oscillatory.

2. Some basic lemmas

We begin with some lemmas which will be useful in the sequel.

Lemma 2.1. ([2, Lemma1]) *Let* (a)–(d) *hold and* $y \in W$ *be a nonoscillatory solution of* (1.1). *Then there exists a* $t_0 \geqslant 0$ *such that* $z(t)$, $y_2(t)$, $y_3(t)$ *are monotone functions of constant sign on the interval* $[t_0, \infty)$.

Let $y = (y_1, y_2, y_3) \in W$ be a nonoscillatory solution of (1.1). Taking into account the Lemma 2.1 we obtain:

$$y_1(t)\, z(t) > 0 \qquad \text{for } t \geqslant t_0 \tag{2.1}$$

or

$$y_1(t)\, z(t) < 0 \qquad \text{for } t \geqslant t_0. \tag{2.2}$$

Denote by N^+ (or N^-) the set of components $y_1(t)$ of all nonoscillatory solutions y of system (1.1) such that (2.1) (or (2.2)) is satisfied.

For the components $y_1(t)$ of the nonoscillatory solutions hold the following lemmas.

Lemma 2.2. ([5, Lemma3]) *Let* (a) *hold and* $y_1(t) \in N^-$. *Then* $\lim\limits_{t\to\infty} y_1(t) = 0$, $\lim\limits_{t\to\infty} z(t) = 0$.

Lemma 2.3. ([3, Lemma2]) *Let* (a) *hold and* $y_1(t) \in N^+$. *If* $\lim\limits_{t\to\infty} z(t) = 0$, *then* $\lim\limits_{t\to\infty} y_1(t) = 0$.

3. Oscillation theorems

Theorem 3.1. *Assume that* $\sigma = -1$ *and*

(A1) $h_3(h_2(h_1(t))) \leqslant t$, $h_i(t)$ *are nondecreasing functions for* $i = 2, 3$;

(A2) $0 < \alpha_1\, \alpha_2\, \alpha_3 < 1$.

If

(A3)

$$\int_0^\infty p_3(v)\Big[\int_0^{h_1(v)} p_1(u)\Big(\int_0^{h_2(u)} p_2(s)ds\Big)^{\alpha_1} du\Big]^{\alpha_3} dv = \infty,$$

(A4)

$$\int_0^\infty p_2(t)\Big(\int_{h_3(t)}^\infty p_3(s)ds\Big)^{\alpha_2} dt = \infty,$$

then every proper solution $y \in W$ of (1.1) *is either oscillatory or* $y_i(t)$, *i=1,2,3 tend monotonically to zero as* $t \to \infty$.

Proof. Let $y(t) \in W$ be a nonoscillatory solution of (1.1). According to Lemma 2.1 there exists a $t_0 \geqslant 0$ such that $z(t)$, $y_2(t)$, $y_3(t)$ are monotone functions of constant sign on the interval $[t_0, \infty)$. Without loss of generality we may assume that $y_1(t) > 0$ for $t \geqslant t_0$. Then either $y_1(t) \in N^+$ or $y_1(t) \in N^-$ for $t \geqslant t_0$.

I. Let $y_1(t) \in N^+, t \geqslant t_0$. Then $z(t) > 0$, $t \geqslant t_0$ and using the assumptions (c), (d) and (b), the third equation of (1.1) implies that $y_3(t)$ is a decreasing function for $t \geqslant t_1 \geqslant t_0$.

I.1 Let $y_3(t) < 0$, $t \geqslant t_2 \geqslant t_1$. In regard of (c) there exists a $t_3 \geqslant t_2$ such that $y_3(h_3(t)) < 0$ for $t \geqslant t_3$. The assumptions (d), (b) and the second equation of (1.1) imply that $y_2(t)$ is a decreasing function for $t \geqslant t_3$.

In view of (c) there exists a $t_4 \geqslant t_3$ such that $h_3(t) \geqslant t_3$ for $t \geqslant t_4$. Using the monotonicity of $y_3(t)$ we have $y_3(h_3(t)) \leqslant y_3(t_3)$ and hence $\mid y_3(h_3(t)) \mid \geqslant K_1$, where $K_1 = -y_3(t_3) > 0$ for $t \geqslant t_4$. Raising this inequality to the power of α_2 and multiplying by $-p_2(t)$ the second equation of (1.1) implies

$$y_2'(t) \leqslant -K_1^{\alpha_2}\, p_2(t), \quad t \geqslant t_4. \tag{3.1}$$

Integrating (3.1) from t_4 to t and in regard of (b) we obtain $\lim\limits_{t\to\infty} y_2(t) = -\infty$. Therefore $y_2(t) < 0$ for $t \geqslant t_5 \geqslant t_4$.

In view of (c) there exists a $t_6 \geqslant t_5$ such that $h_2(t) \geqslant t_5$ for $t \geqslant t_6$. Using the monotonicity of $y_2(t)$ we have $y_2(h_2(t)) \leqslant y_2(t_5)$ and hence $\mid y_2(h_2(t)) \mid \geqslant K_2$, where $K_2 = -y_2(t_5) > 0$, $t \geqslant t_6$. Raising the last inequality to the power of α_1 and multiplying by $-p_1(t)$ the first equation of (1.1) implies

$$z'(t) \leqslant -K_2^{\alpha_1}\, p_1(t), \quad t \geqslant t_6. \tag{3.2}$$

Integrating (3.2) from t_6 to t and in regard of (b) we obtain $\lim\limits_{t\to\infty} z(t) = -\infty$. Therefore $z(t) < 0$ for $t \geqslant t_7 \geqslant t_6$ which is a contradiction with positivity of $z(t)$ for $t \geqslant t_0$.

I.2 Assume that $y_3(t) > 0$ for $t \geqslant t_2 \geqslant t_1$. In view of (c) there exists a $t_3 \geqslant t_2$ such that $y_3(h_3(t)) > 0$ for $t \geqslant t_3$. The assumptions (d), (b) and the second equation of (1.1) imply that $y_2(t)$ is an increasing function for $t \geqslant t_3$.

I.2.a Let $y_2(t) > 0$ for $t \geqslant t_4 \geqslant t_3$. Integrating the second equation of (1.1) from t_4 to t we obtain

$$y_2(t) \geqslant y_2(t) - y_2(t_4) = \int_{t_4}^{t} \Big(y_3(h_3(s))\Big)^{\alpha_2} p_2(s)\, ds, \quad t \geqslant t_4. \tag{3.3}$$

In regard of monotonicity of functions $h_3(t)$, $y_3(t)$ the inequality $t_4 \leqslant s \leqslant t$ may be rewritten as $\Big(y_3(h_3(t_4))\Big)^{\alpha_2} \geqslant \Big(y_3(h_3(s))\Big)^{\alpha_2} \geqslant \Big(y_3(h_3(t))\Big)^{\alpha_2}$. Then from (3.3) we get

$$y_2(t) \geqslant \Big(y_3(h_3(t))\Big)^{\alpha_2} \int\limits_{t_4}^{t} p_2(s)\,ds, \quad t \geqslant t_4.$$

In view of (c) there exists a $t_5 \geqslant t_4$ such that $h_2(t) \geqslant t_4$ for $t \geqslant t_5$. Then the last inequality holds for $h_2(t)$, $t \geqslant t_5$, too:

$$y_2(h_2(t)) \geqslant \Big(y_3(h_3(h_2(t)))\Big)^{\alpha_2} \int\limits_{t_4}^{h_2(t)} p_2(s)\,ds, \quad t \geqslant t_5. \tag{3.4}$$

Raising (3.4) to the power of α_1 and multiplying by $p_1(t)$ the first equation of (1.1) implies:

$$z'(t) \geqslant p_1(t)\,\Big(y_3(h_3(h_2(t)))\Big)^{\alpha_1\alpha_2}\Big(\int\limits_{t_4}^{h_2(t)} p_2(s)\,ds\Big)^{\alpha_1}, \quad t \geqslant t_5.$$

Integrating this inequality from t_5 to t and using the inequality $y_1(t) \geqslant z(t) \geqslant z(t) - z(t_5)$ we have

$$y_1(t) \geqslant \int\limits_{t_5}^{t} p_1(u)\,\Big(y_3(h_3(h_2(u)))\Big)^{\alpha_1\alpha_2}\Big(\int\limits_{t_4}^{h_2(u)} p_2(s)\,ds\Big)^{\alpha_1} du, \quad t \geqslant t_5. \tag{3.5}$$

In regard of monotonicity of functions $h_2(t)$, $h_3(t)$ and $y_3(t)$ the inequality $t_5 \leqslant u \leqslant t$ may be rewritten as

$$\Big(y_3(h_3(h_2(u)))\Big)^{\alpha_1\alpha_2} \geqslant \Big(y_3(h_3(h_2(t)))\Big)^{\alpha_1\alpha_2} \text{for } t \geqslant t_5.$$

Combining the last inequality and (3.5) we obtain

$$y_1(t) \geqslant \Big(y_3(h_3(h_2(t)))\Big)^{\alpha_1\alpha_2} \int\limits_{t_5}^{t} p_1(u)\Big(\int\limits_{t_4}^{h_2(u)} p_2(s)\,ds\Big)^{\alpha_1} du, \quad t \geqslant t_5. \tag{3.6}$$

In view of (c) there exists a $t_6 \geqslant t_5$ such that $h_1(t) \geqslant t_5$ for $t \geqslant t_6$. Then (3.6) holds for $h_1(t)$, $t \geqslant t_6$, too and raising to the power of α_3 we get

$$\Big(y_1(h_1(t))\Big)^{\alpha_3} \geqslant \Big(y_3(h_3(h_2(h_1(t))))\Big)^{\alpha_1\alpha_2\alpha_3}\Big[\int\limits_{t_5}^{h_1(t)} p_1(u)\Big(\int\limits_{t_4}^{h_2(u)} p_2(s)\,ds\Big)^{\alpha_1} du\Big]^{\alpha_3} \tag{3.7}$$

for $t \geqslant t_6$. Multiplying (3.7) by $-p_3(t)$ and using the third equation of system (1.1) we have

$$y_3'(t) \leqslant -p_3(t)\Big(y_3(h_3(h_2(h_1(t))))\Big)^{\alpha_1\alpha_2\alpha_3}\Big[\int\limits_{t_5}^{h_1(t)} p_1(u)\Big(\int\limits_{t_4}^{h_2(u)} p_2(s)\,ds\Big)^{\alpha_1} du\Big]^{\alpha_3} \tag{3.8}$$

for $t \geqslant t_6$. Taking into account (A1) and the monotonicity of $y_3(t)$ we obtain

$$\Big(y_3(h_3(h_2(h_1(t))))\Big)^{\alpha_1\alpha_2\alpha_3} \geqslant (y_3(t))^{\alpha_1\alpha_2\alpha_3} \qquad \text{for } t \geqslant t_6.$$

Therefore (3.8) may be rewritten as

$$\frac{y_3'(t)}{(y_3(t))^{\alpha_1\alpha_2\alpha_3}} \leqslant -p_3(t)\Big[\int\limits_{t_5}^{h_1(t)} p_1(u)\Big(\int\limits_{t_4}^{h_2(u)} p_2(s)\,ds\Big)^{\alpha_1} du\Big]^{\alpha_3}, \quad t \geqslant t_6. \tag{3.9}$$

Integrating (3.9) from t_6 to t and using the substitution $x = y_3(w)$ from (3.9) we get

$$\lim_{t\to\infty}\int\limits_{y_3(t_6)}^{y_3(t)} \frac{dx}{x^{\alpha_1\alpha_2\alpha_3}} \leqslant -\int\limits_{t_6}^{\infty} p_3(v)\Big[\int\limits_{t_5}^{h_1(v)} p_1(u)\Big(\int\limits_{t_4}^{h_2(u)} p_2(s)\,ds\Big)^{\alpha_1} du\Big]^{\alpha_3} dv. \tag{3.10}$$

We know that $y_3(t)$ is a decreasing function and $y_3(t) > 0$. Thus $\lim\limits_{t\to\infty} y_3(t) = K_1 \geqslant 0$ and in view of (A2) we obtain $\lim\limits_{t\to\infty}\int\limits_{y_3(t_6)}^{y_3(t)} \frac{dx}{x^{\alpha_1\alpha_2\alpha_3}} = K_2$, where K_2 is a finite real number. This fact contradicts the assumption (A3).

I.2.b Let $y_2(t) < 0, t \geqslant t_4 \geqslant t_3$. In regard of (c) there exists a $t_5 \geqslant t_4$ such that $y_2(h_2(t)) < 0$, for $t \geqslant t_5$. The assumptions (d), (b) and the first equation of (1.1) imply that $z(t)$ is a decreasing function for $t \geqslant t_5$. On the interval $[t_5, \infty)$ hold:

- $y_1(t) > 0$;
- $z(t)$ is a decreasing function and $z(t) > 0$;
- $y_2(t)$ is an increasing function and $y_2(t) < 0$;
- $y_3(t)$ is a decreasing function and $y_3(t) > 0$.

Therefore exist $\lim\limits_{t\to\infty} y_3(t) = A \geqslant 0$, $\lim\limits_{t\to\infty} y_2(t) = B \leqslant 0$ and $\lim\limits_{t\to\infty} z(t) = C \geqslant 0$. We shall show that $A = 0, B = 0$ and $C = 0$.

(i) Let $A > 0$. Then $y_3(t) \geqslant A$ for $t \geqslant T_0 \geqslant t_5$. In view of (c) and raising to the power of α_2 we have $(y_3(h_3(t)))^{\alpha_2} \geqslant A^{\alpha_2}$ for $t \geqslant T_1 \geqslant T_0$. Integrating the second equation of (1.1) from T_1 to t and using the last inequality we get

$$y_2(t) - y_2(T_1) \geqslant A^{\alpha_2} \int_{T_1}^{t} p_2(s)\, ds, \quad t \geqslant T_1. \tag{3.11}$$

(3.11) and (b) imply that $\lim_{t\to\infty} y_2(t) = \infty$. Therefore $y_2(t) > 0$ for $t \geqslant T_2 \geqslant T_1$, which contradicts $y_2(t) < 0$ for $t \geqslant t_5$. Then $\lim_{t\to\infty} y_3(t) = 0$.

(ii) Assume that $B < 0$. Then $y_2(t) \leqslant B$ for $t \geqslant T_0 \geqslant t_5$ and in regard of (c) we have $y_2(h_2(t)) \leqslant B$ for $t \geqslant T_1 \geqslant T_0$. Hence $\mid y_2(h_2(t)) \mid = -y_2(h_2(t)) \geqslant K_1, K_1 = -B, t \geqslant T_1$. Raising this inequality to the power of α_1, multiplying by $-p_1(t)$ and using the first equation of (1.1) we obtain

$$z'(t) \leqslant -K_1^{\alpha_1}\, p_1(t), \quad t \geqslant T_1.$$

Integrating the last inequality from T_1 to t and in view of (b) we get $\lim_{t\to\infty} z(t) = -\infty$. Therefore $z(t) < 0$ for $t \geqslant T_2 \geqslant T_1$ which is a contradiction with positivity of $z(t)$ for $t \geqslant t_5$.

(iii) Let $C > 0$. Then $z(t) \geqslant C$ for $t \geqslant T_0 \geqslant t_5$. Taking into account the definition of $z(t)$ we are led to $y_1(t) \geqslant z(t) \geqslant C$ for $t \geqslant T_0$. In view of (c) we have $y_1(h_1(t)) \geqslant C$ for $t \geqslant T_1 \geqslant T_0$ and the third equation of (1.1) implies

$$y_3'(t) \leqslant -C^{\alpha_3}\, p_3(t), \quad t \geqslant T_1.$$

Integrating the last inequality from T_1 to t and multiplying by (-1) we obtain

$$y_3(T_1) \geqslant y_3(T_1) - y_3(t) \geqslant C^{\alpha_3} \int_{T_1}^{t} p_3(s)\, ds, \quad t \geqslant T_1.$$

Hence for $t \to \infty$ we get

$$y_3(T_1) \geqslant C^{\alpha_3} \int_{T_1}^{\infty} p_3(s)\, ds. \tag{3.12}$$

In view of (c) there exists a $T_2 \geqslant T_1$ such that $h_3(t) \geqslant T_1$ for $t \geqslant T_2$. Then (3.12) holds for $h_3(t)$, $t \geqslant T_2$, too:

$$y_3(h_3(t)) \geqslant C^{\alpha_3} \int_{h_3(t)}^{\infty} p_3(s)\, ds, \quad t \geqslant T_2 \geqslant T_1.$$

Using the second equation of (1.1) we have

$$y_2'(t) \geqslant C^{\alpha_2\,\alpha_3} p_2(t)\Big(\int\limits_{h_3(t)}^{\infty} p_3(s)\,ds\Big)^{\alpha_2}, \quad t \geqslant T_2. \tag{3.13}$$

Integrating (3.13) from T_2 to t and in regard of (A4) we obtain $\lim\limits_{t\to\infty} y_2(t) = \infty$. Hence $y_2(t) > 0$ pre $t \geqslant T_3 \geqslant T_2$ which is a contradiction with $y_2(t) < 0$ for $t \geqslant t_5$. Therefore $\lim\limits_{t\to\infty} z(t) = 0$ and from Lemma 2.3 we obtain that $\lim\limits_{t\to\infty} y_1(t) = 0$.

II. Let $y_1(t) \in N^-$, $t \geqslant t_0$. Then $z(t) < 0, t \geqslant t_0$. Using the assumptions (c), (d) and (b), the third equation of (1.1) implies that $y_3(t)$ is a decreasing function for $t \geqslant t_1 \geqslant t_0$.

II.1 Assume that $y_3(t) < 0,\ t \geqslant t_2 \geqslant t_1$. Then we can proceed the same way as in the case I.1 to get $\lim\limits_{t\to\infty} z(t) = -\infty$ which is contrary to Lemma 2.2.

II.2 Let $y_3(t) > 0$ for $t \geqslant t_2 \geqslant t_1$. In view of (c) there exists a $t_3 \geqslant t_2$ such that $y_3(h_3(t)) > 0$ for $t \geqslant t_3$. The assumptions (d),(b) and the second equation of (1.1) imply that $y_2(t)$ is an increasing function for $t \geqslant t_3$.

II.2.a Let $y_2(t) > 0$ for $t \geqslant t_4 \geqslant t_3$. In regard of (c) and monotonicity of $y_2(t)$ holds: $y_2(h_2(t)) \geqslant y_2(t_4)$ for $t \geqslant t_5 \geqslant t_4$. Raising this inequality to the power of α_1, multiplying by $p_1(t)$ and using the first equation of (1.1) we get $z'(t) \geqslant M^{\alpha_1} p_1(t)$ where $M = y_2(t_4),\ t \geqslant t_5$. Integrating this inequality from t_5 to t we obtain

$$z(t) - z(t_5) \geqslant M^{\alpha_1} \int\limits_{t_5}^{t} p_1(s)ds, \quad t \geqslant t_5.$$

Hence $\lim\limits_{t\to\infty} z(t) = \infty$ which is a contradiction with Lemma 2.2.

II.2.b Let $y_2(t) < 0$ for $t \geqslant t_4 \geqslant t_3$. In view of assumptions (c), (d), (b) and first equation of (1.1) we get that $z(t)$ is a decreasing function for $t \geqslant t_5 \geqslant t_4$. Therefore $\lim\limits_{t\to\infty} z(t) = A < 0$ which contradicts the Lemma 2.2. □

Theorem 3.2. *Let $\sigma = -1$ and assume that (A1) and (A4) hold. Moreover, let*

(A5) $\alpha_1\,\alpha_2\,\alpha_3 = 1$;

(A6)

$$\int^{\infty} p_3(t)\Big[\int\limits_0^{h_1(t)} p_1(u)\Big(\int\limits_0^{h_2(u)} p_2(s)ds\Big)^{\alpha_1} du\Big]^{(1-\epsilon)\alpha_3} dt = \infty, \quad 0 < \epsilon < 1.$$

Then every proper solution $y \in W$ of (1.1) *is either oscillatory or* $y_i(t)$, *i=1,2,3 tend monotonically to zero as* $t \to \infty$.

Proof. Assume that $y(t) \in W$ is a nonoscillatory solution of (1.1) and $y_1(t) > 0$ for $t \geqslant t_0$. We can proceed exactly as in the proof of Theorem 3.1. We shall discuss only the possibility I.2.a. The proofs of cases I.1, I.2.b and II. are the same.

I. Let $y_1(t) \in N^+, t \geqslant t_0$. Then $z(t) > 0, t \geqslant t_0$ and the third equation of (1.1) implies that $y_3(t)$ is a decreasing function for $t \geqslant t_1 \geqslant t_0$.

I.2 Assume that $y_3(t) > 0$ for $t \geqslant t_2 \geqslant t_1$. The assumptions (c), (d), (b) and the second equation of (1.1) imply that $y_2(t)$ is an increasing function for $t \geqslant t_3$.

I.2.a Let $y_2(t) > 0$ for $t \geqslant t_4 \geqslant t_3$. Then we can proceed the same way as for the case I.2.a of Theorem 3.1 to get (3.7):

$$\Big(y_1(h_1(t))\Big)^{\alpha_3} \geqslant \Big(y_3(h_3(h_2(h_1(t))))\Big)^{\alpha_1\alpha_2\alpha_3} \Big[\int\limits_{t_5}^{h_1(t)} p_1(u) \Big(\int\limits_{t_4}^{h_2(u)} p_2(s)\, ds\Big)^{\alpha_1} du\Big]^{\alpha_3}$$

for $t \geqslant t_6$. In view of monotonicity of $y_3(t)$, assumptions (A1), (A5) and raising to the power of $1-\epsilon$ we are led to

$$(y_1(h_1(t)))^{(1-\epsilon)\,\alpha_3} \geqslant (y_3(t))^{1-\epsilon} \Big[\int\limits_{t_5}^{h_1(t)} p_1(u) \Big(\int\limits_{t_4}^{h_2(u)} p_2(s)\, ds\Big)^{\alpha_1} du\Big]^{(1-\epsilon)\,\alpha_3} \tag{3.14}$$

for $t \geqslant t_6$. The property $y_2(t) > 0, t \geqslant t_4$ and the first equation of (1.1) imply that $z(t)$ is an increasing function for all sufficiently large t. From the proof of Theorem 3.1 we know that $h_1(t) \geqslant t_5$ for $t \geqslant t_6$. Therefore $z(h_1(t)) \geqslant z(t_5)$ for $t \geqslant t_6$ and from $y_1(t) \geqslant z(t)$, $t \geqslant t_0$ we get $y_1(h_1(t)) \geqslant z(t_5),\ t \geqslant t_6$. Hence

$$1 \geqslant \frac{K_1}{(y_1(h_1(t)))^{\alpha_3}}, \quad K_1 = (z(t_5))^{\alpha_3} > 0,\ t \geqslant t_6.$$

Raising to the power of ϵ and multiplying by $(y_1(h_1(t)))^{\alpha_3}$ may be the last inequality rewritten as

$$(y_1(h_1(t)))^{(1-\epsilon)\,\alpha_3} \leqslant K_2\,(y_1(h_1(t)))^{\alpha_3},\ \text{ kde }\ K_2 = K_1^{-\epsilon},\ t \geqslant t_6.$$

Combining this inequality and (3.14), multiplying by $-p_3(t)$ and using the third equation of (1.1) we obtain

$$K_2\,(y_3(t))^{\epsilon-1}\, y_3'(t) \leqslant -p_3(t) \Big[\int\limits_{t_5}^{h_1(t)} p_1(u) \Big(\int\limits_{t_4}^{h_2(u)} p_2(s)\, ds\Big)^{\alpha_1} du\Big]^{(1-\epsilon)\,\alpha_3},\ t \geqslant t_6. \tag{3.15}$$

Integrating (3.15) from t_6 to t we have

$$\frac{K_2}{\epsilon}\Big[(y_3(t))^\epsilon - (y_3(t_6))^\epsilon\Big] \leqslant -\int_{t_6}^{t} p_3(x)\Big[\int_{t_5}^{h_1(x)} p_1(u)\Big(\int_{t_4}^{h_2(u)} p_2(s)\,ds\Big)^{\alpha_1} du\Big]^{(1-\epsilon)\,\alpha_3} dx$$

for $t \geqslant t_6$.

The last inequality and the assumption (A6) imply that $\lim_{t\to\infty}(y_3(t))^\epsilon = -\infty$. But $(y_3(t))^\epsilon$ is a decreasing function and $(y_3(t))^\epsilon \geqslant 0$. Therefore $\lim_{t\to\infty}(y_3(t))^\epsilon = A \geqslant 0$ and this is a contradiction with $\lim_{t\to\infty}(y_3(t))^\epsilon = -\infty$. □

Theorem 3.3. *Assume that $\sigma = 1$ and the assumptions (A3), (A4) of Theorem 3.1 are fulfilled. Then every proper solution $y \in W$ of (1.1) is either oscillatory or $|\, y_i(t)\,|$, $i = 1,2,3$ tend monotonically to infinity as $t \to \infty$ or $y_i(t)$, i=1,2,3 tend monotonically to zero as $t \to \infty$.*

Proof. Let $y(t) \in W$ be a nonoscillatory solution of (1.1). According to Lemma 2.1 there exists a $t_0 \geqslant 0$ such that $z(t)$, $y_2(t)$, $y_3(t)$ are monotone functions of constant sign on the interval $[t_0, \infty)$. Without loss of generality we may assume that $y_1(t) > 0$ for $t \geqslant t_0$. Then either $y_1(t) \in N^+$ or $y_1(t) \in N^-$ for $t \geqslant t_0$.

I. Let $y_1(t) \in N^+, t \geqslant t_0$. Therefore $z(t) > 0$ for $t \geqslant t_0$. Using the assumptions (c), (d) and (b), the system (1.1) implies that the following four cases may occur:

I.1	$y_1(t) > 0$	$y_2(t)$ is increasing and $y_2(t) > 0$	$y_3(t)$ is increasing and $y_3(t) > 0$	$z(t)$ is increasing and $z(t) > 0$
I.2	$y_1(t) > 0$	$y_2(t)$ is increasing and $y_2(t) < 0$	$y_3(t)$ is increasing and $y_3(t) > 0$	$z(t)$ is decreasing and $z(t) > 0$
I.3	$y_1(t) > 0$	$y_2(t)$ is decreasing and $y_2(t) > 0$	$y_3(t)$ is increasing and $y_3(t) < 0$	$z(t)$ is increasing and $z(t) > 0$
I.4	$y_1(t) > 0$	$y_2(t)$ is decreasing and $y_2(t) < 0$	$y_3(t)$ is increasing and $y_3(t) < 0$	$z(t)$ is decreasing and $z(t) > 0$

I.1 In view of (c) and monotonicity of $y_3(t)$ we get $y_3(h_3(t)) \geqslant y_3(t_5)$ for $t \geqslant t_6 \geqslant t_5$. Raising this inequality to the power of α_2, multiplying by $p_2(t)$ and using the second equation of (1.1) we have:

$$y_2'(t) \geqslant L_1^{\alpha_2} p_2(t),\ L_1 = y_3(t_5),\ t \geqslant t_6.$$

Integrating the last equation from t_6 to t we obtain

$$y_2(t) \geqslant y_2(t) - y_2(t_6) \geqslant L_1^{\alpha_2}\int_{t_6}^{t} p_2(s)ds, \quad t \geqslant t_6. \tag{3.16}$$

Hence $\lim\limits_{t\to\infty} y_2(t) = \infty$, i.e. $\lim\limits_{t\to\infty} |y_2(t)| = \infty$.

In regard of (c) and monotonicity of $y_2(t)$ we are led to $y_2(h_2(t)) \geqslant y_2(t_5)$, $t \geqslant t_6 \geqslant t_5$. Raising this inequality to the power of α_1, multiplying by $p_1(t)$ and using the first equation of (1.1) we get:

$$z'(t) \geqslant L_2^{\alpha_1} p_1(t),\ t \geqslant t_6,\ L_2 = y_2(t_5).$$

Integrating the last inequality from t_6 to t and using $y_1(t) \geqslant z(t)$ for $t \geqslant t_0$ we have:

$$y_1(t) \geqslant L_2^{\alpha_1} \int\limits_{t_6}^{t} p_1(s)ds, \qquad t \geqslant t_6.$$

Therefore $\lim\limits_{t\to\infty} y_1(t) = \infty$ and $\lim\limits_{t\to\infty} |y_1(t)| = \infty$.

In view of (c) there exists a $t_7 \geqslant t_6$ such that $h_2(t) \geqslant t_6$ for $t \geqslant t_7$. Then (3.16) holds for $h_2(t), t \geqslant t_7$, too:

$$y_2(h_2(t)) \geqslant L_1^{\alpha_2} \int\limits_{t_6}^{h_2(t)} p_2(s)ds, \quad t \geqslant t_7.$$

Hence we have

$$z'(t) = p_1(t)\Big(y_2(h_2(t))\Big)^{\alpha_1} \geqslant L_3\, p_1(t)\Big(\int\limits_{t_6}^{h_2(t)} p_2(s)ds\Big)^{\alpha_1},\ L_3 = L_1^{\alpha_1\,\alpha_2},\ t \geqslant t_7.$$

Integrating this inequality from t_7 to t and taking into account $y_1(t) \geqslant z(t)$ we get

$$y_1(t) \geqslant L_3 \int\limits_{t_7}^{t} p_1(u)\Big(\int\limits_{t_6}^{h_2(u)} p_2(s)ds\Big)^{\alpha_1} du,\ t \geqslant t_7. \tag{3.17}$$

In regard of (c) the last inequality holds for $h_1(t)$, $t \geqslant t_8 \geqslant t_7$, too:

$$y_1(h_1(t)) \geqslant L_3 \int\limits_{t_7}^{h_1(t)} p_1(u)\Big(\int\limits_{t_6}^{h_2(u)} p_2(s)ds\Big)^{\alpha_1} du,\ t \geqslant t_8.$$

Hence using the third equation of (1.1) we obtain

$$y_3'(t) \geqslant L_4\, p_3(t)\Big(\int\limits_{t_7}^{h_1(t)} p_1(u)\Big(\int\limits_{t_6}^{h_2(u)} p_2(s)ds\Big)^{\alpha_1} du\Big)^{\alpha_3},\ L_4 = L_3^{\alpha_3},\ t \geqslant t_8. \tag{3.18}$$

Integrating (3.18) from t_8 to t we get

$$y_3(t) \geqslant L_4 \int_{t_8}^{t} p_3(v)\Big(\int_{t_7}^{h_1(v)} p_1(u)\Big(\int_{t_6}^{h_2(u)} p_2(s)ds\Big)^{\alpha_1} du\Big)^{\alpha_3} dv, \;\; t \geqslant t_8.$$

In view of (A3) the last inequality implies $\lim\limits_{t\to\infty} y_3(t) = \infty$. Then $\lim\limits_{t\to\infty} |y_3(t)| = \infty$.

I.2 We can proceed the same way as for the case I.1 to get (3.16):

$$y_2(t) \geqslant y_2(t) - y_2(t_6) \geqslant L_1^{\alpha_2} \int_{t_6}^{t} p_2(s)ds, \quad t \geqslant t_6.$$

Therefore $\lim\limits_{t\to\infty} y_2(t) = \infty$, i.e. $y_2(t) > 0$ for $t \geqslant t_7 \geqslant t_6$. But this is a contradiction with $y_2(t) < 0$ for $t \geqslant t_5$.

I.3 Using (c), monotonicity of $z(t)$ and $y_1(t) \geqslant z(t)$ we have: $y_1(h_1(t)) \geqslant L_5$, $L_5 = z(t_5)$, $t \geqslant t_6 \geqslant t_5$. Then the third equation of (1.1) may be rewritten as $y_3'(t) \geqslant L_5^{\alpha_3} p_3(t)$, $t \geqslant t_6$. Integrating this inequality from t_6 to t we obtain:

$$-y_3(t_6) \geqslant y_3(t) - y_3(t_6) \geqslant L_5^{\alpha_3} \int_{t_6}^{t} p_3(s)ds, \quad t \geqslant t_6.$$

Hence for $t \to \infty$ we see that

$$-y_3(t_6) \geqslant L_5^{\alpha_3} \int_{t_6}^{\infty} p_3(s)ds.$$

In regard of (c) the last inequality holds for $h_3(t)$, $t \geqslant t_7 \geqslant t_6$, too:

$$-y_3(h_3(t)) = |y_3(h_3(t))| \geqslant L_6 \int_{h_3(t)}^{\infty} p_3(s)ds, \;\; L_6 = L_5^{\alpha_3}, \, t \geqslant t_7.$$

Hence

$$y_2'(t) = -p_2(t)|y_3(h_3(t))|^{\alpha_2} \leqslant -L_6^{\alpha_2} p_2(t)\Big(\int_{h_3(t)}^{\infty} p_3(s)ds\Big)^{\alpha_2}, \quad t \geqslant t_7,$$

and integrating from t_7 to t we are led to

$$y_2(t) - y_2(t_7) \leqslant -L_6^{\alpha_2} \int_{t_7}^{t} p_2(u)\Big(\int_{h_3(u)}^{\infty} p_3(s)ds\Big)^{\alpha_2} du, \quad t \geqslant t_7.$$

Therefore in view of (A4) we get $\lim\limits_{t\to\infty} y_2(t) = -\infty$. It means that $y_2(t) < 0$ for $t \geqslant t_8 \geqslant t_7$ which is contrary to $y_2(t) > 0$ for $t \geqslant t_5$.

I.4 In regard of (c) and monotonicity of $y_2(t)$ we have $|y_2(h_2(t))| \geqslant L_7$, $L_7 = (-y_2(t_5))$, $t \geqslant t_6 \geqslant t_5$. Hence $z'(t) = -p_1(t)|y_2(h_2(t))|^{\alpha_1} \leqslant -L_7^{\alpha_1} p_1(t)$, $t \geqslant t_6$ and integrating from t_6 to t we obtain

$$z(t) - z(t_6) \leqslant -L_7^{\alpha_1} \int\limits_{t_6}^{t} p_1(s)ds, \quad t \geqslant t_6.$$

Using (b) the last inequality imply that $\lim\limits_{t\to\infty} z(t) = -\infty$. Therefore $z(t) < 0$ for $t \geqslant t_7 \geqslant t_6$ which is a contradiction with $z(t) > 0$ for $t \geqslant t_5$.

II. Let $y_1(t) \in N^-$. Hence $z(t) < 0$ for $t \geqslant t_0$ and the third equation of (1.1) implies that $y_3(t)$ is an increasing function for $t \geqslant t_1$.

II.1 Assume that $y_3(t) > 0$, $t \geqslant t_2 \geqslant t_1$. Then $y_3(h_3(t)) > 0$ for $t \geqslant t_3 \geqslant t_2$ and from the second equation of (1.1) we get that $y_2(t)$ is an increasing function for $t \geqslant t_3$.

II.1.a Let $y_2(t) > 0$ for $t \geqslant t_4$. In view of (c) and monotonicity of $y_2(t)$ we have $(y_2(h_2(t)))^{\alpha_1} \geqslant (y_2(t_4))^{\alpha_1}$ for $t \geqslant t_5 \geqslant t_4$. Integrating the first equation of (1.1) from t_5 to t and using the last inequality we are led to

$$z(t) - z(t_5) \geqslant (y_2(t_4))^{\alpha_1} \int\limits_{t_5}^{t} p_1(s)ds, \quad t \geqslant t_5.$$

Hence in view of (b) we get $\lim\limits_{t\to\infty} z(t) = \infty$ which contradicts Lemma 2.2.

II.1.b Let $y_2(t) < 0, t \geqslant t_4$. Taking into account assumptions (b), (c), (d) the first equation of (1.1) implies that $z(t)$ is a decreasing function for $t \geqslant t_5$. It means that $\lim\limits_{t\to\infty} z(t) = A < 0$ which is contrary to Lemma 2.2.

II.2 Assume that $y_3(t) < 0, t \geqslant t_2 \geqslant t_1$. From the second equation of (1.1) we get that $y_2(t)$ is a decreasing function for $t \geqslant t_3$.

Function $y_3(t)$ is increasing. Therefore exists $\lim\limits_{t\to\infty} y_3(t) = B \leqslant 0$. We shall show that $B = 0$.

Let $B < 0$. Then $y_3(h_3(t)) \leqslant B < 0$ for $t \geqslant t_4 \geqslant t_3$. Hence $|y_3(h_3(t))| \geqslant C$, $C = -B$ and

$$y_2'(t) = -p_2(t)|y_3(h_3(t))|^{\alpha_2} \leqslant -C^{\alpha_2} p_2(t), \quad t \geqslant t_4.$$

Integrating the last inequality from t_4 to t and using (b) we obtain $\lim\limits_{t\to\infty} y_2(t) = -\infty$, i.e. $y_2(t) < 0$, $t \geqslant t_5 \geqslant t_4$. In regard of assumptions (b), (c) and (d) the first

equation of (1.1) implies that $z(t)$ is a decreasing function for $t \geqslant t_6$. Therefore $\lim_{t\to\infty} z(t) = D < 0$ which is a contradiction with Lemma 2.2. Then $\lim_{t\to\infty} y_3(t) = 0$.

II.2.a Let $y_2(t) < 0$, $t \geqslant t_4$. From the first equation of (1.1) we have that $z(t)$ is a decreasing function. Therefore $\lim_{t\to\infty} z(t) = E < 0$ which contradicts Lemma 2.2.

II.2.b If $y_2(t) > 0$, $t \geqslant t_4 \geqslant t_3$, then exists $\lim_{t\to\infty} y_2(t) = F \geqslant 0$. We shall show that $F = 0$.

Assume that $F > 0$. Then $y_2(h_2(t)) > F$, $t \geqslant t_5 \geqslant t_4$ and hence

$$z'(t) = p_1(t)(y_2(h_2(t)))^{\alpha_1} > F^{\alpha_1} p_1(t), \qquad t \geqslant t_5.$$

Integrating the last inequality from t_5 to t and using (b) we obtain $\lim_{t\to\infty} z(t) = \infty$. Therefore $z(t) > 0$ for $t \geqslant t_6 \geqslant t_5$ which is a contradiction with $z(t) < 0$. Then $\lim_{t\to\infty} y_2(t) = 0$.

Because $y_2(t) > 0$, the first equation of (1.1) implies that $z(t)$ is an increasing function such that $z(t) < 0$. In regard of Lemma 2.2 we obtain $\lim_{t\to\infty} z(t) = 0$ and $\lim_{t\to\infty} y_1(t) = 0$. □

References

[1] Györi, I., Ladas, G., Oscillation Theory Of Delay Differential Equations, Clarendon Press, Oxford, 1991.

[2] Marušiak, P., Oscillatory properties of functional differential systems of neutral type, *Czechoslovak Math. J.,* **43** (118), (1993), 649–662.

[3] Mihalíková, B., A note on the asymptotic properties of systems of neutral differential equations, *Stud. Univ. Zilina, Math. Phys. Ser.,* Vol. 13, (2001), 133–139.

[4] Mihalíková, B., Oscillations of neutral differential systems, *Discuss. Math. Differential Incl.,* Vol. 19, (1999), 5–15.

[5] Mihalíková, B., Some properties of neutral differential systems equations, *Bollettino U. M. I.,* Vol. 8 (5-B), (2002), 279–287.

[6] Mihály, T., On the oscillatory and asymptotic properties of solutions of systems of neutral differential equations, *Nonlinear Analysis: Theory, Methods & Applications* (to appear)

[7] Shevelo, V. N., Varech, N. V., Gritsai, A. G., Oscillatory properties of solutions of systems of differential equations with deviating arguments, *preprint no. 85.10,* Institute of Mathematics of the Ukrainian Academy of Sciences Russian, (1985).

[8] Špániková, E., Asymptotic properties of solutions of nonlinear differential systems with deviating argument, Doctoral Thesis, University of Žilina, Žilina, Slovakia, (1990).

[9] Špániková, E., Oscillatory properties of solutions of three-dimensional differential systems of neutral type, *Czechoslovak Math. J.,* **50** (125), (2000), 879–887.

Tomáš Mihály
Department of Mathematical Analysis and Applied Mathematics
Faculty of Science
University of Žilina
Hurbanova 15
010 26 Žilina
Slovakia

Annales Mathematicae et Informaticae
***33** (2006) pp. 109–123*
http://www.ektf.hu/tanszek/matematika/ami

Lebesgue constants in polynomial interpolation

Simon J. Smith

La Trobe University, Bendigo, Australia
e-mail: s.smith@latrobe.edu.au

Submitted 15 September 2005; Accepted 2 June 2006

Abstract

Lagrange interpolation is a classical method for approximating a continuous function by a polynomial that agrees with the function at a number of chosen points (the "nodes"). However, the accuracy of the approximation is greatly influenced by the location of these nodes. Now, a useful way to measure a given set of nodes to determine whether its Lagrange polynomials are likely to provide good approximations is by means of the Lebesgue constant. In this paper a brief survey of methods and results for the calculation of Lebesgue constants for some particular node systems is presented. These ideas are then discussed in the context of Hermite–Fejér interpolation and a weighted interpolation method where the nodes are zeros of Chebyshev polynomials of the second kind.

Keywords: interpolation, Lagrange interpolation, Hermite–Fejér interpolation, Lebesgue constant, Lebesgue function

MSC: 41-02, 41A05, 41A10

1. Introduction

For each integer $n \geqslant 1$, consider n points (*nodes*) $x_{k,n}$ $(k = 1, 2, \ldots, n)$ in $[-1, 1]$ with

$$-1 \leqslant x_{n,n} < x_{n-1,n} < \ldots < x_{2,n} < x_{1,n} \leqslant 1, \tag{1.1}$$

and let X be the infinite triangular matrix

$$X = \{x_{k,n} : k = 1, 2, \ldots, n;\ n = 1, 2, 3, \ldots\}. \tag{1.2}$$

Given $f \in C[-1,1]$, the classical Lagrange interpolation polynomial $L_{n-1}(X,f)$ of degree $n-1$ (or less) for f, based on X, can be written as

$$L_{n-1}(X,f)(x) = \sum_{i=1}^{n} f(x_{i,n})\ell_{i,n}(X,x),$$

where the *fundamental polynomial* $\ell_{i,n}(X,x)$ is the unique polynomial of degree $n-1$ with $\ell_{i,n}(X,x_{k,n}) = \delta_{i,k}$, $1 \leqslant k \leqslant n$. (Here $\delta_{i,k}$ denotes the Kronecker delta.)

Let $\|f\|$ denote the uniform norm

$$\|f\| = \max_{-1\leqslant x\leqslant 1} |f(x)|.$$

When studying the uniform convergence behaviour of the $L_{n-1}(X,f)$ as $n \to \infty$, a crucial role is played by the *Lebesgue function*

$$\lambda_n(X,x) = \max_{\|f\|\leqslant 1} |L_{n-1}(X,f)(x)| = \sum_{i=1}^{n} |\ell_{i,n}(X,x)|$$

and the *Lebesgue constant*

$$\Lambda_n(X) = \max_{\|f\|\leqslant 1} \|L_{n-1}(X,f)\| = \max_{-1\leqslant x\leqslant 1} \lambda_n(X,x)$$

(see, for example, Rivlin [13, Chapter 4] or Szabados and Vértesi [19]).

Now, it is known that for *any* X, $\Lambda_n(X)$ is unbounded with respect to n. A consequence of this is (by the uniform boundedness theorem) Faber's 1914 result [6] that there exists $f \in C[-1,1]$ such that $L_n(X,f)$ does *not* converge uniformly to f. However, if f is not too badly behaved (as measured by the modulus of continuity, for instance) and the $\Lambda_n(X)$ are not too large, then uniform convergence *is* achieved (see, for example, Rivlin [13, Chapter 4]).

Figure 1 illustrates some basic properties of Lebesgue functions for Lagrange interpolation. For example, for any X and $n \geqslant 3$, $\lambda_n(X,x)$ is a piecewise polynomial that satisfies $\lambda_n(X,x) \geqslant 1$ with equality if and only if x is one of the nodes $x_{k,n}$. As well, on each interval $(x_{k+1,n}, x_{k,n})$ for $1 \leqslant k \leqslant n-1$, $\lambda_n(X,x)$ has precisely one local maximum, while $\lambda_n(X,x)$ is decreasing and concave upward on $(-1, x_{n,n})$ and is increasing and concave upward on $(x_{1,n}, 1)$. (For a discussion of these and other properties see, for example, Luttmann and Rivlin [11].)

2. The Lebesgue function for specific node systems

For some particular node systems, the Lebesgue function and constant have been studied in considerable detail. In this section, a summary of some of these results is given — for a more detailed account of many of the results, see the comprehensive survey paper by Brutman [4] and the references therein.

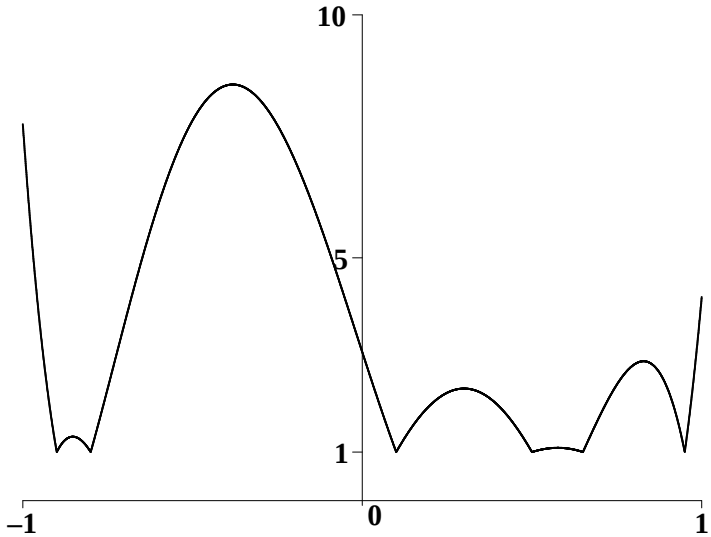

Figure 1: Lebesgue function for Lagrange interpolation based on the six nodes −0.9, −0.8, 0.1, 0.5, 0.65 and 0.95.

2.1. Equally-spaced nodes

Figure 2 illustrates a typical Lebesgue function for Lagrange interpolation based on the equally-spaced nodes

$$E = \{x_{k,n} = 1 - 2(k-1)/(n-1) : k = 1, 2, \ldots, n;\ n = 1, 2, 3, \ldots\}\,.$$

As suggested by the graph, the local maxima of $\lambda_n(E, x)$ are strictly decreasing from the outside towards the middle of the interval $[-1, 1]$, a result that was established by Tietze [20]. Later Turetskii [21] showed that the Lebesgue constant $\Lambda_n(E)$ has the asymptotic expansion as $n \to \infty$,

$$\Lambda_n(E) \sim \frac{2^n}{en \log n}. \tag{2.1}$$

This result has been subsequently refined (to a small extent) by other authors.

2.2. Chebyshev nodes

Figure 3 shows a typical Lebesgue function for Lagrange interpolation based on the Chebyshev nodes

$$T = \{x_{k,n} = \cos(2k-1)\pi/(2n) : k = 1, 2, \ldots, n;\ n = 1, 2, 3, \ldots\}\,.$$

(For each n these nodes are the zeros of the nth Chebyshev polynomial of the first kind.) The graph illustrates that the maximum of the Lebesgue function on $[-1, 1]$

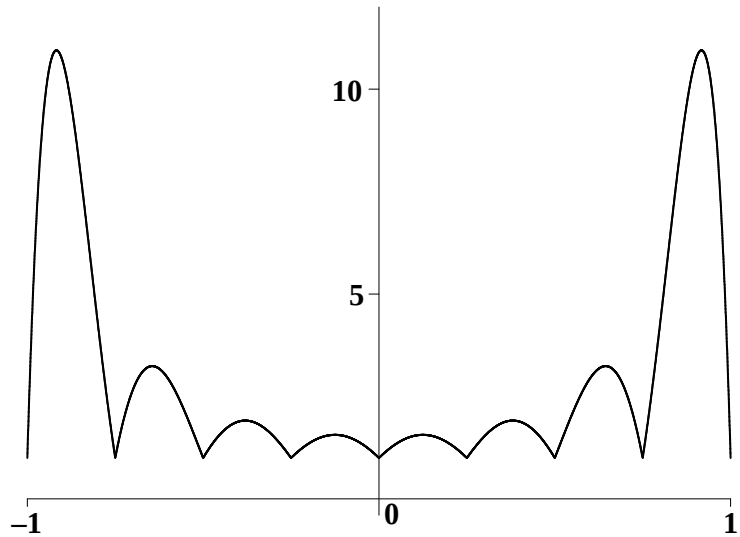

Figure 2: Lebesgue function for Lagrange interpolation on the equally-spaced nodes $x_{k,n} = 1 - 2(k-1)/(n-1)$ [with $n = 9$].

occurs at ± 1, a result due to Ehlich and Zeller [5]. From the representation

$$\Lambda_n(T) = \lambda_n(T, \pm 1) = \frac{1}{n} \sum_{i=1}^{n} \cot(2i-1)\pi/(4n),$$

asymptotic results such as

$$\Lambda_n(T) = \frac{2}{\pi} \log n + \frac{2}{\pi} \left(\gamma + \log \frac{8}{\pi} \right) + O\left(\frac{1}{n^2} \right) \tag{2.2}$$

can be deduced, where γ denotes Euler's constant $0.577\ldots$ (see [4] for references and more precise results). On comparing (2.1) and (2.2) it can be seen that the Lebesgue constant for Chebyshev nodes is *much* smaller than for equally-spaced nodes. This confirms the "bad" status of equally-spaced nodes for Lagrange interpolation, a fact that has become well-known largely because of the example of Runge [15].

Figure 3 also suggests that, as with $\lambda_n(E, x)$, the local maxima of $\lambda_n(T, x)$ are strictly decreasing from the outside towards the middle of the interval $[-1, 1]$. This was proved by Brutman [3] (see also Günttner [8]).

2.3. Extended Chebyshev nodes

The *extended Chebyshev nodes* $\widehat{T}$ are defined by

$$\widehat{T} = \{x_{k,n} = \cos[(2k-1)\pi/(2n)]/\cos[\pi/(2n)] : k = 1, 2, \ldots, n;\ n = 2, 3, 4, \ldots\}.$$

That is, they are obtained by rescaling the Chebyshev nodes so that the nodes of

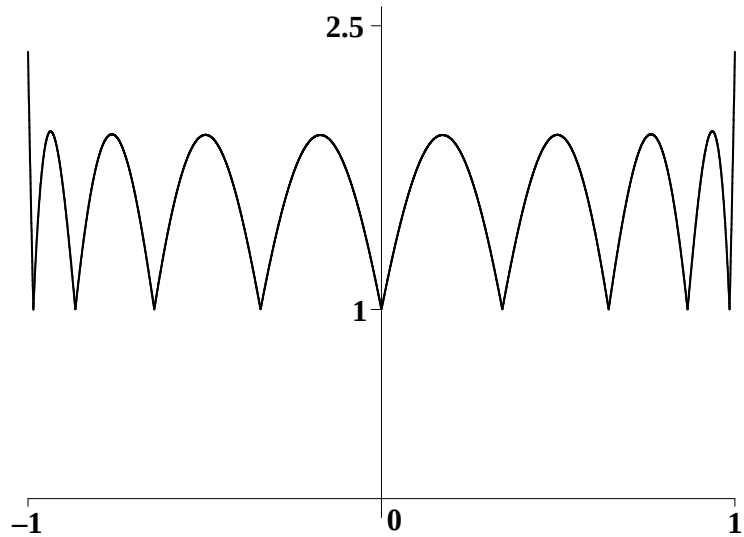

Figure 3: Lebesgue function for Lagrange interpolation on the Chebyshev nodes $x_{k,n} = \cos(2k-1)\pi/(2n)$ [with $n = 9$].

greatest magnitude for each n are at ± 1. Now, it is readily shown that

$$\lambda_n(\widehat{T}, x) = \lambda_n(T, x\cos[\pi/(2n)]).$$

Thus, by the monotonicity result for the local maxima of $\lambda_n(T, x)$, it follows that $\Lambda_n(\widehat{T})$ is strictly less than $\Lambda_n(T)$ and is equal to the maximum of $\lambda_n(T, x)$ on the interval $(\cos 3\pi/(2n), \cos \pi/(2n))$. This characterisation was used by Günttner [9] to obtain an asymptotic result for $\Lambda_n(\widehat{T})$, a simplified version of which is

$$\Lambda_n(\widehat{T}) = \frac{2}{\pi}\log n + \frac{2}{\pi}\left(\gamma + \log\frac{8}{\pi} - \frac{2}{3}\right) + O\left(\frac{1}{\log n}\right). \tag{2.3}$$

2.4. Augmented Chebyshev nodes

Another modification of T is to add ± 1 to each row of the matrix. These *augmented Chebyshev nodes* T_a are given by $x_{1,n+2} = 1$, $x_{n+2,n+2} = -1$ and $x_{k,n+2} = \cos(2k-3)\pi/(2n)$ for $k = 2, 3, \ldots, n+1$.

Now, interpolation polynomials on T and T_a are related by

$$\begin{aligned} L_{n+1}(T_a, f)(x) &= L_{n-1}(T, f)(x) + \\ &\quad T_n(x) \times \{(1+x)[f(1) - L_{n-1}(T, f)(1)] \\ &\quad + (-1)^n(1-x)[f(-1) - L_{n-1}(T, f)(-1)]\}/2 \end{aligned} \tag{2.4}$$

where $T_n(x) = \cos(n \arccos x), -1 \leqslant x \leqslant 1$, is the nth Chebyshev polynomial of the first kind. (To verify (2.4), it is a simple matter to check that the RHS is a polynomial of degree no more than $n+1$ which agrees with f at the nodes $x_{k,n+2}$ for $1 \leqslant k \leqslant n+2$.) Thus if $L_n(T, f) \to f$ uniformly on $[-1, 1]$, then $L_n(T_a, f) \to f$ uniformly on $[-1, 1]$.

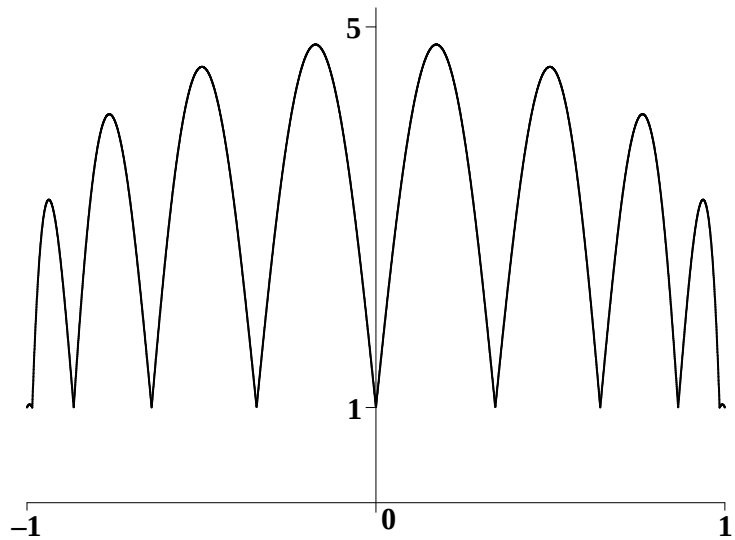

Figure 4: Lebesgue function for Lagrange interpolation on the augmented Chebyshev nodes $\{\cos(2k-3)\pi/(2n) : 2 \leqslant k \leqslant n+1\} \cup \{\pm 1\}$ [with $n = 9$].

Figure 4 appears to show that the local maximum values of $\lambda_{n+2}(T_a, x)$ *increase* from the outside towards the middle of $[-1, 1]$ (which is the reverse of the situation for T). This was proved by Smith [17], who used essentially the method that was employed by Brutman in [3] to establish the monotonic behaviour of the local maxima of $\lambda_n(T, x)$. Smith also obtained the asymptotic result

$$\Lambda_{n+2}(T_a) = \frac{4}{\pi} \log n + \frac{4}{\pi} \left(\gamma + \log \frac{4}{\pi} \right) + 1 + O\left(\frac{1}{n^2} \right) \tag{2.5}$$

which, when compared with (2.2), shows that $\Lambda_{n+2}(T_a)$ is effectively double $\Lambda_n(T)$.

2.5. Optimal nodes

The topic of the *optimal nodes* X^* for Lagrange interpolation, defined by

$$\Lambda_n(X^*) = \min_X \Lambda_n(X), \qquad n = 2, 3, 4, \ldots,$$

has been the subject of much research. Although no explicit formulation of X^* is known, Vértesi [22] showed that $\Lambda_n(X^*)$ has the asymptotic expansion

$$\Lambda_n(X^*) = \frac{2}{\pi}\log n + \frac{2}{\pi}\left(\gamma + \log\frac{4}{\pi}\right) + O\left(\left(\frac{\log\log n}{\log n}\right)^2\right). \tag{2.6}$$

A comparison of (2.3) and (2.6) suggests that $\widehat{T}$ is close to optimal. This point is discussed at some length (and made more precise) in Brutman [4, Section 3].

3. Hermite–Fejér interpolation

Given $f \in C[-1,1]$ and X defined by (1.2), the *Hermite–Fejér interpolation polynomial* $H_{2n-1}(X,f)$ of degree $2n-1$ (or less) for f, based on X, is the unique polynomial of degree no greater than $2n-1$ which interpolates f and has zero derivative at the nodes $x_{k,n}$ for $k = 1, 2, \ldots, n$. It can be written as

$$H_{2n-1}(X,f)(x) = \sum_{i=1}^{n} f(x_{i,n}) A_{i,n}(X,x), \tag{3.1}$$

where the fundamental polynomial $A_{i,n}(X,x)$ is the unique polynomial of degree no greater than $2n-1$ such that $A_{i,n}(X,x_{k,n}) = \delta_{i,k}$ and $A'_{i,n}(X,x_{k,n}) = 0$ for $k = 1, 2, \ldots, n$.

The Lebesgue function for Hermite–Fejér interpolation on X is

$$\lambda_{1,n}(X,x) = \max_{\|f\|\leqslant 1} |H_{2n-1}(X,f)(x)| = \sum_{i=1}^{n} |A_{i,n}(X,x)|$$

and the Lebesgue constant is

$$\Lambda_{1,n}(X) = \max_{\|f\|\leqslant 1} \|H_{2n-1}(X,f)\| = \max_{-1\leqslant x\leqslant 1} \lambda_{1,n}(X,x).$$

For future reference, note that $H_{2n-1}(X,1)(x) = 1$ (from uniqueness considerations), so by (3.1),

$$\sum_{i=1}^{n} A_{i,n}(X,x) = 1. \tag{3.2}$$

3.1. Chebyshev nodes

Interest in Hermite–Fejér interpolation was sparked by Fejér's famous 1916 result (see [7]) that if $f \in C[-1,1]$, then $H_{2n-1}(T,f)$ converges uniformly to f. Thus there is a simple node system for which the Hermite–Fejér method succeeds for all $f \in C[-1,1]$, whereas no such system (simple or otherwise) exists for Lagrange interpolation.

A key point in Fejér's proof is that $A_{i,n}(T,x) \geqslant 0$ for $-1 \leqslant x \leqslant 1$ and $i = 1, 2, \ldots n$. Thus, by (3.2),

$$\lambda_{1,n}(T,x) = \sum_{i=1}^{n} |A_{i,n}(T,x)| = \sum_{i=1}^{n} A_{i,n}(T,x) = 1,$$

and so the Lebesgue constant $\Lambda_{1,n}(T)$ is simply 1.

3.2. A modified Hermite–Fejér method on the augmented Chebyshev nodes

As a "stepping stone" to the study of Hermite–Fejér interpolation on the augmented Chebyshev nodes, consider the following interpolation method.

For $n = 1, 2, 3, \ldots$, write the Chebyshev nodes as

$$t_k = t_{k,n} = \cos(2k-1)\pi/(2n), \qquad k = 1, 2, \ldots, n,$$

and let $t_0 = 1$, $t_{n+1} = -1$. Given $f \in C[-1,1]$, define a polynomial $K_{2n+1}(f)$ of degree $2n+1$ (or less) by

$$\begin{cases} K_{2n+1}(f)(t_k) = f(t_k), & 0 \leqslant k \leqslant n+1, \\ K_{2n+1}(f)'(t_k) = 0, & 1 \leqslant k \leqslant n. \end{cases} \tag{3.3}$$

Thus $K_{2n+1}(f)$ interpolates f on the augmented Chebyshev nodes and has vanishing derivative at the Chebyshev nodes.

An explicit formula for $K_{2n+1}(f)$ in terms of the the Hermite–Fejér interpolation polynomial $H_{2n-1}(T,f)$ is

$$\begin{aligned} K_{2n+1}(f)(x) &= H_{2n-1}(T,f)(x) + \\ &\quad T_n^2(x) \times \{(1+x)[f(1) - H_{2n-1}(T,f)(1)] \\ &\quad + (1-x)[f(-1) - H_{2n-1}(T,f)(-1)]\}/2. \end{aligned} \tag{3.4}$$

(Again, to verify (3.4), it is a simple matter to check that the RHS is a polynomial of degree no more than $2n+1$ that satisfies the conditions (3.3).) From (3.4) it follows immediately by Fejér's result that if $f \in C[-1,1]$, then $K_{2n+1}(f)$ converges uniformly to f.

Now, $K_{2n+1}(f)$ can also be written in terms of fundamental polynomials as

$$K_{2n+1}(f)(x) = \sum_{i=0}^{n+1} f(t_i) B_i(x),$$

where for each $i = 0, 1, \ldots, n+1$, $B_i(x) = B_{i,n}(x)$ is the unique polynomial of degree no greater than $2n+1$ so that $B_i(t_k) = \delta_{i,k}$ for $k = 0, 1, \ldots, n+1$ and $B_i'(t_k) = 0$ for $k = 1, 2, \ldots, n$. The Lebesgue function and constant are respectively

$$\lambda_n(x) = \sum_{i=0}^{n+1} |B_i(x)|, \quad \Lambda_n = \max_{-1 \leqslant x \leqslant 1} \lambda_n(x).$$

By using elementary properties of the Chebyshev polynomials $T_n(x)$ (see, for example, Rivlin [14, Chapter 1]), it is easy to verify that

$$B_0(x) = \frac{1+x}{2}\,T_n^2(x), \quad B_{n+1}(x) = \frac{1-x}{2}\,T_n^2(x) \tag{3.5}$$

and

$$B_k(x) = \frac{(1-x^2)(1+xt_k-2t_k^2)}{n^2(x-t_k)^2(1-t_k^2)}\,T_n^2(x), \quad 1 \leqslant k \leqslant n. \tag{3.6}$$

Observe that for $1 \leqslant k \leqslant n$, the sign of $B_k(x)$ is that of $1+xt_k-2t_k^2$. Thus if $n \geqslant 2$, then $B_1(x)$ (for example) is negative for all values of x in some interval in $[-1,1]$, and so, unlike the Hermite–Fejér method on T, the fundamental polynomials for the modified method are not all non-negative in $[-1,1]$. In terms of the Lebesgue constant, this means that $\Lambda_n > 1$ for all $n \geqslant 2$. On the other hand, since $K_{2n+1}(f)$ converges uniformly to f for all $f \in C[-1,1]$, it follows from the uniform boundedness theorem that the Λ_n are uniformly bounded. In the following theorem, the best possible bound for the Λ_n is derived.

Theorem 3.1. *The Lebesgue constant Λ_n satisfies*

$$\Lambda_n < 3, \qquad n = 1, 2, \ldots \tag{3.7}$$

and

$$\lim_{n\to\infty} \Lambda_n = 3. \tag{3.8}$$

Proof. By (3.5) and (3.6),

$$\lambda_n(x) = T_n^2(x)\left[1 + \frac{(1-x^2)}{n^2}\sum_{k=1}^{n}\frac{|1+xt_k-2t_k^2|}{(x-t_k)^2(1-t_k^2)}\right].$$

Observe that $1+xt_k-2t_k^2 > 0$ if and only if $p(x) < t_k < q(x)$, where

$$p(x) = \frac{x-\sqrt{x^2+8}}{4}, \quad q(x) = \frac{x+\sqrt{x^2+8}}{4}.$$

Let $J_n = \{1, 2, \ldots, n\}$ and to given $x \in [-1,1]$ define

$$\mathcal{R}(x) = \{k \in J_n : p(x) < t_k < q(x)\}, \quad \mathcal{S}(x) = \{k \in J_n : t_k \leqslant p(x) \text{ or } t_k \geqslant q(x)\}.$$

Therefore

$$\lambda_n(x) = T_n^2(x)\left[1 + \frac{(1-x^2)}{n^2}\,F(x)\right], \tag{3.9}$$

where

$$F(x) = \sum_{k\in\mathcal{R}(x)}\frac{1+xt_k-2t_k^2}{(x-t_k)^2(1-t_k^2)} - \sum_{k\in\mathcal{S}(x)}\frac{1+xt_k-2t_k^2}{(x-t_k)^2(1-t_k^2)}.$$

Next employ the partial fraction expansion

$$\frac{1+xt_k-2t_k^2}{(x-t_k)^2(1-t_k^2)} = \frac{x}{(1-x^2)(x-t_k)} + \frac{1}{(x-t_k)^2} - \frac{1/2}{(1-x)(1-t_k)} - \frac{1/2}{(1+x)(1+t_k)}.$$

This leads to

$$\begin{aligned} F(x) = \sum_{k=1}^{n} & \left[\frac{x}{(1-x^2)(x-t_k)} + \frac{1}{(x-t_k)^2} + \frac{1/2}{(1-x)(1-t_k)} + \frac{1/2}{(1+x)(1+t_k)} \right] \\ & - \sum_{k\in\mathcal{R}(x)} \left[\frac{1}{(1-x)(1-t_k)} + \frac{1}{(1+x)(1+t_k)} \right] \\ & - 2 \sum_{k\in\mathcal{S}(x)} \left[\frac{x}{(1-x^2)(x-t_k)} + \frac{1}{(x-t_k)^2} \right]. \end{aligned}$$

Now, from the identity

$$\frac{T_n'(x)}{T_n(x)} = \sum_{k=1}^{n} \frac{1}{x-t_k}$$

and elementary properties of the Chebyshev polynomials (see, for example, Rivlin [14, Chapter 1]), it follows that

$$\sum_{k=1}^{n} \frac{1}{1-t_k} = \sum_{k=1}^{n} \frac{1}{1+t_k} = n^2$$

and

$$\sum_{k=1}^{n} \frac{1}{(x-t_k)^2} = \frac{n^2 - xT_n(x)T_n'(x)}{(1-x^2)T_n^2(x)}.$$

Therefore (3.9) becomes

$$\lambda_n(x) = 1 + 2T_n^2(x) - \frac{2T_n^2(x)}{n^2} \left[\sum_{k\in\mathcal{R}(x)} \frac{1+xt_k}{1-t_k^2} + \sum_{k\in\mathcal{S}(x)} \frac{1-xt_k}{(x-t_k)^2} \right]. \tag{3.10}$$

If $x \in [-1,1]$ the expression in square brackets is positive, so $\lambda_n(x) \leqslant 1 + 2T_n^2(x)$, with equality if and only if $T_n(x) = 0$. In particular, $\lambda_n(x) < 3$, from which (3.7) follows.

To establish (3.8), note that it follows from (3.10) that

$$\lambda_{2n}(0) = 3 - \frac{1}{2n^2} \left[\sum_{k\in\mathcal{R}(0)} \frac{1}{1-t_k^2} + \sum_{k\in\mathcal{S}(0)} \frac{1}{t_k^2} \right], \tag{3.11}$$

where $\mathcal{R}(0) = \{k \in J_{2n} : -1/\sqrt{2} < t_k < 1/\sqrt{2}\}$ and $\mathcal{S}(0) = \{k \in J_{2n} : t_k \leqslant -1/\sqrt{2} \text{ or } t_k \geqslant 1/\sqrt{2}\}$. The sums within the square brackets of (3.11) contain a total of $2n$ terms, each of which is no greater than 2, so

$$\Lambda_{2n} \geqslant \lambda_{2n}(0) \geqslant 3 - \frac{2}{n}.$$

By similar means it can be shown that there exists an absolute constant c so that

$$\Lambda_{2n+1} \geqslant \lambda_{2n+1}\left(\cos\left[n\pi/(2n+1)\right]\right) \geqslant 3 - \frac{c}{2n+1},$$

and hence (3.8) is proved. □

3.3. Hermite–Fejér interpolation on the augmented Chebyshev nodes

If $f \in C[-1,1]$, then by Fejér's result, the Hermite–Fejér interpolation polynomials $H_{2n-1}(T,f)$ converge uniformly to f. In Section 3.2 it was shown that if interpolation conditions at ± 1 are added to the Hermite–Fejér interpolation conditions at the Chebyshev nodes, the resulting interpolation polynomials will still converge uniformly to f. Thus it might be expected that if the full Hermite–Fejér interpolation conditions are applied at ± 1 as well as at the Chebyshev nodes, the resulting polynomials $H_{2n+3}(T_a,f)$ will converge uniformly to f. Perhaps surprisingly, this does not occur.

In fact, Hermite–Fejér interpolation on the augmented Chebyshev nodes exhibits some very bad properties! For example, Berman [1] showed that even for $f(x) = x^2$, $H_{2n+3}(T_a,f)(x)$ diverges as $n \to \infty$ for all $x \in (-1,1)$. (Note that this result doesn't extend to $[-1,1]$ because ± 1 are nodes for all n.) An explanation for "Berman's phenomenon" was provided by R. Bojanić [2], who showed that if $f \in C[-1,1]$ and the left and right derivatives $f'_L(1)$ and $f'_R(-1)$ exist, then $H_{2n-1}(T_a,f) \to f$ uniformly if and only if $f'_L(1) = f'_R(-1) = 0$.

Figure 5 shows a typical Lebesgue function $\lambda_{1,n+2}(T_a,x)$ for Hermite–Fejér interpolation on the augmented Chebyshev nodes T_a. On comparing Figures 4 and 5, it appears that the Lebesgue constant $\Lambda_{1,n+2}(T_a)$ for Hermite–Fejér interpolation is much larger than the Lebesgue constant $\Lambda_{n+2}(T_a)$ for Lagrange interpolation. This was confirmed by Smith [16], who used methods similar to those employed in the proof of Theorem 3.1 of this paper to show

$$\Lambda_{1,n+2}(T_a) = \begin{cases} 2n^2 + 3 + O(1/n), & \text{if } n \text{ is even,} \\ 2n^2 + 3 - \pi^2/2 + O(1/n), & \text{if } n \text{ is odd.} \end{cases}$$

4. A weighted interpolation method

In a paper in 1995, Mason and Elliott [12] studied certain weighted interpolation methods based on the zeros of the Chebyshev polynomials of the second, third and

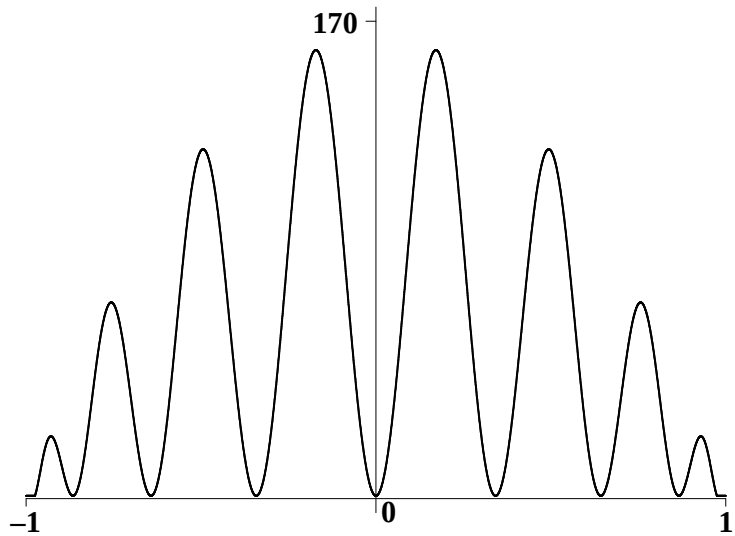

Figure 5: Lebesgue function for Hermite–Fejér interpolation on the augmented Chebyshev nodes $\{\cos(2k-3)\pi/(2n) : 2 \leqslant k \leqslant n+1\} \cup \{\pm 1\}$ [with $n = 9$].

fourth kinds. Although the resulting interpolating functions are not polynomials, there are many similarities between the study of these functions and the study of Lagrange interpolation polynomials. We illustrate Mason and Elliott's ideas by discussing their weighted interpolation method based on the zeros of Chebyshev polynomials of the second kind.

Denote the set of algebraic polynomials of degree at most n by Π_n, let $w(x)$ denote the weight function $w(x) = \sqrt{1-x^2}$, and let X and $x_{i,n}$ be given by (1.1) and (1.2) with $x_{i,n} \neq \pm 1$. We consider the interpolating projection $P_{n-1}(X)$ of $C[-1,1]$ on $w\Pi_{n-1}$ that is defined by

$$P_{n-1}(X)(f)(x) = w(x) \sum_{i=1}^{n} f(x_{i,n}) \ell_{i,n}(X,x)/w(x_{i,n}). \tag{4.1}$$

Also define $\theta_k = \theta_{k,n} = k\pi/(n+1)$ and put

$$U = \{x_{k,n} = \cos\theta_{k,n} : k = 1, 2, \ldots, n;\ n = 1, 2, 3, \ldots\}.$$

(Thus for fixed n, the $x_{k,n}$ are the zeros of the nth Chebyshev polynomial of the second kind.)

Mason and Elliott showed that the projection norm (or Lebesgue constant)

$$\|P_{n-1}(U)\| = \max_{\|f\| \leqslant 1} \|P_{n-1}(U)(f)\|$$

has the representation

$$\|P_{n-1}(U)\| = \max_{0 \leqslant \theta \leqslant \pi} F_n(\theta),$$

where

$$F_n(\theta) = \frac{|\sin(n+1)\theta|}{n+1} \sum_{i=1}^{n} \left| \frac{\sin \theta_i}{\cos \theta - \cos \theta_i} \right|.$$

Based on numerical computations, Mason and Elliott conjectured that the maximum of $F_n(\theta)$ occurs at $\pi/2$ for even n and asymptotically at $n\pi/(2n+2)$ (which is midway between the θ-nodes of $\theta_{(n-1)/2}$ and $\theta_{(n+1)/2} = \pi/2$) for odd n. This conjecture is supported by the graph of $F_7(\theta)$ in Figure 6.

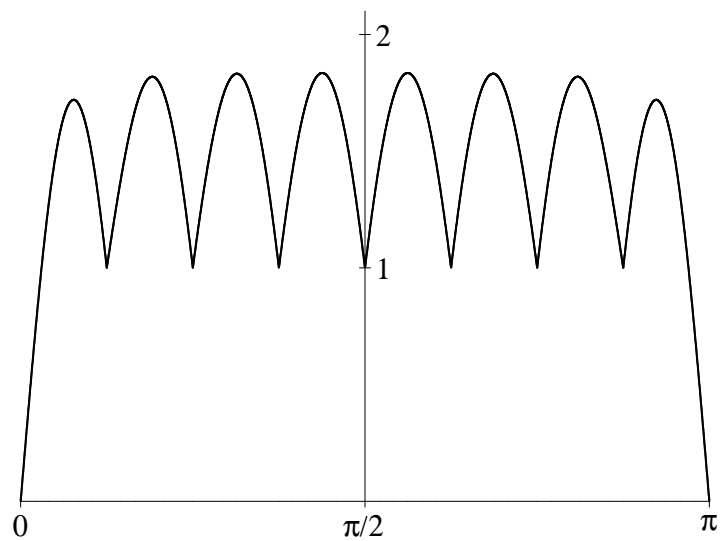

Figure 6: Plot of $F_7(\theta)$

Now, assuming that their conjecture about the maximum of $F_n(\theta)$ is true, Mason and Elliott showed that

$$\|P_{n-1}(U)\| = \frac{2}{\pi} \log n + \frac{2}{\pi} \left(\gamma + \log \frac{4}{\pi} \right) + o(1). \tag{4.2}$$

Smith [18] later established the validity of (4.2), although the proof did not depend on Mason and Elliott's conjecture (which remains unresolved). The result (4.2) means that, to within $o(1)$ terms, $\|P_{n-1}(U)\|$ is equal to $\Lambda_n(X^*)$, the smallest possible Lebesgue constant for *unweighted* Lagrange interpolation (see Section 2.5). Furthermore, by a result of Kilgore [10], the minimum of $\|P_{n-1}(X)\|$ over all X is no smaller than $\Lambda_n(X^*)$. Thus

$$\min_X \|P_{n-1}(X)\| = \|P_{n-1}(U)\| + o(1),$$

which means that for the weighted interpolation method defined by (4.1), there is a simple description of nodes that are essentially optimal.

Acknowledgements. The author thanks the Department of Mathematics and Statistics of La Trobe University and the Analysis Department of the Rényi Mathematical Institute, Budapest (OTKA T037299) for financial support to attend the Fejér–Riesz conference in Eger in June 2005, where a talk based on a draft version of this paper was presented.

References

[1] Berman, D. L., A study of the Hermite–Fejér interpolation process, *Doklady Akad. Nauk USSR* Vol. 187 (1969), 241–244 (in Russian) [*Soviet Math. Dokl.* Vol. 10 (1969), 813–816].

[2] Bojanić, R., Necessary and sufficient conditions for the convergence of the extended Hermite–Fejér interpolation process, *Acta Math. Acad. Sci. Hungar.* Vol. 36 (1980), 271–279.

[3] Brutman, L., On the Lebesgue function for polynomial interpolation, *SIAM J. Numer. Anal.* Vol. 15 (1978), 694–704.

[4] Brutman, L., Lebesgue functions for polynomial interpolation — a survey, *Ann. Numer. Math.* Vol. 4 (1997), 111–127.

[5] Ehlich, H., Zeller, K., Auswertung der Normen von Interpolationsoperatoren, *Math. Ann.* Vol. 164 (1966), 105–112.

[6] Faber, G., Über die interpolatorische Darstellung stetiger Funktionen, *Jahresber. der Deutschen Math. Verein.* Vol. 23 (1914), 192–210.

[7] Fejér, L., Ueber Interpolation, *Göttinger Nachrichten* (1916), 66–91.

[8] Günttner, R., Evaluation of Lebesgue constants, *SIAM J. Numer. Anal.* Vol. 17 (1980), 512–520.

[9] Günttner, R., On asymptotics for the uniform norms of the Lagrange interpolation polynomials corresponding to extended Chebyshev nodes, *SIAM J. Numer. Anal.* Vol. 25 (1988), 461–469.

[10] Kilgore, T., Some remarks on weighted interpolation, in N. K. Govil *et al.*, eds, *Approximation Theory* (Marcel Dekker, New York, 1998) pp. 343–351.

[11] Luttmann, F. W., Rivlin, T. J., Some numerical experiments in the theory of polynomial interpolation, *IBM J. Res. Develop.* Vol. 9 (1965), 187–191.

[12] Mason, J. C., Elliott, G. H., Constrained near-minimax approximation by weighted expansion and interpolation using Chebyshev polynomials of the second, third, and fourth kinds, *Numer. Algorithms* Vol. 9 (1995), 39–54.

[13] Rivlin, T. J., *An Introduction to the Approximation of Functions*, Dover, New York, 1981.

[14] Rivlin, T. J., *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory*, 2nd ed. Wiley, New York, 1990.

[15] Runge, C., Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten, *Z. für Math. und Phys.* Vol. 46 (1901), 224–243.

[16] Smith, S. J., The Lebesgue function for Hermite–Fejér interpolation on the extended Chebyshev nodes, *Bull. Austral. Math. Soc.* Vol. 66 (2002), 151–162.

[17] Smith, S. J., The Lebesgue function for Lagrange interpolation on the augmented Chebyshev nodes, *Publ. Math. Debrecen* Vol. 66 (2005), 25–39.

[18] Smith, S. J., On the projection norm for a weighted interpolation using Chebyshev polynomials of the second kind, *Math. Pannon.* Vol. 16 (2005), 95–103.

[19] Szabados, J., Vértesi, P., *Interpolation of Functions*, World Scientific, Singapore, 1990.

[20] Tietze, H., Eine Bemerkung zur Interpolation, *Z. Angew. Math. and Phys.* Vol. 64 (1917), 74–90.

[21] Turetskii, A. H., The bounding of polynomials prescribed at equally distributed points, *Proc. Pedag. Inst. Vitebsk* Vol. 3 (1940), 117–127 (in Russian).

[22] Vértesi, P., Optimal Lebesgue constant for Lagrange interpolation, *SIAM J. Numer. Anal.* Vol. 27 (1990), 1322–1331.

Simon J. Smith
Department of Mathematics and Statistics
La Trobe University
P.O. Box 199, Bendigo, Victoria 3552
Australia

Annales Mathematicae et Informaticae
***33** (2006) pp. 125–140*
http://www.ektf.hu/tanszek/matematika/ami

Performance modeling tools with applications*

János Sztrik[a], Che Soong Kim[b]

[a]Faculty of Informatics, University of Debrecen
e-mail: jsztrik@inf.unideb.hu

[b]Department of Industrial Engineering, Sangji University
e-mail: dowoo@mail.sangji.ac.kr

Submitted 10 March 2006; Accepted 1 October 2006

Abstract

This paper deals with the role of performance modeling tools. It introduces 3 major tool development centers and shows how a given tool can be applied to investigate the performance of a finite-source retrial queueing system.

Keywords: Performance tools, tool support, finite-source retrial queueing systems.

1. Introduction

The argument for performance engineering methods and tools to be employed in computer-communication systems has always been that such systems cannot be designed or modified efficiently without recourse to some form of predictive model, just as in other fields of engineering. This argument has never been more valid than it is with today's highly complex combination of communication and computer technologies. These have created the Internet, the Grid and diverse types of parallel and distributed computer systems. To be practical, performance engineering relies on tools to render its use accessible to the non-performance specialist, and in turn these depend on sound techniques that include analytical methods, stochastic models and simulation. Tools and techniques also need to be parameterised and validated against real world observations, requiring sophisticated measurement

*Research is partially supported by KOSEF-Hungarian Academy of Sciences Bilateral Scientific Cooperation Grant KOSEF F01-2006-000-10014-0, 2004 and Hungarian Scientific Research Fund-OTKA K60698/2005

techniques in this picosecond cyber-world. The series of "International Conferences on Modelling Techniques and Tools for Computer Performance Evaluation" (TOOLS) has provided a forum for this community of performance engineers with all their diverse interests and selected papers of these conferences have been published in the reputed ***Performance Evaluation*** journal. Many mathematical techniques have been developed to derive various measures from Markov reward models which form the basis of almost all performability models. For evaluation techniques to be used, software tools are needed. The interested reader is referred to, among others, [8, 7] for comprehensive and very detailed surveys on relevant tools.

The paper is organized as follows. Section 2 is devoted to the introduction of some recent well-known tools. In Section 3 tool supported analysis of finite-source retrial queueing systems with a server subject to breakdowns is performed. Finally, some conclusions are made.

2. Some recent modeling tools

In the following 3 major tool development centers are introduced.

2.1. Tools at Faculty of Informatics, University of Dortmund, Germany

This traditionally famous center has developed several software packages which can be downloaded from the site:

`http://ls4-www.informatik.uni-dortmund.de/tools.html`

Parallel to their methodological work the center continuously developed and used tools for performance evaluation. Their intention is to provide facilities for a model description close to the original system specification and hiding details of the analysis techniques. Such tools map the model specification automatically to an analysable model. The set of tools developed by Informatik IV comprises amongst others

- **HIT:** The software tool HIT provides for model-based performance evaluation of computing and communication systems during all phases of their life cycle. Specification of (models of) dynamic, discrete-event, stochastic systems is achieved by particular language- and graphics-based description options. Performance evaluation of accordingly specified models is supported by a variety of techniques of the simulative and analytical types.

- **HiQPN:** HiQPN-Tool supporting the analysis of hierarchical QPN models, a superset of Coloured GSPNs and Queueing Networks. These models can be analysed with respect to qualitative and quantitative aspects.

- **DSPNexpress:** DSPNexpress is a software package for performance and reliability modelling of computer systems with deterministic and stochastic Petri nets (DSPN). DSPNexpress has a user-friendly graphical interface for definition, analysis and graphical animation of DSPNs. The package has been called DSPNexpress because this software package solves complex DSPNs with four orders of magnitude less CPU time than other packages previously introduced.

- **APNN-Toolbox:** The APNN-Toolbox is a software package for functional (invariant, liveness, model checking) and quantitative analysis (APNNsim, NSolve, Parallel, SupGSPN) of GSPNs. Modeling of GSPNs can be realized with the editor APNNed. The editor can start the analysis tools and afterwards a visualization of the results is possible. The communication between APNNed and the analyzers takes place by a common exchange interface so-called Abstract Petri net notation (APNN).

All tools provide a graphical user interface and are available for usual type of workstations. They are tested exhaustively and have been employed for the evaluation of operating systems in the development phase, future hardware architectures, and the assessment of distributed and telecommunication systems. The tools also suit for an evaluation of flexible manufacturing systems. Their practical usability has been proved by a large number of external installations in universities, research institutes and industry.

2.2. Möbius Tool

Möbius is a software tool for modeling the behavior of complex systems, see [5]. It is one of the major research projects of the Performability Engineering Research Group (PERFORM) in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, USA. Although it was originally developed for studying the reliability, availability, and performance of computer and network systems, its use has expanded rapidly. It is now used for a broad range of discrete-event systems, from biochemical reactions within genes to the effects of malicious attackers on secure computer systems, in addition to the original applications. That broad range of use is possible because of the flexibility and power found in Möbius, which come from its support of multiple high-level modeling formalisms and multiple solution techniques. This flexibility allows engineers and scientists to represent their systems in modeling languages appropriate to their problem domains, and then accurately and efficiently solve the systems using the solution techniques best suited to the systems' size and complexity. Time- and space-efficient discrete-event simulation and numerical solution are both supported. Features of the tool is the following.

- **Multiple modeling languages, based on either graphical or textual representations:** Supported model types include stochastic extensions to Petri nets, Markov chains and extensions, and stochastic process algebras.

Models are constructed with the right level of detail, and customized to the specific behavior of the system of interest.

- **Hierarchical modeling paradigm:** Build models from the ground up. First specify the behavior of individual components, and then combine the components to create a model of the complete system. It is easy to combine components in multiple ways to examine alternative system designs.

- **Customized measures of system properties:** Construct detailed expressions that measure the exact information desired about the system (e.g., reliability, availability, performance, and security). Measurements can be conducted at specific time points, over periods of time, or when the system reaches steady state.

- **Study the behavior of the system under a variety of operating conditions:** Functionality of the system can be defined as model input parameters, and then the behavior of the system can be automatically studied across wide ranges of input parameter values to determine safe operating ranges, to determine important system constraints, and to study system behaviors that could be difficult to measure experimentally with prototypes.

- **Distributed discrete-event simulation:** Evaluates the custom measures using efficient simulation algorithms to repeatedly execute the system, either on the local machine or in a distributed fashion across a cluster of machines, and gather statistical results of the measures.

- **Numerical solution techniques:** Exact solutions can be calculated for many classes of models, and advances in state-space computation and generation techniques make it possible to solve models with tens of millions of states. Previously, such models could be solved only by simulation.

The Möbius tool was built based on the belief that no one modeling formalism can be the best way to build all models of systems from across the diverse spectrum of application domains. In addition to the fact that many domain-specific modeling languages are needed, we also need many techniques (for example, simulation, state space exploration, and analytical solution) for analyzing models to study important behaviors of the systems being modeled. Möbius addresses those issues by defining a broad framework in which new modeling formalisms and model solution methods can be easily integrated, and populating that framework with multiple, synergistically combined modeling formalisms and model solution methods. Many advanced modeling formalisms and innovative and powerful solution techniques have been integrated in the Möbius framework. It is available for the following operating systems: Windows XP or 2000, and Linux Fedora Core 3 and later. The package can be found at `http://www.mobius.uiuc.edu/`

2.3. MOSEL Tool

Performance modeling tools usually have their own textual or graphical specification language which depends largely on the underlying modeling formalism. The different syntax of the tool-specific modeling languages implies that once a tool has been chosen it will be difficult to switch to another one as the model has to be rewritten using a different syntax. On the other hand the solution of the underlying stochastic process is performed in most tools by the same standard numerical algorithms. Starting from these observations the creation of MOSEL (MOdeling, Specification and Evaluation Language) tool, developed at the University of Erlangen, Germany, is based on the following idea: Instead of creating another tool with all the components needed for system description, state space generation, stochastic process derivation, and numerical solution, we focus on the formal system description part and exploit the power of various existing and well tested packages for the subsequent stages. MOSEL is a modeling environment which comprises a high-level modeling language that provides a very simple way for system description. In order to reuse existing tools for the system analysis, the environment is equipped with a set of translators which transform the MOSEL model specification into various tool-specific system descriptions [3].

The main features of the MOSEL-environment are the following:

- The modeler inspects the real-world system and generates a high-level system description using the MOSEL specification language. He also specifies the desired performance and reliability measures using the syntax provided by MOSEL. He passes the model to the environment which then performs all following steps without user interaction.
- The MOSEL environment automatically translates the MOSEL model into a tool-specific system description, for example a CSPL-file suitable to serve as input for SPNP.
- The appropriate tool (i.e. SPNP) is invoked by the MOSEL-environment.
- The state space of the model is generated by the tool according to the semantic rules of its modeling formalism out of the static model description.
- The semantic model is mapped onto a stochastic process.
- The stochastic process is solved by one of the standard numerical solution algorithms which are part of the tool. The results of the numerical analysis are saved in a file with a tool-specific structure.
- The MOSEL-environment parses the tool-specific output and generates a result file (sys.res) containing the performance and reliability measures which the user specified in the MOSEL system description. If the modeler requested graphical representation of the results, a second file (sys.igl) is generated by MOSEL [3, 4].

The latest version of MOSEL, called MOSEL-2 has been developed by Jörg Barner and Björn Beutel, see [2]. Moreover, the present version contains the enhanced MOSEL to CSPL translator written by Patrick Wüchner, see [9]. The distribution contains the MOSEL source code (written in C) as well as a collection of examples and installation instructions. Moreover, Björn's diploma-thesis is included in PDF-format. Chapter 4 of his thesis contains an easy-to-read introduction to modeling and performance evaluation with MOSEL-2. MOSEL-2 provides means by which many interesting performance or reliability measures and the graphical presentation of them can be specified straightforwardly. It is especially easy to evaluate a model with different sets of system parameters. The benefit of MOSEL-2 - especially for the practitioner from the industry - lies in its modelling environment: A MOSEL-2 model is automatically translated into various tool-specific system descriptions and then analyzed or simulated by the appropriate tools. This exempts the modeler from the time-consuming task of learning different modeling languages. MOSEL is available under the GNU Public License (GPL). You might read the installation instructions first. All information concerning the package can be found at web address `http://www4.informatik.uni-erlangen.de/Projects/MOSEL/`

System requirements:

- MOSEL-2 should be easily portable to nearly every POSIX system on which an ISO C90 compiler is running.
- You may compile MOSEL-2 using the Portable GNU C Compiler GCC, available from `http://www.gnu.org`.
- It has already been tested under the Linux and Solaris operating system.
- For the graphical display of evaluation results using the IGL interpreter, you need to have Tcl/Tk, version 8.1 or later, installed on your computer system. You can get Tcl/Tk from `http://www.scriptics.com`.
- Last but not least, you'll need an evaluation tool. MOSEL-2 cooperates with the Petri net analysis tools SPNP version 6 (available from `http://www.ee.duke.edu/ kst/softwarepackages.html)` and TimeNET (available from `http://pdv.cs.tu-berlin.de/ timenet`).

3. An application of MOSEL

To show an example for using MOSEL we analyze a retrial queueing system with the following assumptions. Consider a single server queueing system, where the primary calls are generated by K, $1 < K < \infty$ homogeneous sources. The server can be in three states: idle, busy and failed. If the server is idle, it can serve the calls of the sources. Each of the sources can be in three states: free, sending repeated calls and under service. If a source is free at time t it can generate a primary call during interval $(t, t + dt)$ with probability $\lambda dt + o(dt)$. If the server is

free at the time of arrival of a call then the call starts to be served immediately, the source moves into the under service state and the server moves into busy state. The service is finished during the interval $(t, t+dt)$ with probability $\mu dt + o(dt)$ if the server is available. If the server is busy at the time of arrival of a call, then the source starts generation of a Poisson flow of repeated calls with rate ν until it finds the server free. After service the source becomes free, and it can generate a new primary call, and the server becomes idle so it can serve a new call. The server can fail during the interval $(t, t+dt)$ with probability $\delta dt + o(dt)$ if it is idle, and with probability $\gamma dt + o(dt)$ if it is busy. If $\delta = 0, \gamma > 0$ or $\delta = \gamma > 0$ *active or independent breakdowns* can be discussed, respectively. If the server fails in busy state, it either *continues servicing* the interrupted call after it has been repaired or the interrupted request *transmitted to the orbit*. The repair time is exponentially distributed with a finite mean $1/\tau$. If the server is failed two different cases can be treated. Namely, *blocked sources* case when all the operations are stopped, that is neither new primary calls nor repeated calls are generated. In the *unblocked (intelligent) sources* case only service is interrupted but all the other operations are continued (primary and repeated calls can be generated). All the times involved in the model are assumed to be mutually independent of each other.

Our main objective is to continue the investigations which were started in [1] but because of page limitations only some results were presented. The mean number of requests staying in the orbit or in the service, overall utilization of the system and the mean response time of calls are displayed as the function of server's failure and repair rates. To achieve this goal MOSEL is used to formulate and solve the problem.

The system state at time t can be described with the process

$$X(t) = (Y(t); C(t); N(t)),$$

where $Y(t) = 0$ if the server is up, $Y(t) = 1$ if the server is failed, $C(t) = 0$ if the server is idle, $C(t) = 1$ if the server is busy, $N(t)$ is the number of sources of repeated calls at time t. Because of the exponentiality of the involved random variables this process is a Markov chain with a finite state space. Since the state space of the process $(X(t), t \geq 0)$ is finite, the process is ergodic for all reasonable values of the rates involved in the model construction, hence from now on we will assume that the system is in the steady state.

We define the stationary probabilities:

$$P(q; r; j) = \lim_{t \to \infty} P(Y(t) = q, C(t) = r, N(t) = j),$$

$$q = 0, 1, \;\; r = 0, 1, \;\; j = 0, ..., K^*,$$

$$\text{where } K^* = \begin{cases} K - 1 & \text{for blocked case,} \\ K - r & \text{for unblocked case.} \end{cases}$$

Knowing these quantities the main performance measures can be obtained as follows:

- *Utilization of the server*

$$U_S = \sum_{j=0}^{K-1} P(0,1,j).$$

- *Utilization of the repairman*

$$U_R = \sum_{r=0}^{1} \sum_{j=0}^{K^*} P(1,r,j).$$

- *Availability of the server*

$$A_S = \sum_{r=0}^{1} \sum_{j=0}^{K^*} P(0,r,j) = 1 - U_R.$$

- *Mean number of calls staying in the orbit or in service*

$$M = E[N(t) + C(t)] = \sum_{q=0}^{1} \sum_{r=0}^{1} \sum_{j=0}^{K^*} jP(q,r,j) + \sum_{q=0}^{1} \sum_{j=0}^{K-1} P(q,1,j).$$

- *Utilization of the sources*

$$U_{SO} = \begin{cases} \frac{E[K-C(t)-N(t);Y(t)=0]}{K} & \text{for blocked case,} \\ \frac{K-M}{K} & \text{for unblocked case.} \end{cases}$$

- *Overall utilization*

$$U_O = U_S + KU_{SO} + U_R.$$

- *Mean rate of generation of primary calls*

$$\overline{\lambda} = \begin{cases} \lambda E[K - C(t) - N(t); Y(t) = 0] & \text{for blocked case,} \\ \lambda E[K - C(t) - N(t)] & \text{for unblocked case.} \end{cases}$$

- *Mean response time*

$$E[T] = M/\overline{\lambda}.$$

3.1. The MOSEL implementation

Because of the fact that the state space of the describing Markov chain is very large (especially in the heterogeneous model we would like to investigate later), it is difficult to calculate the system measures in the traditional way of solving the system of steady-state equations. To simplify the procedure and to make our study more usable in practice, we used the software tool MOSEL to formulate the model and to calculate the main performance measures. By the help of MOSEL we can use various performance tools (like SPNP – Stochastic Petri Net Package) to get these measures. In this section we show the base MOSEL program and the explanation of its main parts without the technical details of programming. This program belongs to the case of continued service after server's repair and request's generation is blocked during the server repairing. It does not contain the picture section, which is needed to generate various graphical representations of the measures. The figures in the next section are automatically generated by the tool with the corresponding picture part. In the MOSEL program we used the following terminology: The server and the sources are referred to as a CPU and terminals, respectively.

```
/* retrialnr-hom-cpu-cont.msl begins */
/*------------------------------------ Definitions -----------*/
#define NT 6
VAR double prgen;
VAR double prretr;
VAR double prrun;
VAR double cpubreak_idle;
VAR double cpubreak_busy;
VAR double cpurepair;
enum cpu_states {cpu_busy, cpu_idle};
enum cpu_updown {cpu_up, cpu_down};
/*------------------------------------ Node definitions ------*/
NODE busy_terminals[NT] = NT;
NODE retrying_terminals[NT] = 0;
NODE waiting_terminals[1] = 0;
NODE cpu_state[cpu_states] = cpu_idle;
NODE cpu[cpu_updown] = cpu_up;
/*---------------------------------------- Transitions ------*/
IF cpu==cpu_up FROM cpu_idle, busy_terminals
   TO cpu_busy, waiting_terminals W prgen*busy_terminals;
IF cpu==cpu_up AND cpu_state==cpu_busy FROM busy_terminals
   TO retrying_terminals W prgen*busy_terminals;
IF cpu==cpu_up FROM cpu_idle, retrying_terminals
   TO cpu_busy, waiting_terminals W prretr*retrying_terminals;
IF cpu==cpu_up FROM cpu_busy, waiting_terminals
   TO cpu_idle, busy_terminals W prrun;
IF cpu_state==cpu_idle FROM cpu_up TO cpu_down W cpubreak_idle;
IF cpu_state==cpu_busy FROM cpu_up TO cpu_down W cpubreak_busy;
```

```
FROM cpu_down TO cpu_up W cpurepair;
/*------------------------------------------ Results -------*/
RESULT>> if(cpu==cpu_up AND cpu_state==cpu_busy) cpuutil += PROB;
RESULT>> if(cpu==cpu_up) goodcpu += PROB;
RESULT if(cpu==cpu_up) busyterm += (PROB*busy_terminals);
RESULT>> termutil = busyterm / NT;
RESULT>> if(cpu==cpu_up) retravg += (PROB*retrying_terminals);
RESULT if(waiting_terminals>0) waitall += (PROB*waiting_terminals);
RESULT if(retrying_terminals>0)
                              retrall += (PROB*retrying_terminals);
RESULT>> resptime = (retrall + waitall) / NT / (prgen * termutil);
RESULT>> overallutil = cpuutil + busyterm;
/* retrialnr-hom-cpu-cont.msl ended */
```

In the *declaration part* we define the number of terminals (NT), this is the only program code line, that must be modified when modeling larger systems. The terminals have three states: busy (primary call generation), retrying (repeated call generation) and waiting (job service at the CPU). The CPU has two states: idle and busy, and it can be up or failed in both states. We define the three parameters for the terminals: *prgen* denotes the rate of primary call generation, *prretr* references to the rate of repeated call generation and *prrun* denotes the service rate. The *cpubreak_idle*, *cpubreak_busy* and *cpurepair* variables denote the failure rate in the two CPU states and the repair rate.

The *node part* defines the nodes of the system. Our queueing network contains 5 nodes: one node for the number of busy, retrying and waiting terminals, respectively, and two nodes for the CPU. The CPU is idle and up and all the terminals are busy at the starting time.

The *transition part* describes how the system works. The first transition rule defines the successful primary call generation: the CPU moves from the idle state to busy and the terminal from busy to waiting. The second rule shows an unsuccessful primary call generation: if the CPU is busy when the call is generated then the terminal moves to state retrying. The third rule handles the case of a successful repeated call generation: the CPU moves from the idle state to busy and the terminal from retrying to waiting. The fourth rule describes the request service at the CPU. The fifth and sixth rules describe the CPU fail in idle and busy state. The last rule shows the CPU repair.

Finally, the *result part* calculates the requested output performance measures.

3.2. Numerical examples

We used the tool SPNP which was able to handle the model with up to 126 sources. In this case, on a computer containing a 1.1 GHz processor and 512 MB RAM, the running time was approximately 1 second.

The results in the reliable case (with very low failure rate and very high repair rate) were validated by the (a little modified) Pascal program for the reliable case

	retrial (cont.)	retrial (orbit)	reliable [6]
Number of sources:	5	5	5
Request's generation rate:	0.2	0.2	0.2
Service rate:	1	1	1
Retrial rate:	0.3	0.3	0.3
Utilization of the server:	0.5394868123	0.5394867440	0.5394867746
Mean response time:	4.2680691205	4.2680667075	4.2680677918

Table 1: Validations in the reliable case

	retrial (cont.)	retrial (orbit)	non-rel. FIFO
Number of sources:	3	3	3
Request's generation rate:	0.1	0.1	0.1
Service rate:	1	1	1
Retrial rate:	1e+25	1e+25	–
Server's failure rate:	0.01	0.01	0.01
Server's repair rate:	0.05	0.05	0.05
Utilization of the server:	0.2232796561	0.2232796553	0.2232796452
Mean response time:	1.4360656331	1.4360656261	1.4360655471

Table 2: Validations in the non-reliable case

	K	λ	μ	ν	δ, γ	τ
Figure 1	6	0.8	4	0.5	x axis	0.1
Figure 2	6	0.1	0.5	0.5	x axis	0.1
Figure 3	6	0.1	0.5	0.05	x axis	0.1
Figure 4	6	0.8	4	0.5	0.05	x axis
Figure 5	6	0.05	0.3	0.2	0.05	x axis
Figure 6	6	0.1	0.5	0.05	0.05	x axis

Table 3: Input system parameters

given in [6], on pages 272–274. See Table 1 for some test results. The non-reliable case was tested with the non-reliable FIFO model, see Table 2.

In Figures 1–3 we can see the mean response time, the overall utilization of the system and mean number of calls staying in the orbit or in the service for the reliable and the non-reliable retrial system when the server's failure rate increases. In Figures 4–6 the same performance measures are displayed as the function of increasing repair rate. The input parameters are collected in Table 3.

3.3. Comments

In Figure 1, we can see that in the case when the request returns to the orbit at the breakdown of the server, the sources will have always longer response times. Although the difference is not considerable it increase as the failure rate increase. The almost linear increase in $E[T]$ can be explained as follows. In the blocked (non-intelligent) case the failure of the server blocks all the operations and the response time is the sum of the down time of the server, the service and repeated call generation time of the request (which does not change during the failure) thus the failure has a linear effect on this measure. In the intelligent case the difference is only that the sources send repeated calls during the server is unavailable, so this is not an additional time.

In Figure 2 and Figure 5 it is shown how much the overall utilization is higher in the intelligent case with the given parameters. It is clear that the continued cases have better utilizations, because a request will be at the server when it has been repaired.

In Figure 3 we can see that the mean number of calls staying in the orbit or in service does not depend on the server's failure rate in continuous, non-intelligent case, it coincides with the reliable case. It is because during and after the failure the number of requests in these states remains the same. The almost linear increase in the non-continuous, non-intelligent case can be explained with that if the server failure occurs more often the server will be idle more often after repair until a source repeats his call.

In Figure 4, we can see that if the request returns to the orbit at the breakdown of the server, the sources will have longer response times like in Figure 1. The difference is not considerable too, and as it was expected the curves converge to the reliable case.

In Figure 6, it can be seen that the mean number of calls staying in the orbit or in service does not depend on the server's repair rate in continuous, non-intelligent case, it coincides with the reliable case like in Figure 3. It is true for the non-continuous, non-intelligent case too, which has more requests in the orbit on the average because of the non-continuity.

4. Conclusions

This paper introduced some recent performance modeling tools of well-known research centers of famous universities. In Section 3 a finite-source homogeneous retrial queueing system was studied with the novelty of the non–reliability of the server. The MOSEL tool was used to formulate and solve the problem, and the main performance and reliability measures were derived and analyzed graphically. Several numerical calculations were performed to show the effect of server's breakdowns and repairs on the mean response times of the calls, on the overall utilization of the system and on the mean number requests staying in the orbit or in service.

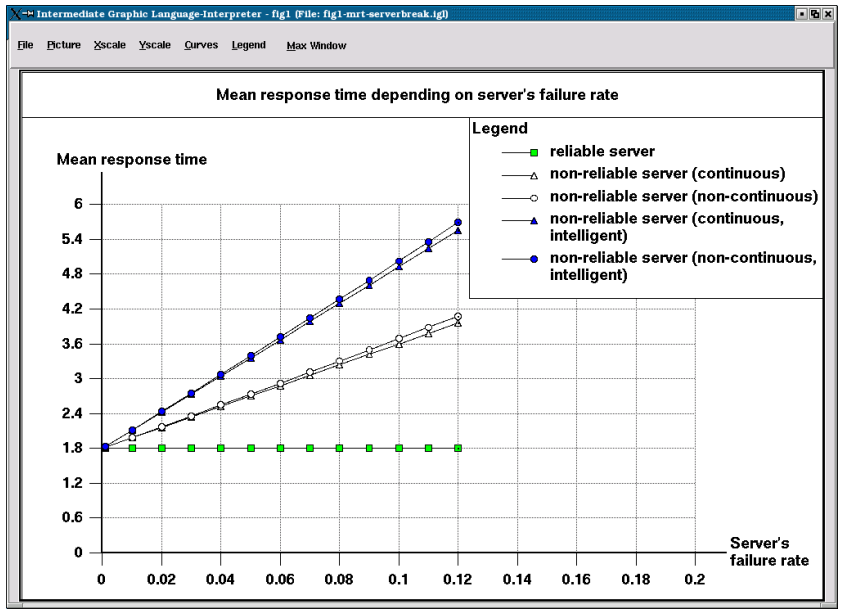

Figure 1: $E[T]$ versus server's failure rate

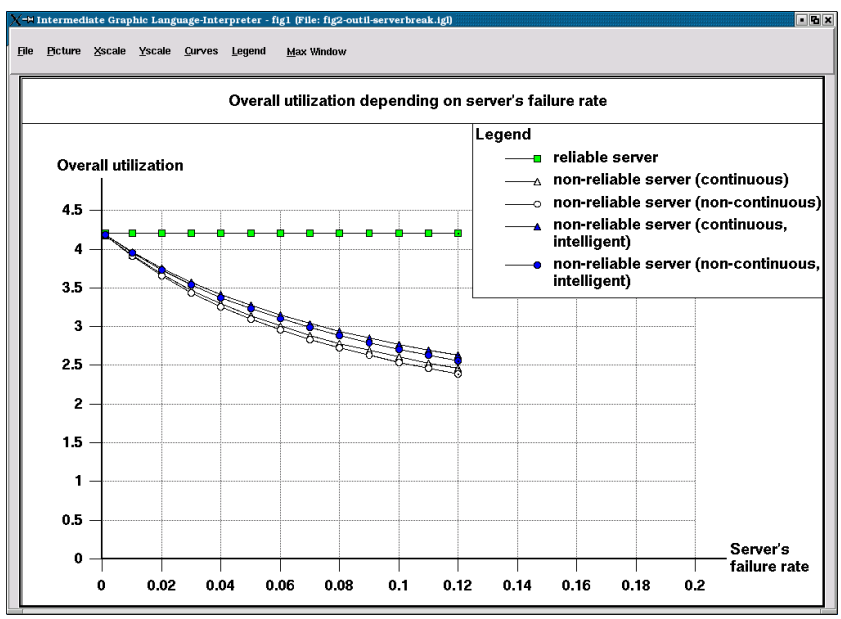

Figure 2: U_O versus server's failure rate

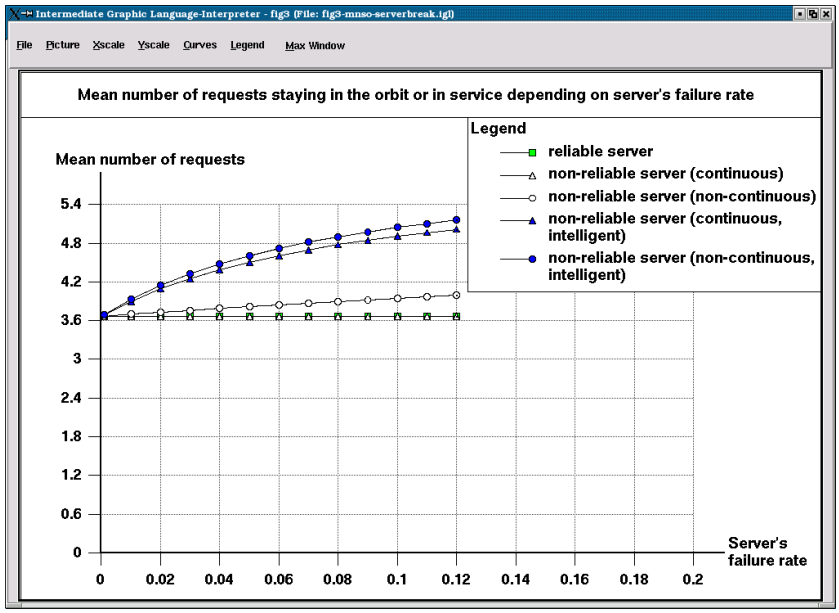

Figure 3: M versus server's failure rate

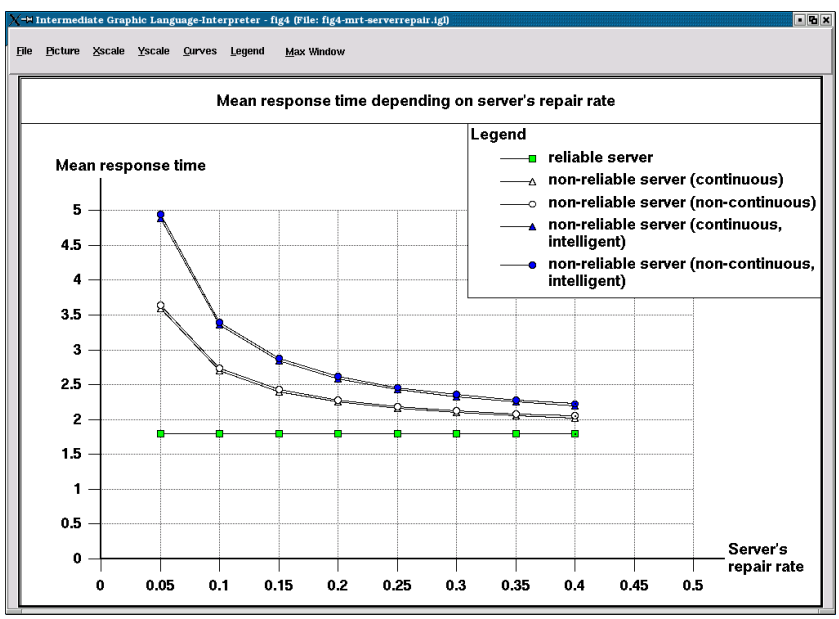

Figure 4: $E[T]$ versus server's repair rate

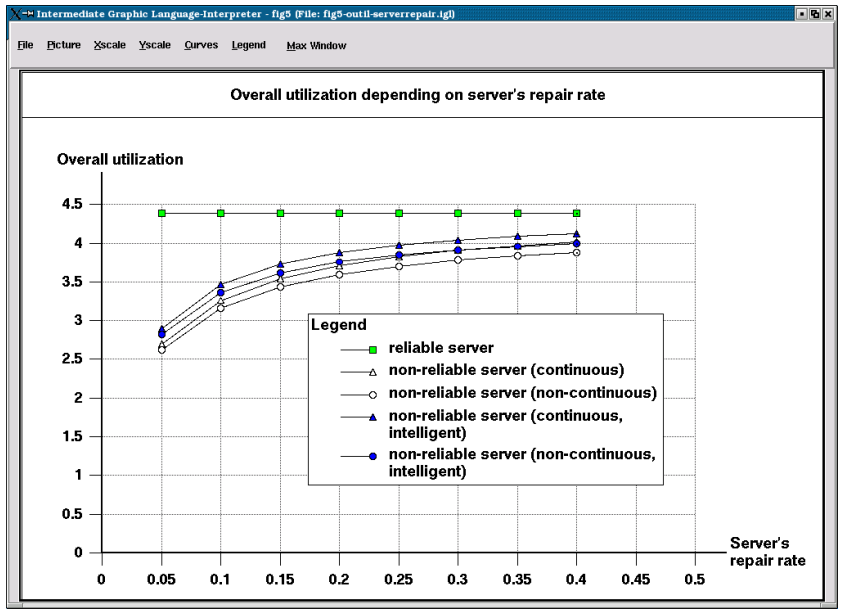

Figure 5: U_O versus server's repair rate

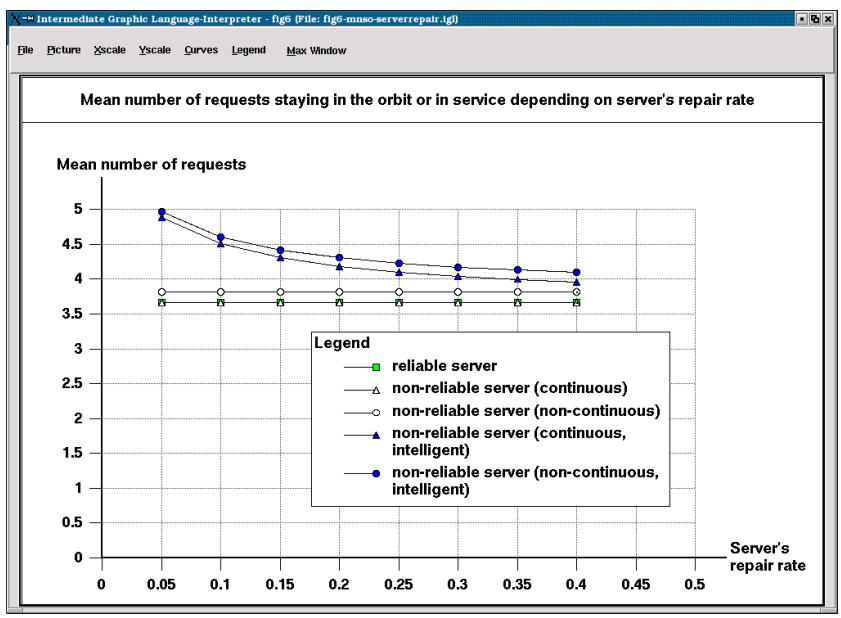

Figure 6: M versus server's repair rate

References

[1] Almási, B., Roszik, J. and Sztrik J., *Homogeneous finite-source retrial queues with server subject to breakdowns and repairs*, Mathematical and Computer Modeling 42 (2005) 673–682.

[2] Barner, J. and Bolch, G., *MOSEL-2-Modeling, Specification and Evaluation Language, Revision 2*, Proceedings of the 13th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation, Performance TOOLS 2003, Urbana-Champaign, Illinois, (2003) 222–230.

[3] Begain, K., Bolch, G. and Herold, H., *Practical performance modeling, application of the MOSEL language*, Kluwer Academic Publisher, Boston, 2001.

[4] Begain, K., Barner, J., Bolch, G. and Zreikat, A., *The Performance and Reliability Modelling Language MOSEL and its Applications*, International Journal on Simulation: Systems, Science, and Technology 3 (2002) 69–79.

[5] Derisavi, S., Kemper, P., Sanders, W. H. and Courtney, T., *The Möbius state-level abstract functional interface*, Performance Evaluation 54 (2003) 105–128.

[6] Falin, G. I. and Templeton, J. G. C., *Retrial queues,* Chapman and Hall, London, 1997.

[7] Haverkort, B. R., Rindos, A., Mainka, V. and Trivedi, K., *Techniques and Tools for Reliability and Performance Evaluation: Problems and Perspectives*, Seventh International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, Vienna, Austria (1994) 1–24.

[8] Haverkort, B. R. and Niemegeers, I. G., *Performability modelling tools and techiques*, Performance Evaluation 25 (1996) 17–40.

[9] Wüchner, P., de Meer, H., Barner, J. and Bolch, G., *MOSEL-2 - A Compact But Versatile Model Description Language and Its Evaluation Environment*, Proceedings of the Workshop: MMBnet'05, University of Hamburg, Germany, 2005.

János Sztrik
Faculty of Informatics
University of Debrecen
P.O. Box 12
H-4010 Debrecen
Hungary

Che Soong Kim
Department of Industrial Engineering
Sangji University
Wonju, 220-702
Korea

Annales Mathematicae et Informaticae
***33** (2006) pp. 141–149*
http://www.ektf.hu/tanszek/matematika/ami

A Hájek–Rényi type inequality and its applications

Tibor Tómács[a], Zsuzsanna Líbor[b]

[a]Department of Applied Mathematics, Eszterházy Károly College
e-mail: tomacs@ektf.hu

[b]Department for Methodology of Economic Analysis, Szolnok College
e-mail: liborne@szolf.hu

Submitted 8 August 2006; Accepted 18 September 2006

Abstract

A general method is presented to obtain strong laws of large numbers. Then it is applied for certain dependent random variables to obtain some strong laws.

1. Introduction

It is well-known that the Hájek–Rényi inequality (see [7]) is a generalization of the Kolmogorov inequality. In this paper we show (Theorem 2.1) that Kolmogorov's inequality implies a certain Hájek–Rényi type inequality. Using this fact we give a general method to obtain strong laws of large numbers (Theorem 2.4). Actually our method is the same as the one applied in Fazekas and Klesov [5] and Fazekas et al. [6] but here we use probabilities instead of moments. In the proof we follow the lines of [5].

Our theorem offers a general tool: if a maximal inequality is known for a certain sequence of random variables then one can easily obtain a strong law of large numbers. Our scheme helps to find the conditions and the normalizing constants.

In section 3 we apply our theorem to give alternative proofs for some known strong laws of large numbers. We deal with associated, negatively associated random variables and demimartingales.

2. Results

Let $\mathbb{N}$ be the set of the positive integers and $\mathbb{R}$ the set of real numbers. If $a_1, a_2, \ldots \in \mathbb{R}$ then in case $A = \emptyset$ let $\max_{k\in A} a_k = 0$ and $\sum_{k\in A} a_k = 0$. Let $\{X_k,\ k \in \mathbb{N}\}$ be a sequence of random variables defined on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and $S_k = \sum_{i=1}^{k} X_i$ for all $k \in \mathbb{N}$.

Theorem 2.1. *Let $\{\alpha_k,\ k \in \mathbb{N}\}$ be a sequence of nonnegative real numbers and $r > 0$. Then the following two statements are equivalent.*

(i) *There exists $c > 0$ such that for any $n \in \mathbb{N}$ and any $\varepsilon > 0$*

$$\mathrm{P}\Big(\max_{k\leqslant n} |S_k| \geqslant \varepsilon\Big) \leqslant c\varepsilon^{-r} \sum_{k=1}^{n} \alpha_k.$$

(ii) *There exists $c > 0$ such that for any nondecreasing sequence $\{\beta_k,\ k \in \mathbb{N}\}$ of positive real numbers, any $n \in \mathbb{N}$ and any $\varepsilon > 0$*

$$\mathrm{P}\Big(\max_{k\leqslant n} |S_k|\, \beta_k^{-1} \geqslant \varepsilon\Big) \leqslant c\varepsilon^{-r} \sum_{k=1}^{n} \alpha_k \beta_k^{-r}.$$

Proof. The proof is based on the idea of the proof of Theorem 1.1 in Fazekas and Klesov [5]. It is clear that (ii) implies (i). Now we turn to (i) $\Rightarrow$ (ii). Let $0 < \beta_1 \leqslant \beta_2 \leqslant \ldots$, $n \in \mathbb{N}$ and $\varepsilon > 0$ are fixed. Without loss of generality we can assume that $\beta_1 = 1$. Introduce the following notation

$$A_i = \{m\ :\ 1 \leqslant m \leqslant n \text{ and } 2^i \leqslant \beta_m^r < 2^{i+1}\},\quad i = 0, 1, 2, \ldots,$$
$$I = \max\{i\ :\ A_i \neq \emptyset\},$$
$$m_i = \begin{cases} \max A_i, & \text{if } A_i \neq \emptyset, \\ m_{i-1}, & \text{if } A_i = \emptyset, \end{cases} \qquad i = 0, 1, 2, \ldots \text{ and } m_{-1} = 0.$$

Then we have

$$\begin{aligned}
&\mathrm{P}\Big(\max_{k\leqslant n} |S_k|\, \beta_k^{-1} \geqslant \varepsilon\Big) \leqslant \sum_{i=0}^{I} \mathrm{P}\Big(\max_{k\in A_i} |S_k| \geqslant \varepsilon 2^{i/r}\Big) \\
&\leqslant \sum_{i=0}^{I} \mathrm{P}\Big(\max_{k\leqslant m_i} |S_k| \geqslant \varepsilon 2^{i/r}\Big) \leqslant \sum_{i=0}^{I} c\varepsilon^{-r} 2^{-i} \sum_{k=1}^{m_i} \alpha_k \\
&= c\varepsilon^{-r} \sum_{k=0}^{I} \sum_{j\in A_k} \alpha_j \sum_{i=k}^{I} 2^{-i} \leqslant 2c\varepsilon^{-r} \sum_{k=0}^{I} 2^{-k} \sum_{j\in A_k} \alpha_j \\
&\leqslant 2c\varepsilon^{-r} \sum_{k=0}^{I} \sum_{j\in A_k} \alpha_j 2\beta_j^{-r} = 4c\varepsilon^{-r} \sum_{k=1}^{n} \alpha_k \beta_k^{-r}.
\end{aligned}$$

Thus the theorem is proved. □

The following two lemmas are due to Fazekas and Klesov (see [4, Lemma 2.1 and Lemma 2.2]).

Lemma 2.2. *Let* $\{\lambda_k,\ k \in \mathbb{N}\}$ *be a sequence of nonnegative real numbers. Assume that* $\sum_{k=1}^{\infty} \lambda_k 2^{-k} < \infty$. *Then there exists a nondecreasing unbounded sequence* $\{\gamma_k,\ k \in \mathbb{N}\}$ *of positive real numbers such that*

$$\sum_{k=1}^{\infty} \lambda_k \gamma_k^{-1} < \infty \quad \textit{and} \quad \lim_{k\to\infty} \gamma_k 2^{-k} = 0. \tag{2.1}$$

Proof. If finitely many λ_k are positive then the statements are obvious. Suppose that there are infinitely many positive λ_k. Let $z = \sum_{k=1}^{\infty} \lambda_k 2^{-k}$ and let n_i be the smallest integer such that

$$\sum_{k=n_i}^{\infty} \lambda_k 2^{-k} \leqslant z2^{-i}, \quad i = 0, 1, \dots.$$

Let $q_{-1} = 0$, $q_i = \min\{n_j \ :\ j = 0, 1, \dots \text{ and } n_j > q_{i-1}\}$ $(i = 0, 1, \dots)$,

$$B_i = \{k \in \mathbb{N} \ :\ q_i \leqslant k < q_{i+1}\} \quad (i = 0, 1, \dots)$$

and $\gamma_k = 2^{k-i/2}$ for $k \in B_i$. Property $\gamma_k \leqslant \gamma_{k+1}$ has to be verified only for $k = q_{i+1} - 1$, $i = 0, 1, \dots$. In this case $\gamma_{k+1}/\gamma_k = \sqrt{2}$ so $\{\gamma_k,\ k \in \mathbb{N}\}$ is nondecreasing. This equality implies $\lim_{i\to\infty} \gamma_{q_i} = \infty$, so $\{\gamma_k,\ k \in \mathbb{N}\}$ is unbounded. Now we turn to (2.1).

$$\sum_{k=1}^{\infty} \lambda_k \gamma_k^{-1} = \sum_{i=0}^{\infty} \sum_{k\in B_i} \lambda_k \gamma_k^{-1} \leqslant \sum_{i=0}^{\infty} 2^{i/2} \sum_{k=n_i}^{\infty} \lambda_k 2^{-k} \leqslant z \sum_{i=0}^{\infty} 2^{-i/2} < \infty.$$

The last statement follows from the definition of γ_k. □

Lemma 2.3. *Let* $\{\alpha_k,\ k \in \mathbb{N}\}$ *be a sequence of nonnegative real numbers,* $\{b_k,\ k \in \mathbb{N}\}$ *a nondecreasing unbounded sequence of positive real numbers and* $r > 0$. *Assume that* $\sum_{k=1}^{\infty} \alpha_k b_k^{-r} < \infty$. *Then there exists a nondecreasing unbounded sequence* $\{\beta_k,\ k \in \mathbb{N}\}$ *of positive real numbers such that*

$$\sum_{k=1}^{\infty} \alpha_k \beta_k^{-r} < \infty \quad \textit{and} \quad \lim_{k\to\infty} \beta_k b_k^{-1} = 0. \tag{2.2}$$

Proof. Let $w_0 = 0$, $w_i = \max\{k \in \mathbb{N} \ :\ b_k^r \leqslant 2^i\}$ $(i \in \mathbb{N})$,

$$C_i = \{k \in \mathbb{N} \ :\ w_{i-1} + 1 \leqslant k \leqslant w_i\} \quad (i \in \mathbb{N})$$

and $\lambda_i = \sum_{k\in C_i} \alpha_k$. Since

$$\sum_{k=1}^{\infty} \alpha_k b_k^{-r} = \sum_{i=1}^{\infty} \sum_{k\in C_i} \alpha_k b_k^{-r} \geqslant \sum_{i=1}^{\infty} \lambda_i 2^{-i}$$

we get that $\sum_{i=1}^{\infty} \lambda_i 2^{-i} < \infty$. So all conditions of Lemma 2.2 are satisfied. Let $\{\gamma_k,\ k \in \mathbb{N}\}$ be fixed by Lemma 2.2. Now we put

$$\beta_k = \gamma_i^{1/r} \quad \text{for} \quad k \in C_i.$$

Then

$$\infty > \sum_{i=1}^{\infty} \lambda_i \gamma_i^{-1} \sum_{i=1}^{\infty} \sum_{k \in C_i} \alpha_k \gamma_i^{-1} = \sum_{k=1}^{\infty} \alpha_k \beta_k^{-r}.$$

The other statements are obvious. □

Theorem 2.4. *Let $\{\alpha_k,\ k \in \mathbb{N}\}$ be a sequence of nonnegative real numbers, $r > 0$ and $\{b_k,\ k \in \mathbb{N}\}$ a nondecreasing unbounded sequence of positive real numbers. Assume that*

$$\sum_{k=1}^{\infty} \alpha_k b_k^{-r} < \infty$$

and there exists $c > 0$ such that for any $n \in \mathbb{N}$ and any $\varepsilon > 0$

$$\mathrm{P}\Big(\max_{k \leqslant n} |S_k| \geqslant \varepsilon\Big) \leqslant c\varepsilon^{-r} \sum_{k=1}^{n} \alpha_k. \tag{2.3}$$

Then

$$\lim_{n \to \infty} S_n b_n^{-1} = 0 \quad \textit{almost surely (a.s.).}$$

Proof. The proof is based on the idea of the proof of Theorem 2.1 in Fazekas and Klesov [4]. Let $\{\beta_k,\ k \in \mathbb{N}\}$ be fixed by Lemma 2.3. Then (2.3) and Theorem 2.1 imply that there exists $c > 0$ such that for any $n \in \mathbb{N}$ and any $\varepsilon > 0$

$$\mathrm{P}\Big(\max_{k \leqslant n} |S_k|\, \beta_k^{-1} \geqslant \varepsilon\Big) \leqslant c\varepsilon^{-r} \sum_{k=1}^{n} \alpha_k \beta_k^{-r}.$$

By this fact we get for any fixed $m \in \mathbb{N}$

$$\mathrm{P}\Big(\sup_k |S_k|\, \beta_k^{-1} > \varepsilon_m\Big) \leqslant \lim_{n \to \infty} \mathrm{P}\Big(\max_{k \leqslant n} |S_k|\, \beta_k^{-1} \geqslant \varepsilon_m\Big) \leqslant c\varepsilon_m^{-r} \sum_{k=1}^{\infty} \alpha_k \beta_k^{-r},$$

where $\{\varepsilon_m,\ m \in \mathbb{N}\}$ a nondecreasing unbounded sequence of positive real numbers. So we have by (2.2)

$$\lim_{m \to \infty} \mathrm{P}\Big(\sup_k |S_k|\, \beta_k^{-1} > \varepsilon_m\Big) = 0.$$

Hence, using continuity of probability, we have

$$\mathrm{P}\Big(\sup_k |S_k|\, \beta_k^{-1} > \varepsilon_m \text{ for all } m \in \mathbb{N}\Big) = 0.$$

Consequently $\sup_k |S_k|\, \beta_k^{-1} < \infty$ a.s. Thus by (2.2) we get

$$\lim_{k \to \infty} |S_k(\omega)|\, b_k^{-1} = \lim_{k \to \infty} \big(|S_k(\omega)|\, \beta_k^{-1}\big)\, \big(\beta_k b_k^{-1}\big) = 0$$

for almost every $\omega \in \Omega$. Thus the theorem is proved. □

3. Some applications

We shall prove that some known results (i.e. Theorem 3.3, Theorem 3.4, Theorem 3.7, Theorem 3.8 and Theorem 3.12) are special cases of Theorem 2.4.

Associated random variables

Definition 3.1 (Esary et al. [3]). A finite family $\{X_1, \dots, X_n\}$ of random variables is called *associated* if

$$\operatorname{cov}\big(f(X_1, \dots, X_n), g(X_1, \dots, X_n)\big) \geqslant 0$$

for any real coordinatewise nondecreasing functions f, g on $\mathbb{R}^n$ such that the above covariance exists. An infinite family of random variables is associated if its every finite subfamily is associated.

Lemma 3.2 (Matuła [11], Lemma 1). *Assume that $X_1, \dots, X_n$ are associated zero mean random variables with finite second moments. Then for every $\varepsilon > 0$*

$$\mathrm{P}\Big(\max_{k\leqslant n} |S_k| \geqslant \varepsilon\Big) \leqslant 8\varepsilon^{-2}\,\mathrm{E}\,S_n^2.$$

Theorem 3.3 (Matuła [11], Theorem 1). *Let $\{X_k,\ k \in \mathbb{N}\}$ be a sequence of associated random variables with finite second moments and $\{a_k,\ k \in \mathbb{N}\}$ a sequence of positive real numbers satisfying $\sum_{k=1}^{\infty} a_k = \infty$. Let $b_n = \sum_{i=1}^{n} a_i$. Assume that*

$$\sum_{j=1}^{\infty} \sum_{i=1}^{j} a_i a_j \operatorname{cov}(X_i, X_j) b_j^{-2} < \infty.$$

Then

$$\lim_{n\to\infty} (S_n^* - \mathrm{E}\,S_n^*) b_n^{-1} = 0 \quad a.s.,$$

where $S_n^* = \sum_{i=1}^{n} a_i X_i$.

Proof. Without loss of generality we can assume that $\mathrm{E}\,X_k = 0$ for all $k \in \mathbb{N}$. Let $\alpha_k = \mathrm{E}\,S_k^{*2} - \mathrm{E}\,S_{k-1}^{*2}$, where $S_0^* = 0$. Then for all $k \in \mathbb{N}$

$$0 \leqslant \alpha_k \leqslant 2 \sum_{i=1}^{k} a_i a_k \operatorname{cov}(X_i, X_k),$$

so we have

$$\sum_{k=1}^{\infty} \alpha_k b_k^{-2} \leqslant \sum_{k=1}^{\infty} \sum_{i=1}^{k} 2 a_i a_k \operatorname{cov}(X_i, X_k) b_k^{-2} < \infty.$$

It is easy to see that $\{a_k X_k,\ k \in \mathbb{N}\}$ is associated thus, by Lemma 3.2,

$$\mathrm{P}\Big(\max_{k\leqslant n} |S_k^*| \geqslant \varepsilon\Big) \leqslant 8\varepsilon^{-2}\,\mathrm{E}\,S_n^{*2} = 8\varepsilon^{-2} \sum_{k=1}^{n} \alpha_k$$

for any $\varepsilon > 0$. Consequently, by Theorem 2.4, we get $\lim_{n\to\infty} S_n^* b_n^{-1} = 0$ a.s. $\square$

Theorem 3.4 (Birkel [1], Theorem 2 and Christofides [2], Corollary 2.2). *Let $\{X_k,\ k \in \mathbb{N}\}$ be a sequence of associated random variables with finite second moments. If*

$$\sum_{k=1}^{\infty} k^{-2} \operatorname{cov}(X_k, S_k) < \infty$$

then

$$\lim_{n\to\infty} (S_n - \mathrm{E}\, S_n) n^{-1} = 0 \quad a.s.$$

Proof. Without loss of generality we can assume that $\mathrm{E}\, X_k = 0$ for all $k \in \mathbb{N}$. Let $\alpha_k = \operatorname{cov}(X_k, S_k)$, $b_k = k$ and $S_0 = 0$. Then, by Lemma 3.2, we have

$$\mathrm{P}\Big(\max_{k\leqslant n} |S_k| \geqslant \varepsilon\Big) \leqslant 8\varepsilon^{-2}\, \mathrm{E}\, S_n^2 = 8\varepsilon^{-2} \sum_{k=1}^{n} \big(\mathrm{E}\, S_k^2 - \mathrm{E}\, S_{k-1}^2\big) \leqslant 16\varepsilon^{-2} \sum_{k=1}^{n} \alpha_k.$$

Thus Theorem 2.4 implies the statement. □

Negatively associated random variables

Definition 3.5 (Joag-Dev and Proschan [8]). A finite family $\{X_1, \dots, X_n\}$ of random variables is called *negatively associated* if for any disjoint nonempty subsets $A, B \subset \{1, \dots, n\}$, $A = \{i_1, \dots, i_l\}$, $B = \{i_{l+1}, \dots, i_n\}$ and any real coordinatewise nondecreasing functions f on $\mathbb{R}^l$ and g on $\mathbb{R}^{n-l}$

$$\operatorname{cov}\big(f(X_{i_1}, \dots, X_{i_l}), g(X_{i_{l+1}}, \dots, X_{i_n})\big) \leqslant 0.$$

An infinite family of random variables is negatively associated if every finite subfamily is negatively associated.

The following lemma is a special case of Theorem 2.1 of Liu et al. [9]. (See Lemma 1 of Matuła [10], too.)

Lemma 3.6. *Assume that $X_1, \dots, X_n$ are negatively associated zero mean random variables with finite second moments. Then for every $\varepsilon > 0$*

$$\mathrm{P}\Big(\max_{k\leqslant n} |S_k| \geqslant \varepsilon\Big) \leqslant 32\varepsilon^{-2} \sum_{k=1}^{n} \mathrm{E}\, X_k^2.$$

Theorem 3.7 (Matuła [11], Theorem 2). *Let $\{X_k,\ k \in \mathbb{N}\}$ be a sequence of negatively associated random variables with finite second moments and $\{a_k,\ k \in \mathbb{N}\}$ a sequence of positive real numbers satisfying $\sum_{k=1}^{\infty} a_k = \infty$. Let $b_n = \sum_{i=1}^{n} a_i$. Assume that*

$$\sum_{k=1}^{\infty} a_k^2 b_k^{-2}\, \mathrm{D}^2\, X_k < \infty.$$

Then

$$\lim_{n\to\infty} (S_n^* - \mathrm{E}\, S_n^*) b_n^{-1} = 0 \quad a.s.,$$

where $S_n^* = \sum_{i=1}^{n} a_i X_i$.

Proof. Without loss of generality we can assume that $\mathrm{E}\, X_k = 0$ for all $k \in \mathbb{N}$. Let $\alpha_k = a_k^2 \,\mathrm{E}\, X_k^2$. It is clear that $\{a_k X_k,\ k \in \mathbb{N}\}$ is negatively associated, so by Lemma 3.6 we have

$$\mathrm{P}\Big(\max_{k\leqslant n} |S_k^*| \geqslant \varepsilon\Big) \leqslant 32\varepsilon^{-2} \sum_{k=1}^{n} \alpha_k$$

for any $\varepsilon > 0$. Thus Theorem 2.4 implies the statement. □

Theorem 3.8 (Liu et al. [9], Theorem 3.1). *Let $\{X_k,\ k \in \mathbb{N}\}$ be a sequence of negatively associated random variables with finite second moments and $\{b_k,\ k \in \mathbb{N}\}$ a nondecreasing and unbounded sequence of positive real numbers. Assume that*

$$\sum_{k=1}^{\infty} b_k^{-2} \,\mathrm{D}^2\, X_k < \infty.$$

Then

$$\lim_{n\to\infty} (S_n - \mathrm{E}\, S_n) b_n^{-1} = 0 \quad a.s.$$

Proof. Without loss of generality we can assume that $\mathrm{E}\, X_k = 0$ for all $k \in \mathbb{N}$. Let $\alpha_k = \mathrm{E}\, X_k^2$. Then Lemma 3.6 and Theorem 2.4 imply the statement. □

Demimartingales

We shall use the following notations:

$$X^+ = \max\{0, X\} \quad \text{and} \quad X^- = -\min\{0, X\}.$$

Definition 3.9 (Newman and Wright [12]). Let $\{S_k,\ k \in \mathbb{N}\}$ be an L^1 sequence of random variables. Assume that for $j \in \mathbb{N}$

$$\mathrm{E}\big((S_{j+1} - S_j)\, f\,(S_1, \ldots, S_j)\big) \geqslant 0$$

for all coordinatewise nondecreasing functions f on $\mathbb{R}^j$ such that the expectation is defined. Then $\{S_k,\ k \in \mathbb{N}\}$ is called a *demimartingale*. If in addition the function f is assumed to be nonnegative, the sequence $\{S_k,\ k \in \mathbb{N}\}$ is called a *demisubmartingale*.

Lemma 3.10 (Christofides [2], Theorem 2.1). *Let $\{S_k,\ k \in \mathbb{N}\cup\{0\}\}$ be a demisubmartingale with $S_0 = 0$. Let $\{b_k,\ k \in \mathbb{N}\}$ be a nondecreasing sequence of positive real numbers. Then for all $\varepsilon > 0$*

$$\mathrm{P}\Big(\max_{k\leqslant n} S_k b_k^{-1} \geqslant \varepsilon\Big) \leqslant \varepsilon^{-1} \sum_{k=1}^{n} b_k^{-1} \,\mathrm{E}\left(S_k^+ - S_{k-1}^+\right).$$

The following lemma is a corollary of Lemma 2.1 and Corollary 2.1 of Christofides [2].

Lemma 3.11. *If* $\{S_k,\ k \in \mathbb{N}\}$ *is demimartingale then* $\{(S_k^+)^r,\ k \in \mathbb{N}\}$ *and* $\{(S_k^-)^r,\ k \in \mathbb{N}\}$ *are demisubmartingales for all* $r \geqslant 1$.

Theorem 3.12 (Christofides [2], Theorem 2.2). *Let* $\{S_k,\ k \in \mathbb{N} \cup \{0\}\}$ *be a demimartingale with* $S_0 = 0$. *Let* $\{b_k,\ k \in \mathbb{N}\}$ *be a nondecreasing and unbounded sequence of positive real numbers. Let* $r \geqslant 1$ *and* $\mathrm{E}\,|S_k|^r < \infty$ *for each* $k \in \mathbb{N}$. *Assume that*

$$\sum_{k=1}^{\infty} b_k^{-r}\,\mathrm{E}\left(|S_k|^r - |S_{k-1}|^r\right) < \infty.$$

Then

$$\lim_{n\to\infty} S_n b_n^{-1} = 0 \quad a.s.$$

Proof. Let $\alpha_k = \mathrm{E}\left(|S_k|^r - |S_{k-1}|^r\right)$ for all $k \in \mathbb{N}$ and $\varepsilon > 0$. By Lemma 3.11 and 3.10

$$\begin{aligned}\mathrm{P}\Big(\max_{k\leqslant n}|S_k| \geqslant \varepsilon\Big) &\leqslant \mathrm{P}\Big(\max_{k\leqslant n}(S_k^+)^r \geqslant \varepsilon^r/2\Big) + \mathrm{P}\Big(\max_{k\leqslant n}(S_k^-)^r \geqslant \varepsilon^r/2\Big)\\ &\leqslant 2\varepsilon^{-r}\sum_{k=1}^{n}\mathrm{E}\left((S_k^+)^r + (S_k^-)^r - (S_{k-1}^+)^r - (S_{k-1}^-)^r\right) = 2\varepsilon^{-r}\sum_{k=1}^{n}\alpha_k.\end{aligned}$$

Thus Theorem 2.4 implies the statement. □

Acknowledgements. Our paper was inspired by the ideas of Oleg Klesov and István Fazekas. The authors would like to thank István Fazekas for several helpful discussions and for his attention to our paper.

References

[1] Birkel, T., Moment bounds for associated sequences, *Ann. Probab.* 16 (1988) 1184–1193.

[2] Christofides, T. C., Maximal inequalities for demimartingales and a strong law of large numbers, *Stat. & Prob. Letters* 50 (2000) 357–363.

[3] Esary, J., Proschan, F. and Walkup, D., Association of random variables with applications, *Ann. Math. Statist.* 38 (1967) 1466–1474.

[4] Fazekas, I. and Klesov, O., A general approach to the strong law of large numbers, Technical Report No. 4/1998, Universitas Debrecen, Hungary.

[5] Fazekas, I. and Klesov, O., A general approach to the strong laws of large numbers, *Theory of Probab. Appl.*, 45/3 (2000) 568–583.

[6] Fazekas, I., Klesov, O. I., Noszály, Cs., Tómács, T., Strong laws of large numbers for sequences and fields, (Proceedings of the Third Ukrainian-Scandinavian Conference in Probability Theory and Mathematical Statistics 8–12 June 1999. Kyiv, Ukraine) *Theory of Stochastic Processes*, Vol.5 (21) no. 3–4 (1999) 91–104.

[7] Hájek, J. and Rényi, A., Generalization of an inequality of Kolmogorov, *Acta Math. Acad. Sci. Hungar.* 6 no. 3–4 (1955) 281–283.

[8] Joag-Dev, K. and Proschan, F., Negative association of random variables with applications, *Ann. Statist.* 11 (1983) 286–295.

[9] Liu, J., Gan, S. and Chen, P., The Hájeck-Rényi inequality for the NA random variables and its application, *Stat. & Prob. Letters* 43 (1999) 99–105.

[10] Matuła, P., A note on the almost sure convergence of sums of negatively dependent random variables, *Stat. & Prob. Letters* 15 (1992) 209–213.

[11] Matuła, P., Convergence of weighted averages of associated random variables, *Prob. and Math. Statist.* 16, Fasc. 2 (1996) 337–343.

[12] Newman, C. M. and Wright, A. L., Associated random variables and martingale inequalities, *Z. Wahrsch. Verw. Geb.* 59 (1982) 361–371.

Tibor Tómács
Department of Applied Mathematics
Károly Eszterházy College
P.O. Box 43
H-3301 Eger
Hungary

Zsuzsanna Líbor
Department for Methodology of Economic Analysis
Szolnok College
Ady Endre u. 9.
H-5000 Szolnok
Hungary

Annales Mathematicae et Informaticae
***33** (2006) pp. 151–165*
http://www.ektf.hu/tanszek/matematika/ami

Variations for spanning trees

László Zsakó

Faculty of Informatics ELTE
e-mail: Zsako@ludens.elte.hu

Submitted 20 June 2006; Accepted 8 December 2006

Abstract

Coursebooks discussing graph algorithms usually have a chapter on minimum spanning trees. It usually contains Prim's and Kruskal's algorithms [1, 2] but often lacks other applications. This type of problem is rarely present at informatics competitions or in tests in secondary or higher level informatics education This article is aimed at describing some competition tasks that help us prove that the application of the above algorithms are well-suited for both competition and evaluation purposes.

The Hungarian National Informatics Competition for Secondary School Students look back on a history of 20 years and so does the International Olympiad in Informatics. Basically, informatics competitions rely on algoritmization tasks [3, 4], the circle of which is continually developing though showing a surprisingly great constancy at the same time.

At both types of competitions there are often tasks connected to graphs or problems that can be reduced to representation of graphs. International competitions nearly lack tasks in connection with minimum spanning trees. On the other hand, at national competitions (Hungarian National Informatics Competitions for Secondary School Students and the Selecting Competition for the International Olympiad) the scientific committees deciding on tasks (mainly Gyula Horváth and László Zsakó) has set this kind of problem several times. You could also have met a similar one for instance at the 11th Lithuanian Olympiad in Informatics (Winter in The Kingdom of Fancy).

This article is about the experience gained so far: with the help of the tasks below, we would like to show that their solution is not a simple description of the two classical spanning tree algorithms (Prim's and Kruskal's) but their creative application [5, 6].

As being a methodologist, in this article I do not intend to invent fundamentally new algorithms but rather discover the applicability and limits of algorithms known so far.

When Hungarian secondary school students are being prepared for competitions, we rely on two fundamental books on algorithms [1, 2]. Therefore, our stepping stone here are two basic algorithms described in these books (one is taken from one of the books while the other is from the other one). This article is not aimed at describing further algorithms that are too complicated for this age group but the studying of the applicability of the two basic procedures.

Definition 1.13. Let $G = (V, E)$ be a connected undirected graph. An acyclic connected $F(V, E')$ subgraph of a G graph is a spanning tree of the graph where $E' \subseteq E$ (A spanning tree is a tree that contains all the nodes of G and its edges are taken from the edges of G.) Let a weight function $c : E \to \mathbb{R}$ be given. Then an F spanning tree of G is a minimum spanning tree if its cost (the sum of the weights of its edges) is the minimum of all the spanning trees of G.

Prim's procedure [1]: Let $G = (V, E)$ be a connected graph, $V = \{1, \ldots, N\}$. Starting from node 1, expand the tree containing this node until – expanded with all the nodes – you get the minimum spanning tree.

```
Procedure Prim(G: graph; var F: set of edges);
  var U: set of nodes;
    u,v: nodes;
  begin
    F:={};
    U:={1};
    while U ≠ G.V do begin
      let (u,v) be a minimum weight edge, for which u ∈ U and v ∈ G.V\U;
      F := F ∪ {(u,v)};
      U := U ∪ {v};
    end;
  end;
```

Kruskal's procedure [2]: Sort the edges in increasing order by weight. Create an N-member spanning forest from node N. Check the edges and if there are any that connect various spanning trees, join the spanning trees.

```
Procedure Kruskal(G: graph; var A: set of edges)
  A := ∅;
  for all v ∈ G.V do Create-a-set(v);
  sort the edges of G.E in increasing order by weight w;
  for all (u,v) ∈ G.E
    do if Find-set(u) ≠ Find-set(v)
         then begin A := A ∪ {(u,v)}; Join(u,v); end;
  end;
```

In this article we are only dealing with problems that can be solved by secondary school students participating at informatics competitions if they are familiar with the above basic graph algorithms. The essence of the method is the analysis of two starting situations:

- Let us modify the text of the task by adding or dropping conditions then examine whether the above algorithms are applicable and if so what modifications are necessary.
- Let us modify either of the above algorithms then find a sensible problem to match it, which can thus be solved.

First let us see some competition tasks from the past few years and the solutions we expected to receive. You can find examples for both strategies among them as well as among the ideas that follow them.

Problem 1.14 (Plant – spanning forest).[1] *A company manufactures goods in its plants based in K cities, which are then to be transported to N cities. The transport routes must be strengthened so that heavy lorries will be able to run along them. Therefore, the transport routes are meant to be arranged in such a way that the total length of the transport routes to be strengthened will be the smallest possible. Write a program that determines the minimum total length of the roads to be strengthened as well as from which city to which city the goods are to be transported.*

At this problem it is feasible to start with Prim's algorithm. The basic variant takes an arbitrary node (right node 1) as a starting tree. In this task take a starting forest with nodes $1, 2, \ldots, K$. Here you do not even need to assume that the graph is connected. The only thing is that in its every connected component there must be at least one plant.

In the solution you will need the spanning trees themselves. The starting nodes of the spanning trees are the nodes with plants. The task expects you to determine for all the nodes of the graph from which nodes of a given spanning tree they can be reached directly.

Therefore, you have to modify the algorithm at two parts–at the ones in italics:

```
Procedure Prim(G: graph; var F: set of edges);
  var U: set of nodes;
      u,v: nodes;
  begin
    F := {};
    U := {1, 2, ..., K};
    for all u ∈ U do w(u) := u;
    while U ≠ G.V do begin
      let (u, v) be the cheapest edge for which u ∈ U and v ∈ G.V\U;
```

[1] The Selecting Competition for the International Olympiad in 2002. Competition tasks are printed in italics. From the original tasks we have omitted the descriptions of the imput and the output as well as the limitations and the conditions.

```
        F := F ∪ {(u, v)}; w(v) := u;
        U := U ∪ {v};
    end;
end;
```

Here Kruskal's procedure is a bit awkward to use because you should work around so as not to join two sub-trees in both of which there is a plant.

Problem 1.15 (Mill – spanning forest).[2] *The Milling Company has flour milled and packed in K Mills. The flour should be transported to N cities in such a way that the total transportation cost be the lowest possible. Write a program that gives the minimum transportation cost, and for each town from which mill the flour is to be transported.*

The problem is fairly similar to the previous one: you have to construct a spanning forest here, as well [7]. As for time, it was set after that one at the competition but it is a bit tricky because in this problem it is not the total length of the edges that should be minimum but the total length of the paths deriving from the roots (mill) of trees. Therefore, it is possible that the shortest edge of the graph is not an element of the spanning forest as the figure above shows (Kruskal's procedure would first join these two nodes). According to the previous problem (Mill,A), (A,B) would be the minimum spanning tree, whereas for this problem the correct solution is (Mill,A), (Mill,B).

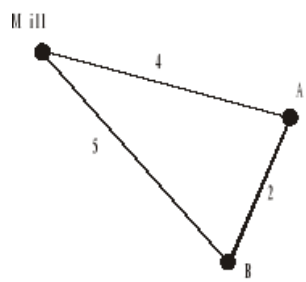

Figure 1:

Thus you need to store in the modified Prim's algorithm for each node of the spanning trees to-be-built, the length of the shortest path leading from the root (mill) to it, then you need to attach a node to one of its spanning trees to which the path leading from the root is the shortest. A further modification is that here for each node of the spanning tree you need to store its root (i.e. the starting node of the path leading there instead of the edge leading there).

Now you have to modify the algorithm at three places; the modifications here are in italics, as well:

[2]The Hungarian National Informatics Competition for Secondary School Students in 2005.

```
Procedure Prim(G: graph; var F: set of edges);
  var U: set of nodes;
      u,v: nodes;
  begin
    F := ;
    U := {1, 2, ..., K}; for all u ∈ U do w(u) := u;
    while U ≠ G.V do begin
      Let (u, v) be an edge where u ∈ U and v ∈ G.V\U and
            Min(u) + weight(u, v) are minimum;
      F := F ∪ {(u, v)}; w(v) := w(u);
      U := U ∪ {v}; Min(v) := Min(u) + weight(u, v);
    end;
  end;
```

Remark 1.16. If you started from one node, you would actually get a variant of Dijkstra's algorithm to determine the minimum paths starting from one node.

Problem 1.17 (Pipeline – a spanning forest for a complete graph).[3] *In a country oil refineries were built at K places, where petrol is produced from crude oil. There are N places with filling stations where the petrol should be transported through pipeline. Both of them are given with their co-ordinates. The pipeline must be designed in such a way that minimum length of pipes are to be laid. The pipeline can fork only at filling stations or refineries and can be led only horizontally or vertically in the map. The refineries are numbered from 1 to K while filling stations from K+1 to K+N. Write a program that determines between which nodes (oil refineries or filling stations) the pipeline should be constructed in a way that the lenght of the pipeline will be minimum.*

The problem is nearly identical with the "Plant" problem. There is just one little peculiarity here: there is an edge between any nodes of the graph and the length i.e. the weight of the edge is equal to the Manhattan-distance of its two extreme nodes.

Problem 1.18 (Road – a spanning tree with starting components).[4]
In a city there are several squares. Some of them are connected with roads while others are not and they should be constructed. Between the squares there are areas allocated by the local government for roads. Only after the construction of roads had started did it turn out that there would not be enough money to build all the roads. On the other hand, the reconversion of those that already exist is also terribly expensive. Write a program that determines which roads the local government should build so that it will cost the least and in the meantime you could travel from each square to each square.

What is peculiar about the problem is that there are starting connected components (not necessarily trees). The components can be defined relying on the

[3]The Selecting Competition for the International Olympiad in 2000.
[4]The Selecting Competition for the International Olympiad in 2003.

already strengthened roads. Therefore, it is not the minimum spanning tree of the starting graph that you should create but the minimum spanning tree of another graph that consists of the components. And from now on you can choose among three solutions:

A. Take a graph consisting of the components (i.e. combine the nodes of the components into the nodes of the new graph), then with the aid of the non-strengthened roads apply any algorithm that gives you a minimum spanning tree.

B. Applying Kruskal's procedure, first join the sub-trees between which there is a strengthened road without any further conditions then relying on the basic algorithm, join those between which there is a non-strengthened road.

Let F be the set of strengthened and E the set of not yet strengthened edges. The modifications are shown in italics:

```
Procedure Kruskal(G: graph;F: set of edges; var A: set of edges)
  A := {};
  for all v ∈ G.V do Make-a-set(v);
  for all (u, v) ∈ F do if Find-a-set(u)≠Find-a-set(v)
       then Join(u,v);
  sort the edges of G.E in increasing order by weight w;
  for all (u, v) ∈ G.E
    do if Find-a-set(u)≠Find-a-set(v)
         then begin A := A ∪ {(u, v)}; Join(u,v); end;
  end;
```

C. As a third solution, the edges of set F could be added to set E with 0 weight. Then the basic algorithm could do nearly without any modifications. The only thing you need to take care about, however, is that you must not include the edges with 0 weight in the edge set A arising as a solution. The solution would be a bit slower since there is a larger edge set here to be sorted than in the case above.

Problem 1.19 (Network – online spanning tree).[5] *In Fairyland a network is to be constructed to connect N cities. During K weeks you keep receiving tenders for constructing a direct connection between two cities. For every week–if possible– give a plan, based on the tenders received so far, regarding which cities are to be connected directly so that every city will be accessible from every other city directly or via other cities but such that the total cost of construction will be minimum. The plan contains the pairs of cities to be connected and the total cost of constructing the network.*

The problem is the online version of creating a minimum spanning tree. The difficulty the competitors are to face is that they rarely, if ever, meet such algorithms at school or at the preparatory courses for the competitions [1].

The problem, however, can be deduced to the following case: if there is a minimum spanning forest and one single new graph edge, how can you make a

[5]The Selecting Competition for the International Olympiad in 2004.

minimum spanning forest? For example, a modification to Kruskal's algorithm could be the following:

```
Procedure online(G: graph; var A: set of edges)
  A := {};
  for all v∈ G.V do Make-a-set(v);
  for all (u,v) ∈ G.E
    do if Find-a-set(u) ≠ Find-a-set(v)
         then begin A := A ∪ {(u, v)}; Join(u,v); end
         else begin
                (p, q):=the maximum weight edge of path (u, v);
                if weight(p, q) > weight(u, v)
                  then A := A ∪ {(u, v)} − {(p, q)};
              end;
  end;
```

Further ideas

Below you will find problem ideas that the scientific committees of the competitions concerned have already dealt with and probably they will be set as competition tasks in the future.

Idea$_1$

The pipeline problem can also be formulated such that for the sake of safety there should be constructed a pipeline where every filling station can be supplied with petrol from at least two oil refineries.

A possible solution is: the solution of the original pipeline problem then connecting K components somehow.

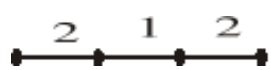

Figure 2:

It means that if you take the graph consisting of components, define the minimum spanning forest consisting of at least 2-element trees. It is easy to see that you cannot apply the greedy strategy here: you must not include the edge with weight 1 you can see in the figure into the spanning forest of the component-graph.

Therefore, this problem cannot be used as a variant of a greedy strategy minimum spanning tree, no matter how much it resembles the basic problem.

Idea$_2$

So far we have been dealing with undirected graphs. And what about directed ones? If you assume that a graph connected and directed i.e. it has at least one node from which there is a path to all other nodes then the problem can be solved.

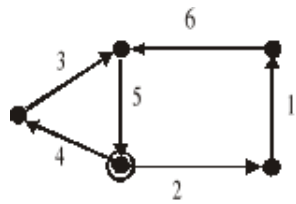

Figure 3:

If there is one starting node (i.e. one to which does not lead an edge), the solution is easy: start constructing the tree from that node following Prim's algorithm.

If there are several such nodes, they are definitely in a cycle (since starting from any node any of them can be accessed). In this case, create for each node their spanning trees, then choose the best one out of them. If you look at the figure, the good starting node is circled. Although all the nodes of the graph can be accessed from every node, if you start from any other nodes, you will get higher-cost spanning trees.

There is just one question left: to determine the nodes from which all the other nodes are accessible.

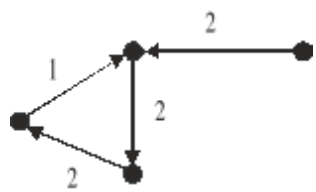

Figure 4:

If you take graph with several starting nodes and you are aware that all the nodes can be accessed from at least one node, you can apply Prim's algorithm (e.g. just like in the "Plant" problem), as well.

Interestingly, if you try to use Kruskal's algorithm, however, you may arrive in a dead-end street as its simple application can result in a wrong spanning tree. For example, as for the graph in the figure the edge of weight 1 is not in the spanning tree.

Idea$_3$

If you are unlucky, in the "Plant" problem you may even have to transport goods to cities with no plants from one single plant whereas from all the other plants nowhere i.e. tree K-1 of the minimum spanning forest consists of only one node whereas the Kth tree contains all the other nodes.

Modify the problem such that every sub-tree contains at least M nodes.

Figure 5:

It is easy to see that this problem cannot be solved with the so far applied greedy strategy. Let the black circles in the figure denote the starting nodes and let all the trees of the spanning forest contain 2 nodes. If you include the edge of weight 1 into any of the sub-trees, the cost of the spanning forest can be only 10 but the correct solution contains only the edges of weights 2 and 3.

For M$\geqslant$4 the problem is NP complete [10]. Therefore, it cannot be used as a variant of a greedy strategy minimum spanning tree although that is a natural extension of the original problem.

Idea$_4$

In an archipelago the islands are connected with bridges. The bridges of a maximum length of H can be re-newed. Write a program that determines the minimum total length of the bridges to be renewed in a way that you can travel from any island to any island. The problem is a modification of the "Plant" problem: the spanning forest must consist of maximum H long edges. Actually, you must determine a subgraph of a graph (excluding edges longer than H) then apply e.g. Kruskal's algorithm on them [8]. When sorting the edges, you can exclude the unnecessary ones.

Idea$_5$

An important feature of Kruskal's algorithm is that the spanning forest being built is always of the lowest cost. The number of the trees of the forest is being decreased by 1 at each step. Relying on this, you can create a K-element minimum spanning forest. The problem is similar to the "Plant" problem, just the nodes of the K-element spanning forest are not given in advance. In this case, the solution is to follow Kruskal's procedure until you have exactly K trees after the joining.

Idea$_6$

You can also modify Kruskal's algorithm in a way that in the meantime you subtract spanning trees of a certain size from further processing. Do this to trees that contain at least M nodes. (We do not expect, however, that the spanning forest created this way will be of minimum cost.)

Lemma 2.20. *Every tree subtracted from further processing can contain maximum* $2^{(M-1)}$ *nodes.*

Proof. According to the task, you will have to deal with trees that contain maximum $M-1$ nodes. The algorithm reduces two trees at each step so the number of elements of the tree can be maximum double the $M-1$. □

You should modify Kruskal's algorithm in a way that you exclude the edges leading out from reduced trees with minimum M nodes (the essence of the modification is in italics):

If *free(u) and free(v) and*
Find-a-set(u)≠ Find-a-set(v) then ...

Idea$_7$

The previous idea shows clearly that if you use Kruskal's procedure, it is far from being guaranteed that you will have an exactly M-element spanning tree in the course of progress.

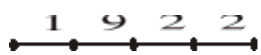

Figure 6:

Prim's procedure is suitable for this in principle but the result you get is not necessarily the minimum M-element spanning tree (the sub-graph containing M arbitrary nodes). If you take the left node in the figure as a starting point, you will get a 10-cost long spanning tree with the left 3 nodes but the cost of the spanning tree on the right containing 3 nodes is only 4. The figure shows that even if you started from the two ends of the shortest edge as a starting tree, it would not help, either.

Relying on Prim's algorithm, however, it may be a sensible task to create a minimum M-element spanning tree that starts from a given "starting node".

Idea$_8$

You can get the second best spanning treea from the minimum spanning tree. Take the minimum spanning tree. Take the edges not belonging to the tree in turn. If you add them to the tree, you will get a cycle. Omit the longest edge of the cycle. If you follow this procedure for each possible edge, from the trees obtained this way that one will be the second best spanning tree in which the difference in length of its new edge and omitted edge is minimum [1, 2].

Idea$_9$

You can formulate the minimum spanning tree problem by saying that you have to construct a spanning tree the longest edge of which is minimum. In this case you do not need a new algorithm as Kruskal's algorithm will give you the correct solution just like for the original problem [8].

Idea$_{10}$

The minimum spanning tree problem can easily be transformed into a maximum spanning tree one. An easy solution: the sum of the weights is maximal when/if the sum of their negatives is minimal.

In another formulation of the problem (previous idea): for the shortest edge to be maximal, the longest among the edges of the respective negative lengths should be minimal.

Another solution: sort the edges in Kruskal's algorithm in decreasing order [9].

Idea$_{11}$

An important feature of Kruskal's algorithm is that the spanning forest being constructed is always minimum and the number of trees in it is decreasing by one at each step. Thus it would be a sensible task to determine a spanning forest consisting of the fewest trees with less than a given cost.

Idea$_{12}$

We are going to connect big cities in Hungary with pipelines suitable for transporting petrol. A petrol pipeline can be built between two cities but the pipelines can be joined at any of their points. Every pipeline connects either two cities or joins a pipeline already constructed. The task is to construct the pipeline network so that you use the fewest pipes possible. Write a program that reads in the co-ordinates (real numbers) of big cities then displays the length of the minimum length pipeline network.

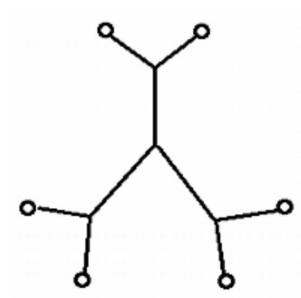

Figure 7:

The problem just sligthly differs from the "Pipeline" problem but its solution is completely different: you have to create a Steiner tree, which is far from being secondary school level.

Big cities of Hungary are to be connected with pipelines suitable for transporting petrol. First a petrol pipeline is constructed between two cities but a new city can be joined to the ready network not only at another city but at any point of the pipeline. Every pipeline connects either two cities or joins a pipeline already constructed. The task is to construct the pipeline network so that you use the least pipes possible. Write a program that reads in the co-ordinates (real numbers) of the cities then displays the length of the minimum length pipeline network.

It is the online version of the problem. It is greedy but it has nothing to do with classical spanning tree algorithms since in this case you have to determine in what order the cities are connected to the network. Undoubtedly, you have to construct the pipeline between the first two cities. Whereas regarding any other city, you

Figure 8:

have to determine the shortest length between the new city and the already existing pipes and then to connect the city to the closest node (city or pipeline). (In the figure you can see the connection sequence.)

Remark 2.21. the result you get this way will not be identical with the Steiner tree of the previous problem. What is more, you will get different spanning trees if the sequences of nodes are different.

Instead of a conclusion – survey experience

With the help of our survey experience, I would like to prove that these problems can be solved by the élite of secondary school students. The above shows that they can be varied highly therefore they are well-suited to be used at informatics competitions.

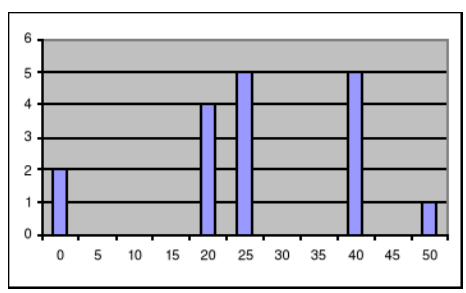

Figure 9: Pipeline

The slight deviation from the maximum score mostly derives from efficiency or from omitting the study of special cases.

This spanning tree task was the first of its kind set at a selecting competition for the student olympiad. The distribution of scores is interesting: about half of the

competitors handed in a 50% solution and another significant percentage achieved an 80% performance. Since it was the first task of its kind, it took the competitors by surprise. The 80% solutions rose from good but not efficient algorithms i.e. as the diagram shows nobody but one student knew (or discovered) the efficient minimum spanning tree algorithm. Behind the 40-50% performance lie solutions of not full value.

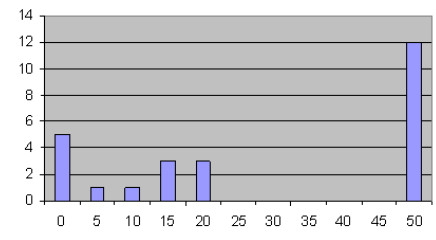

Figure 10: Plant

This diagram shows that in 2002 about 50% of the students participating at the selecting competition for the student olympiad provided a perfect solution while the other half did not deserve a score or just a little. When analysing their solutions, it turned out that they did not know either of the two original algorithm. They attempted at various solutions and in some special cases they got a right one. The students who received 15 to 20 scores, typically devised a correct algorithm, but with their minimum cubic or even worse algorithms they were not able to supply a correct result within the time limit of the competition. It was a definite progress within two years.

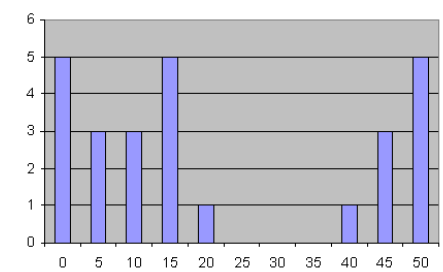

Figure 11: Road

This problem was set at a competition a year after the previous one and it shows a seemingly surprising distribution of scores. Compared to the previous one,

many competitors fell into the 0-40% category. Analysing the solutions, you can see that about half of the 17 competitors belonging to this group did not take into account the ready components. However, they received a correct result for some test cases even this way. Most of those with 0-5 scores started working following Solution A described in the article and they either did not manage to finish work in time or their programs were correct only in a few cases. Half of those with high scores chose a solution similar to Variant B, while the other half to Variant C.

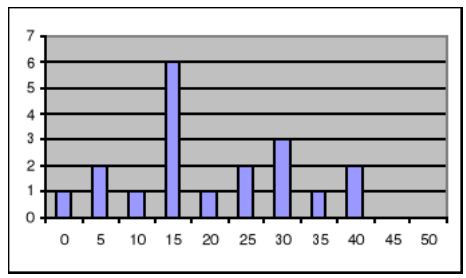

Figure 12: Network

As we mentioned above in the article, the online version took the competitors by surprise. In general, they are not familiar with online algorithms, which is clearly shown by the scores they achieved. Interestingly, about 50% of the competitors were the same as the participants of the competition a year before so they had already seen a spanning tree problem and some of them had given a good solution then.

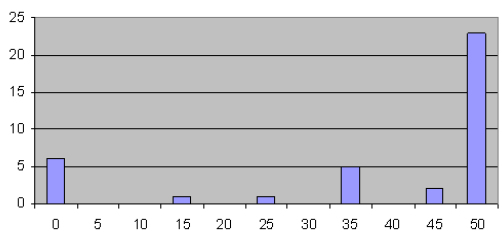

Figure 13: Mill

In 2005 this task was solved by a broader circle that could have seen the solutions of the previous tasks. Typically, here we also had about 50% with a maximum score and 10% with 0 score. Those with scores in between are different from the other task: typically, here they knew the original algorithms that give solutions to this group of problems but they applied them incorrectly.

References

[1] Rónyai Lajos, Ivanyos Gábor, Szabó Réka: Algoritmusok. Typotex, 1999.

[2] T. H. Cormen, C. E. Leiserson, R. L. Rivest: Introduction to Algorithms, MIT, 1990.

[3] `http://olympiads.win.tue.nl/`

[4] `http://ceoi.inf.elte.hu/`

[5] Szlávi Péter, Zsakó László: Módszeres programozás: Gráfok, ELTE Informatikai Kar, 2004.

[6] Szlávi Péter, Zsakó László: Gráfokkal kapcsolatos algoritmusok, NJSZT – Neumann János Tehetséggondozó Program, 2004.

[7] Katona Gyula, Recski András, Szabó Csaba: A számítástudomány alapjai. Typotex, 2002.

[8] Steven Skiena: Analysis of Algorithms, SUNY Stony Brook distance learning program, 1996.

[9] R. Muhammad: Spanning Tree and Minimum Spanning Tree, web-publications 2005, `http://www.personal.kent.edu/~rmuhamma/Algorithms/MyAlgorithms/Greedy/ms.htm`

[10] Xiaoyun Ji, John E. Mitchell: Minimum Weight Constrained Forest Problem, 2005 Optimization Days, Montreal, Canada, 2005.

[11] `http://ims.mii.lt/olimp/?lang=en&sk=pasirengimas&id=0610` – Lithuanian Olympiad in Informatics, tasks

László Zsakó
Faculty of Informatics ELTE
1117 Budapest
Pázmány Péter sétány 1/C
Hungary

Methodological papers

Annales Mathematicae et Informaticae
***33** (2006) pp. 169–174*
http://www.ektf.hu/tanszek/matematika/ami

Gender differences in spatial visualization among engineering students

Brigitta Németh[a], **Miklós Hoffmann**[b]

[a] Department of Descriptive Geometry and Computer Science
Szent István University
e-mail:nemeth.brigitta@ymmfk.szie.hu

[b] Institute of Mathematics and Computer Science
Károly Eszterházy College
e-mail:hofi@ektf.hu

Submitted 24 August 2006; Accepted 12 November 2006

Abstract

Spatial visualization of engineering students is of greatest importance in terms of their professional achievement, thus evaluation of this skill is essential. Mental Cutting Test (MCT) is one of the most widely used evaluation method for spatial abilities. In this study we present an analysis of MCT results of first-year engineering students, with special emphasis on gender differences. Similarly to other international projects, significant difference is observed between male and female students, which is statistically analyzed in this paper.

Keywords: spatial visualization, spatial skills, MCT, gender differences

MSC: 51N05

1. Introduction

Spatial visualization is defined by McGee as "the ability to mentally manipulate, rotate, twist or invert pictorally presented stimuli" [8]. The five components of spatial skills are

- spatial perception
- spatial visualization
- mental rotations

- mental relations
- spatial orientation

To ascertain students' mental skills, some standardized tests have been developed, among which Mental Rotation Test (MRT) and Mental Cutting Test (MCT) are of greatest importance. Mental Rotation Test is introduced by Vanderberg and Kuse [9], while Mental Cutting Test, originally developed for entrance examination in the United States [10], has a long history and widely used for testing the spatial ability of students at any level.

There are other tests like Objective Test on Orthographic Projection (OTO) evaluating the effects of the education in orthographic view [1], or Space Imagination Test (TPP) developed by international cooperation in VEGA project [5].

The aim of this paper is to evaluate classical MCT test results of first-year engineering students in Hungary, with special emphasis on gender differences.

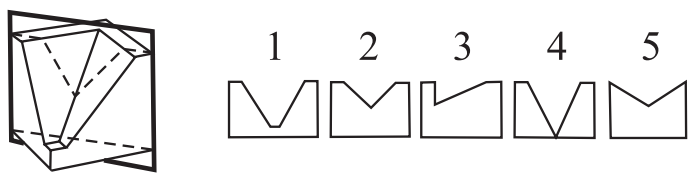

Figure 1: An example of MCT problems (the correct answer is 2).

2. The Mental Cutting Test

In our project we used the standard Mental Cutting Test, which consists of 25 problems. In each problem a perspective drawing of a solid body is given, which has been cut by a plane (c.f. Fig.1). Students are asked for choosing the cross section among the 5 given alternatives, always one being correct.

Basically there are two different types of problems can be found in MCT: pattern recognition problems and quantity problems [2]. In the first category one can find strongly different alternatives of possible cross sections, thus the right solution can be found simply by recognizing the pattern of the section from the spatial figure. In the quantity problems, however, some of the cross sections are similar (more precisely affine) to the correct one, thus the right answer can be determined only by guessing the relative quantities, like ratios of lengths or angles between the edges.

Most of the solids in MCT have relatively complicated, unusual forms, some of them are truncated cubes, others are curved objects, like cylinders. As Tsutsumi et al. reported in [3], failures are mostly based on the fact that students are not able to recognize the spatial form of the object.

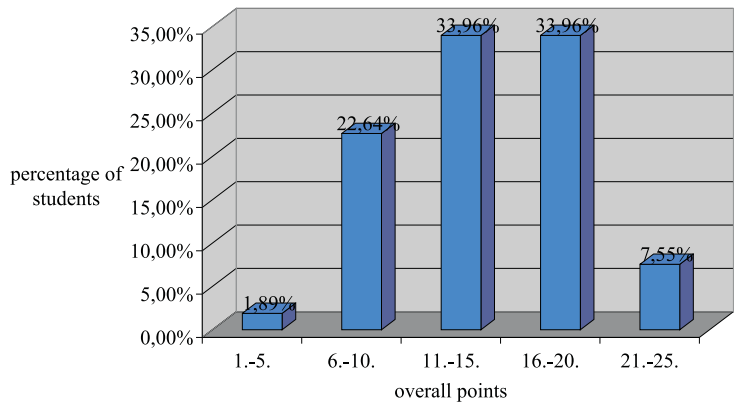

Figure 2: Overall results of male students.

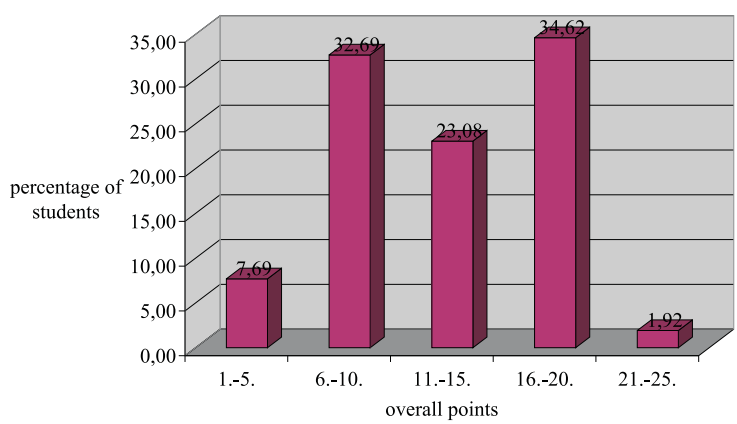

Figure 3: Overall results of female students.

3. Results - Gender Differences in MCT

As it is already stated in [3], females "are much less likely to get high scores in the standard MCT". Similar results have been observed in several countries from Japan through Germany to Poland in an international project by Gorska et al. [4], [7], [5] and even in a recent longitudinal research in Cracow University of Technology [6]. The MCT test in a stereographic circumstances was also used for female students by Tsutsumi et al. [3].

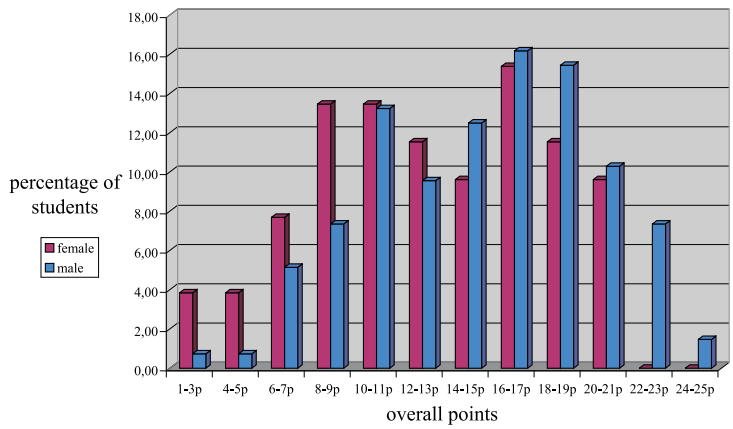

Figure 4: Detailed results of male and female students.

Our purpose was to ascertain the gender differences of Hungarian students in terms of their spatial skills. Here we used the classical MCT test for first-year engineering students of Szent István University.

The test has been filled by 187 students, approximately third of them being female students. Each of the 25 problems counts 1 point, thus the perfect solution yields 25 points. The overall results of the test have been summarized in a way that the range of points has been subdivided into 5 equal parts, 1-5 points, 6-10 points etc. The overall results of male students can be seen in Fig. 2 while Fig. 3 shows the overall results of female students. As one can easily observe from these diagrams, results of female students are strictly worse than that of male students. Especially remarkable the difference in the highest portion: among the students with best spatial skills the number of females are extremely low. Actually the female student with best result had 21 points, just falling into the highest part. This fact can even be better examined in the detailed results (c.f. Fig.4). Note, that the same difference can be seen in the lowest interval (1-5 points), but in the opposite meaning: less than 2% of male students achieved 5 or less points, while amongst female students this rate was more than 7.5%.

Statistical analysis of the results also proves this difference. The means and

standard deviances are summarized in Table 1.

	male	female
mean	14.99	12.69
st.dev.	4.71	5.02

Table 1: Basic statistical analysis of the results.

4. Conclusion and further research

Spatial imagination of engineering students has been studied in this paper. Their abilities were tested by the standard MCT tool which is widely used for this purpose. In accordance with the international experiences we observed relevant differences in male and female students' abilities. Future work will be focused on the possible gender differences in improvement of their spatial skills through their studies. This work requires longitudinal research with regular testing periods along their university studies.

References

[1] Takeyama, K., Maeguchi, R., Chibana, K., Yoshida, K., Evaluation of Objective Test using a pair of orthographic projections for descriptive geometry. *Journal for Geometry and Graphics*, **3** (1999), 99-109.

[2] Tsutsumi, E., A Mental Cutting Test using drawings of intersections, *Journal for Geometry and Graphics*, **8** (2004), 117-126.

[3] Tsutsumi, E., Shiina, K., Suzaki, A., Yamanouchi, K., Takaaki, S., Suzuki, K., A Mental Cutting Test on female students using a stereographic system. *Journal for Geometry and Graphics*, **3** (1999), 111-119.

[4] Gorska, R., Sorby, S., Leopold, C., Gender differences in visualization skills - an international perspective. *The Engineering Design Graphics Journal*, **62**, (1998), 9-18.

[5] Juscaková, Z., Gorska, R., A pilot study of a new testing method for spatial abilities evaluation. *Journal for Geometry and Graphics*, **7** (2003), 237-247.

[6] Gorska, R., Spatial imagination - an overview of the longitudinal research at Cracow University of Technology, *Journal for Geometry and Graphics*, **9** (2005), 201-208.

[7] Gorska, R., Sorby, S., Leopold, C., International comparisons of gender differences in spatial visualization and the effect of graphics instruction on the development of these skills. *Proc. of the 8th Intl. Conf. of Engineering Comp. Graph. and Descriptive Geom. (ICECGDG), Austin, USA*, 1998, 261-266.

[8] McGee, M.G., Human Spatial Abilities: Psychometric studies and environmental, genetic, hormonal and neurological influences, *Psychological Bulletin*, **86**, 899-918.

[9] Vanderberg, S.G., Kuse, A.R., Mental Rotations, a group test of three dimensional spatial visualization. *Perceptual and Motor Skills*, **47** (1978), 599-604.

[10] CEEB Special aptitude test in spatial relations. College Entrance Examination Board, USA, 1939.

Brigitta Németh
Department of Descriptive Geometry and Compute Science
Szent István University
Thököly str. 74.
H-1146 Budapest, Hungary

Miklós Hoffmann
Institute of Mathematics and Computer Science
Károly Eszterházy College
Leányka str. 4.
H-3300 Eger, Hungary

Annales Mathematicae et Informaticae
***33** (2006) pp. 175–187*
`http://www.ektf.hu/tanszek/matematika/ami`

Making slides for lecture by LaTeX*

Péter Olajos[a], Erzsébet Orosz[b]

[a] EKF, Institute of Mathematics and Informatics
e-mail: olaj@ektf.hu

[b] EKF, Institute of Mathematics and Informatics
e-mail: ogyne@ektf.hu

Submitted 26 September 2006; Accepted 23 November 2006

Abstract

Our purpose is giving the possibilities of making professional lectures. Tour of our paper consists of several lecture styles and we hope that everybody can find the best one from these ones. All of these lectures were made in format `pdf`, that is they are compatible and portable documents. By our paper it can be seen that making precise, taxing lectures with dynamical elements in arbitrarily themes is succesful only using LaTeX.

Keywords: Format pdf, TeX, LaTeX, METAPOST, package `texpower`

MSC: Primary: 97U70 Secondary: 97U50, 97U80

1. Introduction – what is TeX?

1.1. Knuth and plainTeX

The history of TeX began in 1977 when a mathematician from Stanford, Donald E. Knuth prepared a program system for making and printing documents in professional way. This system had knowledge of several century in typography using computers and had a lot of developing in this topic. For solving typographic problems TeX used packages which consisted of macros. Knuth made a lot of macro packages and it was plainTeX. Later two other system were born[1]: first was $\mathcal{AMS}$-TeX by Michael Spivak ([5]) and it was supported by American Mathematical Society ($\mathcal{AMS}$) and the second program system was LaTeX by Leslie Lamport

*Supported in part by Grant T-48945 and T-48791 from the Hungarian National Foundation for Scientific Research.

[1] In using of the system TeX the authors have also got through the following steps: plainTeX→ $\mathcal{AMS}$-TeX → LaTeX.

([7]). So, the name TEX means the entire system that is plainTEX and LATEX are members of it. In the following we will deal with possibilities of LATEX.

1.2. LATEX

Since 80's LATEX have been developped dynamically and have had a lot of changes (see: section Introduction in [7]). Nowadays LATEX is a uniform program language and by using it we can easily solve every problems in word processing. The structure of LATEX is a system of packages that is if we want to use a special command, we use a package and after loading of it the command will be usable. Number of packages of TEX is hard to calculate, because it depends on operating system and distribution[2], e.g. MikTEX, TEXlive, teTEX.

A very important property of system of TEX that our source making in LATEX is directly convertible into format `pdf`. Using this main property our task is to introduce a free and useful program system which can realize all possibilities in making slides like other non-free programs, e.g. PowerPoint, Scientific Workplace. Unfortunately these programs are uneasy in many cases, e.g. when we want to make mathematical formulas in correct way. Examples (that is packages, see below) of this paper show that these possibities are well usable in other topics, too.

1.3. What does format pdf mean?

Meaning of `pdf` is Portable Document Format. It was developped by Adobe for making a really portable format and to replace other document formats (see for details: [7]). They have done it almost all. So, it is worth saving our files into format `pdf`, because these files will also work correctly in future.

2. Possibilities of LATEX

From the following examples we will see that the lectures made in LATEX are well-structured, correct, brilliant, simple, stylish. So, let us consider the problems below that is what types of tasks we can solve by LATEX.

2.1. Equations

On a mathematical lecture or practice we use equations or system of equations. Making an equation and a system of equations in source is the following:

```
\begin{equation}
4x+3y=\frac{3}{4}
\end{equation}
```

```
\begin{align}
4x+3y&=\frac{3}{4}\\
3x-11y&=2\nonumber\\
-x-13y&=-\frac{11}{4}\\
\end{align}
```

[2] A package from one of these distributions is able to be embedded into a different system easily.

Using the interpreter of LaTeX we get:

$$4x + 3y = \frac{3}{4}. \tag{2.1}$$

and

$$4x + 3y = \frac{3}{4} \tag{2.2}$$

$$3x - 11y = 2$$

$$-x - 13y = -\frac{11}{4} \tag{2.3}$$

We remark that in LaTeX numbering of equations is handled automatically and we can use references to get these numbers in other lines.

2.2. Pictures and other special objects

While preparing a lecture we usually have one ore more pictures. In LaTeX pictures have a special format (which name is called) `eps` and we need the package `graphicx` (or `epsfig`, `mfpic`) to insert them . Consider the simple example below (source and layout (figure 1)):

```
\includegraphics[angle=-45,width=8cm]{gyik.eps}
```

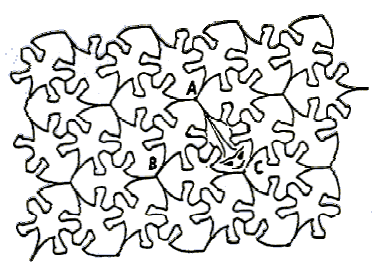

Figure 1: An inserted `eps`

A a Á a B b C c Cs
cs D d Dz dz Dzs dzs E e É é
F f G g Gy gy H h I i Í í J j K
k L l Ly ly M m N n Ny ny
O o Ó ó Ö ö Ő ő P p Q q R
r S s Sz sz T t Ty ty U u
Ú u Ü ü Ű ű V v Z z
Zs zs X x Y y
♡

Figure 2: An example of `shapepar`

For details see [7].

The second example is a special paragraph from package `shapepar`. Source is the following (layout: figure 2):

```
\heartpar{A a \'{A} \'{a} B b ...}}.
```

For more details see [6].

The last example is a special graphic package which give us almost unlimited possibilities to draw an arbitrary picture or figure. This package is `mfpic` and it belongs to program `metapost` (see [4]). Let us see the following simple example (source and and layout: figure 3):

```
\opengraphsfile{fuggveny}
\begin{mfpic}[20]{-3}{3.1}{-3}{3}
\axes \function{-2.4,2.2,0.1}{((x**3)-3x)/4}
\tlabel(1,2.5){$f(x):=\dfrac{(x^3-3x)}{4}$}
\end{mfpic}
\bigskip
\begin{mfpic}[5]{-10}{10}{-10}{10}
\ellipse[6]{(0,0),6,10} \tlabel(3,7){This is an ellipse!}
\end{mfpic}
\closegraphsfile
```

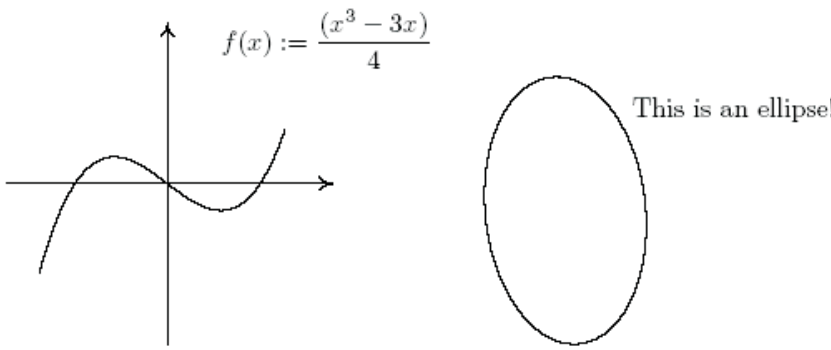

Figure 3: Functions by package `mfpic`

Consider a few main properties of package `mfpic` that is how to use these things for teaching:

- We can use elementary functions as embedded functions, that is precise figures and graphs are able to be seen on slides of lecture.
- We can use options of erasuring and filling (several coding systems of colour), that is we get shielding and three-dimensional figures (to develope use of high-dimensional objects).
- Using arrows, lines or dashed (dotted) lines.

2.3. Problems of TEX

Programmers using TEX have met the problems below which have already been solved or have also had more different solutions. It is important to know and prepare them.

A few trivial mistakes:

1. Problems of package `babel`: system TeX support several different languages, that is rules of typography are embedded into TeX. Unfortunately using option `magyar` of package `babel` we can see that default files are wrong, so use prepared ones (see for details: [7] and [2])

2. In different distributions of TeX a few package are not embedded. Because of this do not use special packages to avoid mistakes during the interpretation of source.

3. There are a lot of old and out of date packages. Do not use them because they will induce incompatibility.

4. Take care of individual command definitions, because we can easily overwrite fundamental commands and it will lead to wrong working.

3. `Texpower`: a basic package

Using of slides has a lot of profitable properties:

- summary of knowledge
- partial summary
- excepting of important objects
- demonstration of an example
- considering each steps of an algorithm
- possibilities of differentiation
- placing of keywords and fundamental definitions in lecture

Before birth of `texpower` there was a class (`slides`) which was usable to make all slides of a lecture. Unfortunately in this case we did not have variable and dynamic elements to make interesting lectures. The package `texpower` has solved these problems and because of it this package is usable with several classes. Thus we can realize this property in all the examples below.

3.1. Settings

There are only a few data to set and these settings need minimal time, because TeX will set everything using only these few parameters. For example: for setting of layout we have to get only margins of top and bottom and margins of left and right. After this the layout will be determined automatically. Source is the following:

```
\renewcommand{\slidetopmargin}{3.7mm}
\renewcommand{\slidebottommargin}{0mm}
\renewcommand{\slideleftmargin}{5mm}
\renewcommand{\sliderightmargin}{5mm}
```

We can set also measure of magnification. Source:

```
\slidesmag{5} %%% Number of magnification: 5
```

Using this command we can use huge, large or small letters that is every student can see our lectures very well. Using of colours is very important in teaching. We can give a lot of special colour to make colourful objects on a slide. Source:

```
\pagecolor[rgb]{1,1,0.7}          %%% Colour of page
\definecolor{LB1}{rgb}{0.1,0.1,1} %%% Colour LightBlue
```

We have another commands to determine time of a slide that is we can make a time-table for lecture. It is very useful, because we can determine how many objects have to be placed on a slide (see [3]).

3.2. How to step?

In a lot of cases we need to step our objects e.g. equation, figure or picture. Practical investigations show that we should not use more than one objects or information together (because of students' concentration). So, this rule is true in LaTeX because of the following commands: `\pause` and `\stepwise` (or `\liststepwise`). It is important that these commands are usable for almost all objects, that is during our talk we can show the symbol or the word preferential on the slide. Consider these commands (see examples in section 5):

- Stepping of lists: use command `\pause` (see figures 9 and 10).
- Stepping (and highlighting) of equations: use command `\stepwise` (see figures 11 and 12).
- Highlighting (and stepping) of lists: use command `\liststepwise`. So, the information we talk is highlighted from the others (see figures 13).

4. Other packages and classes

In the following we will introduce a few other and easily usable styles to make professional lectures. We usually prepare format `pdf` of our lecture in a direct method (by `pdflatex`), but there have been exceptions. In many cases we have to make other conversions on the first `pdf` file to get the correct slides of lecture at last. One of these exceptions is style `ppower4`.

4.1. Package ppower4

The class of this package is `foils` which is ceirtainly usable with package `texpower` also. For making final `pdf` file at first we have to use `pdflatex` and after it a conversion of programming language `java` (see for more details in [6]). Consider a few tricks and solutions from this package:

– Settings of background: we give in command \`definecolor` the name of the colour (e.g. `blue`) and parameters in coding `rgb` (e.g. {1,0,0.6}). After this definition we can use the command \`vpagecolor` to change brightness of all slides or one slide in our lecture. Using colorful background is very important. For example: in investigations of teaching it was proven that colour blue is very useful in understanding of mathematical objects (see [1]). So, using command \`definecolor` we can make colourful slides which enhance understanding and attention of our lecture that is our slides will be well sctructured and aesthetic. Let us see an example (source):

```
\definecolor{blue}{rgb}{0.17,0.22,0.4}
\vpagecolor{blue}
```

– Pictures in the background: we can place arbitrary pictures in the background for all slides or only one. Use command \`bgaddcenter` to interest students for listening. Source and layout (figure 4):

```
\bgaddcenter{
\includegraphics[width=290truemm,height=240truemm]
{kep.jpg}}
```

Figure 4: Command \`bgaddcenter`

4.2. Package pdfscreen and pdfwin

If you want to find other colourful and dynamic styles, consider a few slides from examples of package **pdfscreen** and package **pdfwin**. All of these examples were made by class **article**. Let us consider one slide from these packages:

– First slide of **pdfscreen**: (figure 5)

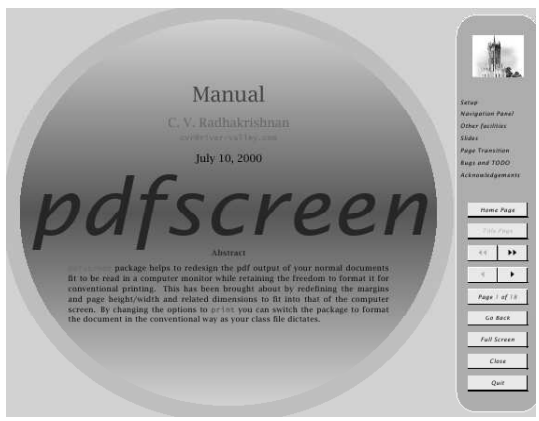

Figure 5: Package **pdfscreen**

– Frame and formulas from **pdfwin**: (figure 6)

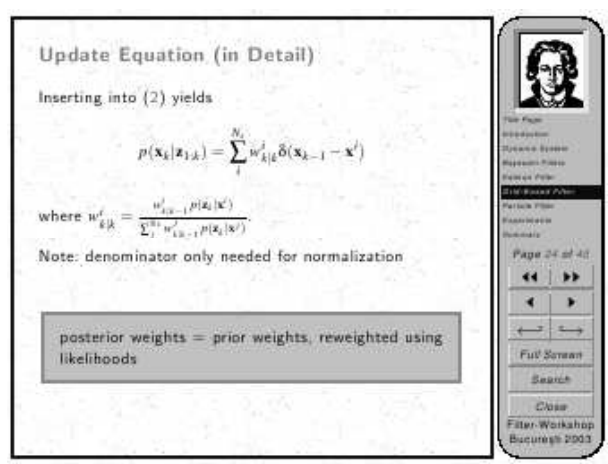

Figure 6: Package **pdfwin**

4.3. Package `prosper` and `ha-prosper`

Other interesting packages are `prosper` and `ha-prosper`. In these cases source have to be interpreted by method `latex` → `dvi` → `dvitopdf` and `latex` → `dvi` → `dvitops` → `pstopdf`. Two simple examples are the following:

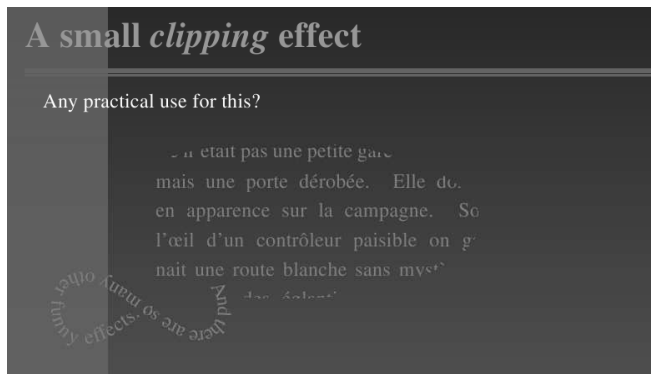

Figure 7: Package `prosper`

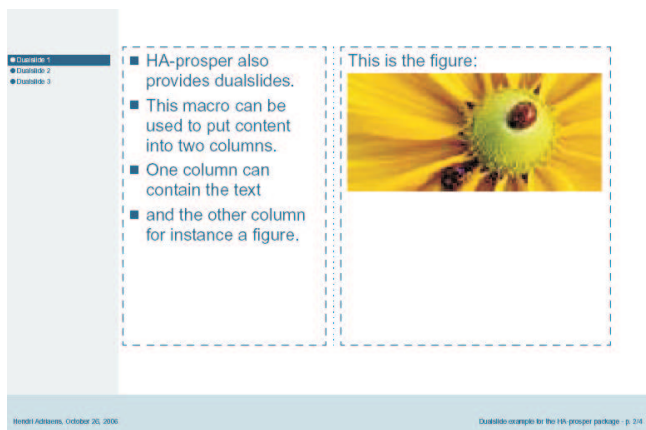

Figure 8: Package `ha-prosper`

We remark that colours and backgrounds are predefined in these packages that is we have to use only options to choose them.

5. Examples and experiences

Before subsections of examples and experiences it is important to know that the methods above and below have been used by authors since 2003. In several

conferences (e.g. Chech Republic, Netherland) and lectures (e.g Linear algebra, Applied Mathematics in Eszterházy Károly College) they were used to show the main objects and give simple and clear slides. So, in the following we get a summary of examples and experiences to show why we recommend to use LaTeX for making lectures.

5.1. Examples: `texpower`

Using tricks and commands of `texpower` consider the following examples. Use of command `\pause` is simple and easy. Source:

```
\begin{itemize}
\item foo\pause
\item bar\pause
\item baz
\end{itemize}
```

Layout of two slides (after one stepping):

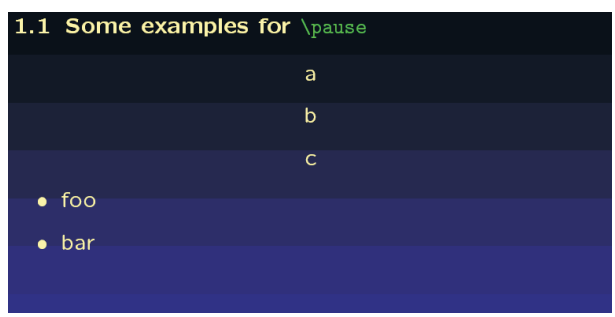

Figure 9: Command `\pause (first)`

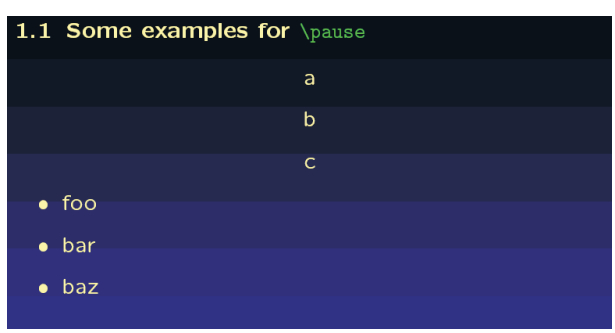

Figure 10: Command `\pause (second: after stepping)`

Consider the example of command `\stepwise` (higlighting of list and stepping parts of system of equations).

Layout of one slide:

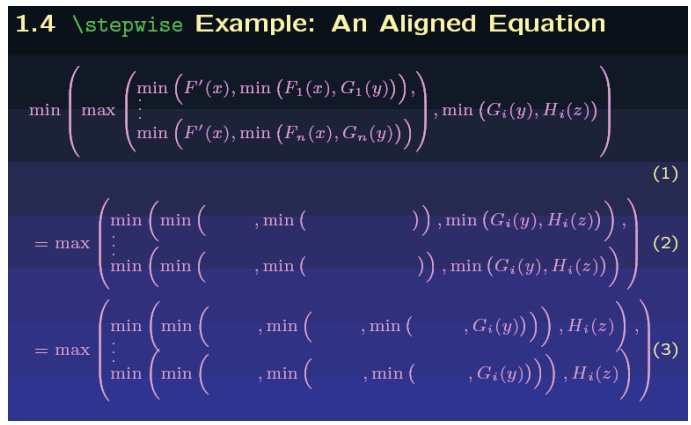

Figure 11: Command `\stepwise (first)`

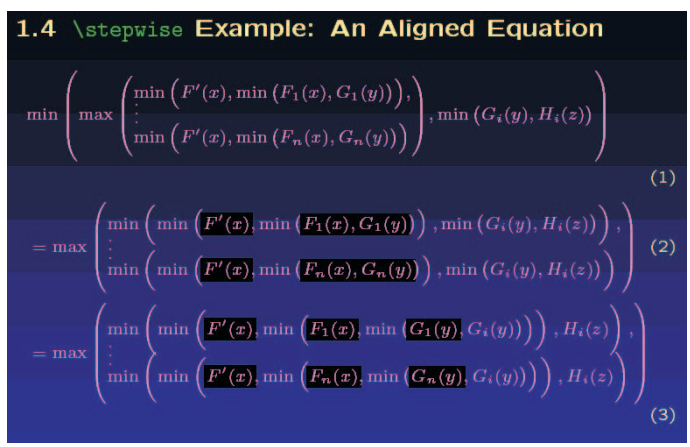

Figure 12: Command `\stepwise (second: after stepping)`

Consider the example of command `\liststepwise`. Source:

```
Instead of displaying incrementally, we can just 'flip through' some
items by highlighting them: \liststepwise* { \begin{stepitemize}
  \item Item 1
  \item Item 2
  \item Item 3
  \end{stepitemize} }
```

Layout of one slide:

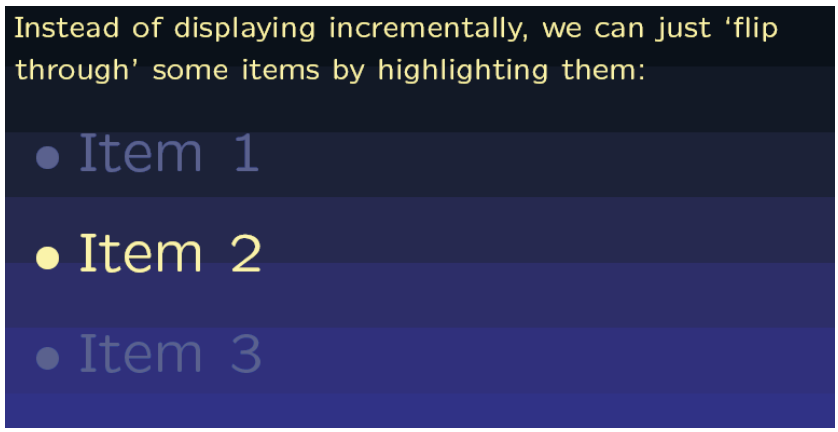

Figure 13: Command \liststepwise

5.2. Experiences of lectures

Our experiences are the following:

- Using this system and packages we have more time to talk about details of e.g. definitions or theorems. So, the first advantage is more time and if the slides are made by a known book, students will be able to follow the detailed explanations and they will not deal with writing. That is they will also have enough time to listen and understand our lecture.

- The second one is that the lecture is colourful. Using different colours for e.g. definitions or theorems students can easily see the difference between texts. Use of colours also helps them to study the important elements of the lecture, because they can remember the colour of an element.

- The next advantage is that our lecture is dynamic. We show elements of lecture step by step and due to it students can import and understand immediate information easier than by using older technology.

- The fourth one is the following: hardwer conditions of TEX are very basic, that is we can use almost any computer to make our lectures. So, to change the structure of a lecture or a property of an element from the lecture is really simple. For example: if you want to use larger letters, just modify argument of command \slidesmag and every letter will be changed. There is a simple question: **how to do this by PowerPoint?**

- Using pictures as backgrounds (different slide – different picture) students will be motivated for a long time. If we change pictures, students will not get tired or bored even after more than 2 hours.
- Combining LaTeX-presentations with using interactive board helps students clearly understan mathematical proofs. Since this technology does not break the mathematical text, the essential steps of proofs remain unseparated.
- Last one from experiences is that large lectures need a lot of slides because of stepping (about 400-500). Due to this large number we suggest to use simple package-class pair (e.g package `texpower` with class `powersem`). If only a few slides are needed (e.g. for a conference talk), it is useful to make it with another package-class pair (e.g. package `ppower4` with class `foils`). It was induced by time, because we do not usually have enough time to make it.

References

[1] Ambrus András, Matematikadidaktikai tanulmányok, *Tankönyvkiadó*, Budapest ELTE, (1989).

[2] BME math LaTeX, `http://www.math.bme.hu/latex/`

[3] Bruno Buchberger, Thinking Speaking Writing, *Springer-Verlag*, London Paris Tokyo. or instrumental music, (2005), `http://icking-music-archive.org/`.

[4] John D. Hobby, Drawing Graphs with METAPOST, (1990), `CTAN systems/msdos/metapost/doc/mpgraph.ps`. (2005).

[5] Michael Spivak, The joy of TeX, 2. kiad., *AMS*, (1990).

[6] TeX Catalogue, `http://www.tug.org/tex-archive/help/Catalogue/brief.html`

[7] Wettl Ferenc–Mayer Gyula–Szabó Péter, LaTeX kézikönyv, *Panem Kiadó*, Budapest, (2004).

Péter Olajos
Eszterházy Károly College,
Institute of Mathematics and Informatics,
P.O. Box 43
H-3300, Eger, Hungary

Erzsébet Orosz
Eszterházy Károly College,
Institute of Mathematics and Informatics,
P.O. Box 43
H-3300, Eger, Hungary

Annales Mathematicae et Informaticae
***33** (2006) pp. 189–205*
http://www.ektf.hu/tanszek/matematika/ami

Mathematics teachers and differentiation - results of a survey concerning Hungarian secondary schools

Réka Szász

Alfréd Rényi Institute of Mathematics
e-mail: reka@renyi.hu

Submitted 25 December 2005; Accepted 3 November 2006

Abstract

Differentiation is a method where teachers adapt their teaching to the needs of individual students. In Hungary it had not been an issues until recently, although it is becoming a more and more acute need. The paper presents the findings of a survey, the aim of which was to examine the present situation of differentiation in Hungarian secondary schools.

Keywords: differentiation

MSC: 97C90

1. Introduction

There is a diversity of students in each and every classroom: students differ in ability, aims, learning styles and many other factors. Hence teachers must adapt their teaching methods to the needs of individual students, which process is called differentiation. This involves setting activities where students may do different tasks, or tasks in a different way, and also using a variety of teaching styles, which ensures that each student profits from some of them. To realize this, teachers need appropriate professional skills, and also practical help, such as appropriate teaching material, or an assistant teacher.

In Hungary, differentiation has not been an issue until recently, although every teacher uses it intuitively to some degree. With increasing differences among students, the greater emphasis on equal opportunities, and the introduction of the two-level secondary final examination it is becoming a more and more acute need. The present National Curriculum emphasizes the need for differentiation both in

the description of its general principles and in its standards for Mathematics teaching [1]. However, the National Institute of Public Education (Országos Közoktatási Intézet) found in a survey on secondary mathematics teaching in 2003 that although teachers differentiate, they face both professional and practical difficulties concerning the issue [16].

This paper presents the findings of a survey, the aim of which was to examine the present situation of differentiation in Hungarian secondary schools. As the two-level final examination is in the centre of these issues, I introduce the acronyms CLFE and ELFE for the core and the extended level final examinations respectively.

2. Differentiated teaching - literature review

Here I give a short summary of the theory of differentiated teaching. A more detailed presentation can be found in [8].

2.1. The diversity of students

Results of several surveys, such as the international PISA [22], and the national MONITOR [21] showed that variation in mathematical attainment has increased in Hungary. Czeglédy [6] finds that there is a striking **heterogeneity** in students' mathematics achievement not only between, but within classes, as well.

There are many **factors** underlying mathematical achievement, which educators need to bear in mind when trying to cope with the diversity of students. Cangelosi [4] describes the following factors:

- **Mathematical competence**
  - Mathematical ability
  - Cognitive stage
  - Prior mathematical learning
  - Communication skills
- **Attitude towards mathematics**
  - Motivation
  - Self-confidence
- **Way of learning**
  - Learning style
  - Study skills
- **General factors**
  - Gender
  - Social background
  - Special needs

2.2. Differentiation as an answer to diversity

An obvious answer to diversity is ability grouping, which is strongly present in Hungarian secondary schools. There is much evidence, however, that this does not solve the problem ([14] and [11]), and it reinforces social inequalities [18]. The alternative to teaching the whole class as a unit is **differentiation**, 'the process of identifying, with each learner, the most effective strategies for achieving agreed targets' ([19], p. 129). We can realize differentiation from the following aspects [7]:

- **Content**
  - **Task**: students work on different tasks
  - **Pace of learning**: students work on the same tasks, but proceed at a different pace
  - **Outcome**: students work on a task that can be answered on different levels
- **Way of learning**
  - **Lesson form**: students can work individually, in pairs, or groups simultaneously
  - **Teaching method**: using a different form of dialogue
  - **Use of aids**: the quantity and nature of aids can be different
- **Checking and assessment**: checking work with different thoroughness, setting different requirements

Differentiation requires a great amount of attention and work from teachers, but there are certain tools they can make use of. It is essential to have an appropriate **curriculum**, which identifies core material, to be acquired by every student, and enrichment topics for those who are able to move faster [12] (in Hungary, core topics include the CLFE requirements, and enrichment topics can be ELFE requirements, cultural topics, etc.). The most important help is a **textbook** that is built on the curriculum and takes every aspect of differentiation into consideration [12]. Other tools are **assistant teachers** (student teachers or volunteers) [9] and **computers** [2], which make all of the above mentioned aspects of differentiation easier. I should also add that certain **lesson forms** are more appropriate for differentiation than others. Individual work and homogeneous pairing or grouping is a way to realise differentiation by task or pace, while heterogeneous pairs and groups facilitate differentiation by outcome. In the latter case, students either solve subtasks of different difficulty, or take roles that require different skills. They can also teach each other: more advanced students deepen their knowledge by explaining to their peers, while these benefit from their explanation. Whole-class work, on the other hand, is quite inappropriate for differentiation [15].

Finally, as educators cannot differentiate all the time and in every aspect of teaching, they should teach in a way that is **appropriate to a diversity of**

students. This means using a variety of instructional methods and materials [17], and **real-life** context, which is able to bridge the gap between different students [18].

3. Teachers in Hungary and differentiation

This section presents the results of the survey, which was carried out in 2005, and involved a representative set of secondary mathematics teachers in Hungary (details of methodology are described in the next section). The aim was to find out how acute the problem of diversity is, how teachers handle it, and what kind of help they need most.

3.1. Diversity of students

First, I wanted to know about the extent and nature of heterogeneity in classrooms, as this is the major motivation for differentiation. Teachers turned out to teach in class sizes with an average of 20. In years 9 and 10 many of them teach whole classes with more than 35 students, but many other schools are fortunate enough to have smaller groups from the beginning. In years 11 and 12 groups are usually smaller, especially special (fakultációs) groups preparing for the ELFE.

I also asked teachers to rate the **heterogeneity** of their groups from 1 (homogeneous) to 5 (very heterogeneous). Table 1 shows the results and a fitted bell-curve. The mean is 3.53 with a standard deviation of 1.04. So we can say that teachers

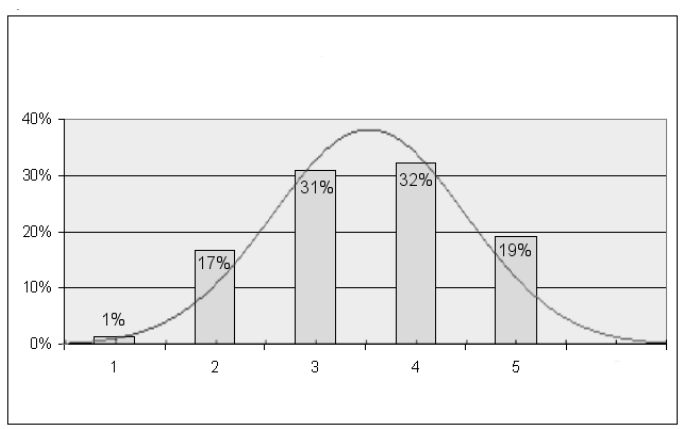

Table 1: Heterogeneity of groups.

perceive their groups to be quite heterogeneous. Some teachers voiced difficulties caused by heterogeneity in words, too:

- It is almost impossible in groups of 35 students [to handle students with different aims]
- In year 9 I give separate worksheets to those behind and good students. It takes a lot of time to design, correct, and assess these.
- It is difficult [to handle student differences]. I have to work more.
- Sometimes I feel that it is unmanageable! [about handling differences]

I also asked teachers what **factors** of difference cause the greatest difficulties in teaching. Let me first present Cangelosi's [4] categories adding a percentage value of corresponding factors in teachers' answers (the value of the main categories includes answers relating to the category as a whole besides those relating to sub-categories):

- **Mathematical competence 63%**
  - mathematical ability 32%
  - cognitive stage 1%
  - prior mathematical learning 25%
  - communication skills 3%
- **Attitude towards mathematics 26%**
  - motivation 18%
  - self-confidence
- **Way of learning 8%**
  - learning style 7%
  - study skills 1%
- **General factors 3%**
  - gender
  - social background 2%
  - special needs

As *mathematical competence* has the biggest weight, and because many teachers referred to skills that cannot be related to any of its sub-categories, I divided it into ability&cognitive stage and knowledge&skills (including Cangelosi's *prior mathematical learning, communication skills*, and *mathematical competence in general*). Table 2 shows teachers' answers clustered this way. I find these results important for two main reasons. On one hand, they help in devising effective aids to differentiation. On the other hand, it draws attention to factors that teachers do not recognise enough, or at all. They tend to look at their students either from the

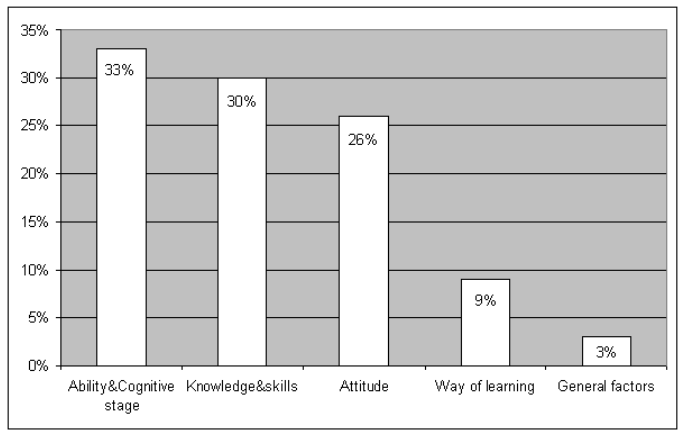

Table 2: Differences between students.

side of output (knowledge&skills), or they simplify input as being mostly ability and attitude (and they often regard negative attitude the fault of the student). However, some of the factors they recognise less are such that they could have a positive influence on, such as communication skills (very important in the new final examination), self-confidence, and study skills. It would also be important for teachers to be more aware of students' social background and learning styles, in order to motivate them effectively and to find suitable ways to teach them. It is natural that teachers do not mention special needs students, as they are not usually integrated into secondary teaching at the moment, but later teachers will have to care for them, too.

Another interesting aspect of teachers' view on student differences appears in their definitions of differentiation, where a lot of them indicate some kind of factor as a motive for differentiation. I put these in the same clusters adding *goal* as a new category (although this is part of motivation I kept it separate to show its appearance). Table 3 compares the importance of factors when teachers look at differences pragmatically (report on their students) and theoretically (giving a definition). It is apparent that in theory teachers recognise a much narrower spectre of differences. While knowledge&skills have the same weight as before, the importance of input factors shifts almost exclusively to ability. However, goal appears as a new factor, which teachers probably recognise because of the two-level final examination. The fact that they do not list it in pragmatics shows that probably they do not distinguish between their students according to their goals, which becomes apparent in the next section indeed.

On the whole, the task seems to be to widen teachers' view on the factors of student differences. To this teacher training (both pre- and in-service) can have the greatest contribution. Although this is outside the reach of this study, I would like to state is an important future goal. But this fact also highlights the importance

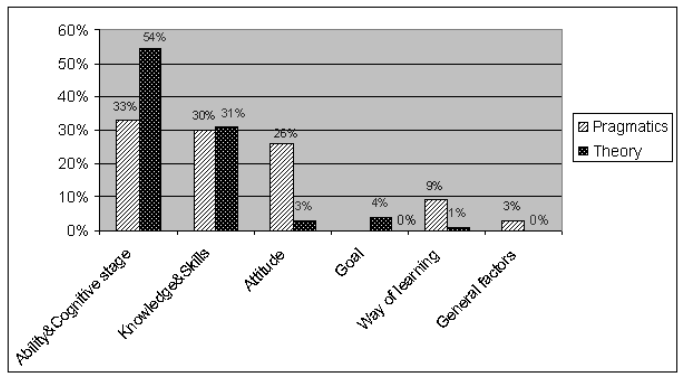

Table 3: Teachers' view on student differences.

of designing teaching materials that serve a wide variety of students even without teachers specifically aiming at this.

3.2. Extent and ways of differentiation

To the first glance the results of this question are positive: two-thirds of teachers differentiate to some extent. Let us take a closer look. The questionnaire had two questions strongly related to differentiation, a pragmatic one ("How do you handle these differences?"), and a theoretical one ("What do you think the term differentiation means?"). As in the case of diversity, the pragmatic question is closer to real teaching, and it gives more information, so I will start with that and analyse the results more deeply.

It is difficult to calculate a reliable percentage of teachers who differentiate, but the answers suggest that 63% of teachers asked do differentiate, and 29% do not differentiate in their lessons (but 17% among them differentiate in homework, or give extra-curricular remediation or enrichment classes). It is more informative to look at all the methods they mention, and compare the frequency of these, as shown in Table 4. Again, numbers suggest that differentiation is widely used to handle differences both in and outside lessons, and teachers also use other methods to handle differences. Before analysing them, let us take a closer look at what kind of answers these terms cover. *Differentiation by task* usually means that stronger students work on difficult problems while weaker ones practice or get further teacher explanation. *Differentiation by pace* means that faster students get extra problems to solve. *Differentiation by outcome/role* means using a lesson form where students profit from the same activity in a different way, e.g groupwork, study-pairs, or presentations. *Differentiation by teaching methods* includes techniques when teachers show other solutions to stronger, pitfalls to weaker students, or keep stronger control over less diligent ones. *Motivation* includes keeping students' interest, helping them to have a sense of achievement, giving good points, or convincing them that

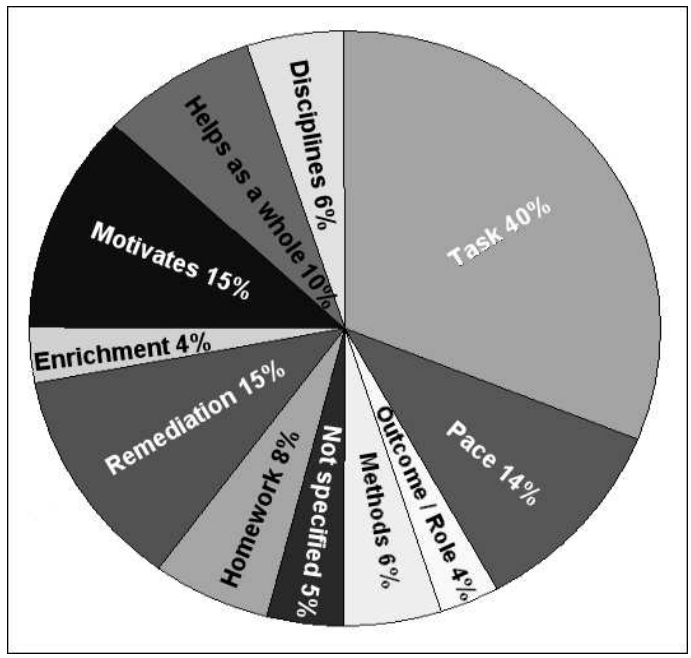

Table 4: Handling differences.

hard work leads to success. *Helping the group as a whole* includes re-teaching and revision, extra practice, answering questions, or interpreting problems together.

To analyse ways of differentiation it is interesting to compare the above pragmatic answers to teachers' theoretical answers, as shown in Table 5. (Note that although teachers worded the answers to the two questions quite differently, the results are very similar). Differentiation by task is naturally the most popular way of differentiation, followed by differentiation by pace. Differentiation by level/material can be realised by task and pace, I will deal later with the fact that it is only present in theory. Differentiation by outcome, although used in practice, is not present in theoretical answers, probably because these are methods teachers use incidentally. Although from the above answers it seems that teachers recognise differentiation by lesson form in theory only, from the answers to other questions it is obvious that they use it in practice, too[1].

A relatively high number of teachers mention differentiation by teaching methods both in theory and practice, while that by use of aids is missing, maybe because the scarce use of aids in secondary school, anyway. *Differentiation by checking answers* is a way that would deserve more emphasis, although teachers might use

[1]In their answers to the question about lesson forms (later in this section) 3% of teachers explain that stronger students work in pairs on difficult problems while others work with the teacher, or individually.

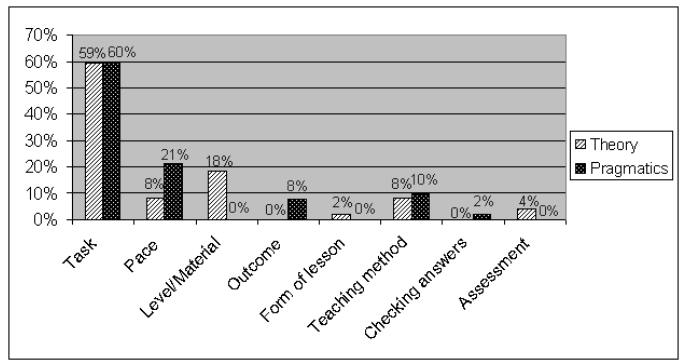

Table 5: Ways of differentiation.

it when differentiating by task without mentioning it explicitly. *Differentiation by assessment* is only recognised in theory, this is mainly a problem concerning groups preparing for two different levels of the final examination.

Let me now examine the different tools of differentiation. As the **differentiated curriculum** is concerned, it seems that teachers do not use it widely. The first evidence is the fact that differentiation by task/material only appears in theory. From the descriptions I presented for differentiation by task and by pace, it is also clear that when students are involved in different activities, they work on problems of different difficulty within the same material. Teachers do not mention teaching different material to students, or even giving strong students problems that lead to new material. This fact is also made clear by the lack of differentiation by assessment in practice. Another strong evidence for the lack of differentiated curriculum can be seen from the following question:

> *Do you have a group where students want to reach different goals (e.g a year 11 or 12 group which involves students preparing for both levels of the final exam)? If yes, how do you handle the situation?*

30% of teachers answered yes, and these all teach groups preparing for both levels for the final examination. Table 6 summarises their methods. Teachers who differentiate do it by giving different problems to students. Those who do not differentiate prepare for the level most students need, or to a level in between the two. The high percentage of those who do not differentiate in a situation where the need is so evident, clearly shows their inability to use a differentiated curriculum.

Only 11% of teachers have someone **assisting** in part of their lessons, and this is always a teacher trainee. Only 28% of teachers use a **computer**, and part of them only as a tool for representation through a projector. Many of those who do not, refer to the absence of appropriate facilities.

Concerning **methods appropriate for a diversity of students**, let us look back at results shown in the previous subsection. I found that Ability&Cognitive

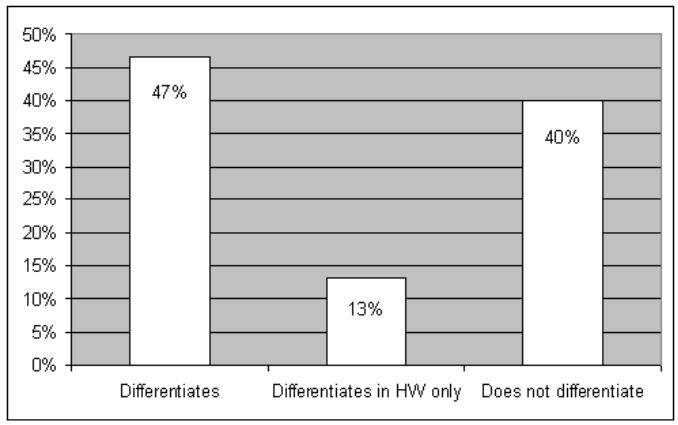

Table 6: Groups preparing for two levels.

Stage, Knowledge&Skills, and Attitude are the categories that weigh most, while Way of learning and General Factors also have a role. Let us examine the connection of these to their methods of differentiation. With students differing in the first two big categories, that is in *mathematical competence*, clearly *differentiation by content* is appropriate, which teachers use a lot. However, Table 7 shows that teaching is more appropriate for weaker student than for stronger ones. These numbers were

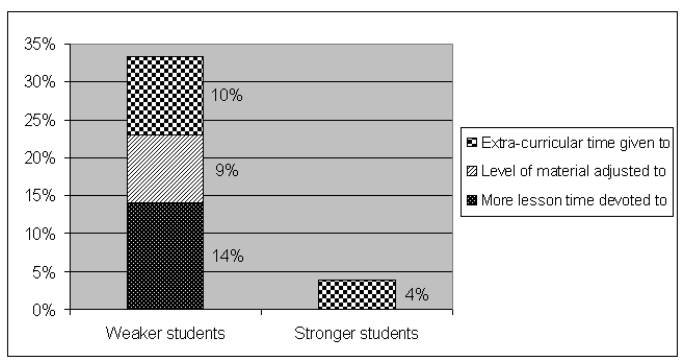

Table 7: Adjusting to diverse needs in mathematical competence.

all inferred from the way teacher handle differences. Those who devote more time to weaker students differentiate by task or by pace, but leave stronger students working on extra problems while they spend extra time with weaker ones. Those who adjust material to the level of weaker students either say so explicitly, or belong to teachers who *help the group as a whole* by re-teaching, revision, extra practice,

answering questions, and interpreting problems together. In sum, teachers who favour one level with their teaching methods all help weaker students, but the rest seem to care equally about their students.

Teachers seem to handle differences in *attitude* effectively, as *motivating* is an important tool. Differentiation by *teaching methods* seems appropriate for students with a different *way of learning*, or differences in *general factors*, of course there are many other ways to do this.

On the whole, the recognition of factors and the connected methods of handling differences seem to correlate a lot. We can interpret this statement both in a positive and in a negative way. From an optimistic point of view, teachers find appropriate methods to handle student differences. From a pessimistic one, teachers naturally handle factors of difference they recognise, hence it is a great problem that, as I observed before, that they are not aware of many of them. Finally, let me examine the **lesson forms** teachers use, as this tells us a lot about the extent and way they differentiate. I asked teachers to tell how much part of the lesson they spend with each form, and what kind of activities they use these for. Table 8 shows each activity type within a lesson form as a percentage of all lesson time. Most time

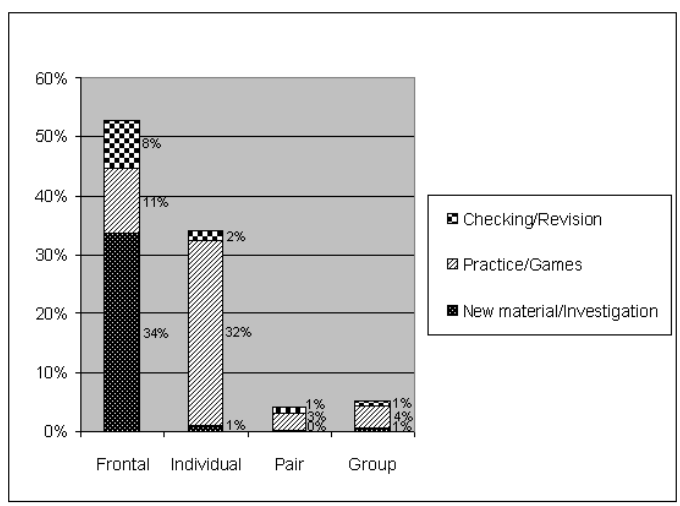

Table 8: Lesson forms - Activities.

is spent with whole-class work, and this is basically the only way students acquire new material, check answers or revise. This means, that teachers think their person as an active participant is indispensable in such activities. From a general point of view, this is a problem because student activity is quite low during whole-class work. And from the present point of view, it makes differentiation impossible [15]. Practice is mainly individual, when, as we saw before, teachers use differentiation by task or pace. A quite large proportion of practice is done with the whole class,

partly in form of solving model problems, but some of the teachers simply use this lesson form for practice, too. In my opinion it would be favourable if teachers used less whole-class, and more pair- and groupwork, thus enabling their students to work according to their needs, and also to substitute teachers by taking a role of explainer [ibid.].

Now let us look at activities more closely. Table 9 shows each activity type and teachers' favoured lesson forms. Teaching new material, as we saw before,

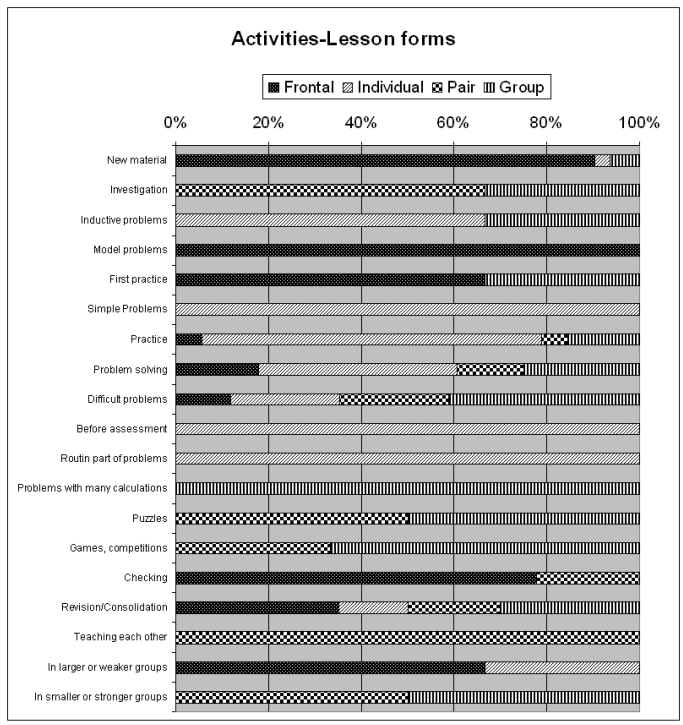

Table 9: Activities - Lesson forms.

is mostly whole-class, but some teachers also let their students investigate and solve inductive problems. As we look at different problem solving activities, we see that while the group advances from model problems to first practice, simple problems, and practice, teachers let the control out of their hands and trust their students doing individual work and a little pair- and groupwork. Then, as they get to problem solving and difficult problems, they either get the control back, or let students help each other in pairs and groups. The second version is of course more favourable for differentiation. Then, we can see that teachers favour certain lessons forms with certain activities. For instance, they trust their students to solve routine parts of problems individually, they want them to share work with problems requiring many calculations, and favour pair- and groupwork with fun,

such as puzzles, games and competitions. Checking work is mostly whole-class, but teachers recognise the other forms for revision and consolidation, and a few also the effectiveness of peer-teaching. Finally, it is interesting that teachers prefer whole-class and individual work with larger or weaker groups, presumably because it is easier for them to keep control that way. However, it would be worth teaching their students how to work in pairs and groups and accept less control, as that would mean more activity and attention (from peers) for their weaker students, and members of large groups, too [15].

4. Methods of data collection and analysis

4.1. Designing the questionnaire

The content of the questionnaire was based on the theory presented in Section 2. The first question,

> *1. Have you heard of the term* differentiation *(or differentiated teaching)? If yes, what do you think it means?*

wanted to find out the theoretical knowledge of teachers about differentiation. The next two questions,

> *2. What is the average class size you teach?*
>
> *3. (a) How heterogeneous are the classes you teach? 1 2 3 4 5*
> *(1: homogenous, 5: very heterogeneous)*
> *(b) What kind of differences cause the greatest difficulties in mathematics lessons?*

aimed at finding out the acuteness of the need for differentiation. Question 4 asked about how teachers differentiate in practice,

> *4. How do you handle these differences?*

and question 5 asked about the specific case of preparing for the two-level final examination:

> *5. Do you have a group where students want to reach different goals (e.g a year 11 or 12 group which involves students preparing for both levels of the final exam)? If yes, how do you handle the situation?*

Questions 6, 7 and 9 asked if specific aids are available or used:

> *6. Which textbooks do you use?*
>
> *7. Do you use a computer in your lessons? If yes, how often and what for?*

8. How much, and for what activities do you use the following lesson forms:

whole-class work:

individual work:

pairwork:

groupwork:

9. Does it happen that someone attends to students besides you during lessons (assistant, teacher trainee, parent, etc.)?

4.2. Methods of data collection

The first step of designing the survey was selecting a representative sample of Hungarian secondary schools. I selected 25 schools (4% of all 636 secondary schools) using the database of the Information Office of Education (Közoktatási Információs Iroda) [10]. The database can be searched according to school type (secondary in the present case), county and maintainer (community council, city council, council of town of county rack, county council, capital council, capital district council, church, foundation, state university, etc). My aim was to select a sample where the number of schools from each county and maintainer is proportional to the corresponding number of all secondary schools. The number of secondary schools in each county is about 25, except for the following cases. In four counties it is less, but I could pair them up so as to get around 25 together. In the county of Pest the number is double, and in the capital it is 7 times as much. So I selected one school from each average size county, 1 from each pair of small ones, 2 from Pest and 7 from the capital. I selected these schools randomly: selected a random number between 1 and the length of the list, and chose the school with that serial number in the list. Then I had to make sure that the sample was representative of maintainers. If a maintainer was over-represented, I dropped a random school of that maintainer, and selected a random school (regardless of maintainer) from the county dropped. After repeating this step a few times I obtained a set that was fully representative.

The next step was to have mathematics teachers of the selected schools fill out the questionnaire. Larger schools teaching more students should have had a greater weight, which would naturally happen as they would have more mathematics teachers asked. So I contacted the head of mathematics department in each school, and sent them the appropriate number of questionnaires. Teachers of 21 schools (84% of the ones asked) filled out the questionnaire, 78 teachers altogether.

4.3. Methods of analysis

The analysis of questions 2, 3/a, 6, 7 and 9 is quite obvious. The methods used with the rest are presented in the order that follows their analysis in the previous section. Student differences appeared in question 3, and partly in question 1, and were presented in Table 2 and Table 3. For each question, I counted the differences

that occurred in the answers, and calculated the ratio of the number of occurrences and the total number of occurrences. Hence the percentages obtained were a weight in that factor within all factors, and not a percentage of teachers who listed that factors. When making clusters of categories, these numbers better reflected the importance of each category.

When analysing **differentiation** based on questions 1, 3 and 4, (Table 4 and Table 5) values were calculated as a percentage of teachers using each method, as I was interested not only in weights, but the ratio of teachers using certain methods. These values, however, are not additive, so for clustering I had to recalculate values.

Lesson forms were analysed based on question 8, and represented in . First I averaged percentages teachers gave for time spent on each lesson form. Within each form, I counted the occurrences of activities, and calculated their ratio within all occurrences. I used these ratios as estimates of time spent with these activities within a certain lesson form. Multiplying activity ratios within a form with the time ratio of the given lesson form, I obtained an estimate of the percentage of lesson time spent with the given activity in the given form. These values are presented in Table 9. Table 8 presents lesson forms within each activity. Unlike in the previous table, I used a 100% stacked chart. Although this is less informative then a plain bar chart, as it does not show how time spent with each activity relates to each other, I used it because otherwise some bars would have been too small to see well.

5. Conclusion

On the whole, I found that a large part of teachers differentiate, and in many ways they differentiate effectively. However, I also found ground for change.

First, most teachers do not differentiate the curriculum, although this would be essential with the two level final examination, as students study together until year 10, and sometimes even after that. Then, teachers are not aware of many factors in which students differ, and hence probably do not teach appropriately to these. And some do not even teach appropriately to students differing in the most recognised factor, mathematical competence, as they favour weaker students with their teaching methods. Finally, examining lesson forms, I found that they use whole-class work in the greatest part of lesson time, which renders differentiation impossible. At the same time, they scarcely use pair- and groupwork, although these are very favourable forms for differentiation.

Many of these problems stem for teaching traditions in Hungary, which are quite achievement-oriented and favour whole-class teaching [3]. Part of the change is needed to be made in teachers' attitude, which is the task of teacher training (both pre- and in-service) [13]. The other, equally important part of the change would be in the materialistic situation of teaching, that is, the circumstances and aids available for teacher, such as printed and electronic teaching materials, assistant teachers, IT facilities. There are initiatives to give such practical help. One example is a set of textbook that has appeared recently with the aim of supporting differentiation (Mathematics by Czeglédy, Hajdu, Hajdu, Kovács & Róka) [8] and

[5]. Another example is an initiative to have teacher trainees assisting in classrooms [20].

References

[1] *A Kormány 243/2003. (XII.17.) Kormányrendelete a Nemzeti alaptanterv kiadásáról, bevezetéséről és alkalmazásáról*, www.om.hu/main.php?folderID=391&articleID=1478&ctag=articlelist&iid=1.

[2] Ainly, J., Adjusting to the newcomer: Roles for the computer in mathematics classrooms, in *Issues in Mathematics Teaching*, Peter Gates (Ed.), Routledge, London, (2001).

[3] Benda, J., A kooperatív pedagógia szocializációs sikerei és lehetőségei Magyarországon I., *Új Magyar Pedagógiai Szemle*, September, (2002), 26-37.

[4] Cangelosi, J. S., *Teaching mathematics in secondary and middle school: an interactive approach*, (2nd ed.), Merill, Englewood Cliffs, (1996).

[5] Czeglédy, I., Hajdu, S., Hajdu, S. Z., Róka, S., Kovács, A., *Matematika, A Felzárkóztatástól a Tehetséggondozásig*, Budapest, Műszaki Könyvkiadó, (2003).

[6] Czeglédy, I., Matematika tantárgy mérés 7. osztályban, *Miskolci Pedagógus*, 26, (1990), 10-16.

[7] Czeglédy, I., Orosz, Gy., Szalontai, T., Szilák, A., *Matematika tantárgypedagógia I.*, Bessenyei György Könyvkiadó, Nyíregyháza, (2000).

[8] Czeglédy, I., Szász, R., The mathematics textbook as an aid to differentiation: A first Hungarian example, *Teaching Mathematics and Computer Science*, 3/1, (2005), 35-53.

[9] Hedrick, W. B., Pre-service teachers tutoring one-on-one within the school setting, *Reading Research and Instruction* 38/3, (1999), 211-19.

[10] *Közoktatási Információs Iroda*, www.kir.hu/intezmeny.

[11] Kulik, C. C., Kulik, J. A., Effects of Ability Grouping on Secondary School Students: A Meta-analysis of Evaluation Findings, *American Educational Research Journal*, 19/3, (1982), 415-428.

[12] Meiring, S. P., Rubenstein, R. N., Schultz, J. E., de Lange, J., Chambers, D. L., *Curriculum and Evaluation Standards for School Mathematics, Addenda Series, Grades 9-12: A Core Curriculum*, National Council of Teachers of Mathematics, Reston, Virginia, (1992).

[13] Poór, Z., Pedagógusképzés és -továbbképzés a változó pedagógusszerepek tükrében, *Új Magyar Pedagógiai Szemle*, May (2003), 50-54.

[14] Slavin, R. E., Achievement Effects of Ability Grouping in Secondary Schools: A Best Evidence Synthesis, *Review of Educational Research*, 60/3, (1990), 471-499.

[15] Slavin, R. E., Research on Cooperative Learning and Achievement: What We Know, What We Need to Know Contemporary Educational Psychology, *Review of Educational Research*, 21/1, (1996), 43-69.

[16] Somfai, Zs., A matematikatanítás helyzete a középiskolában - A 2003-as obszervációs felmérés tapasztalatai, *Országos Közoktatási Intézet*, www.oki.hu/oldal.php?tipus=cikk&kod=kozepfoku-somfai-matematikatanitas

[17] Stiff, L. V., Johnson, J. L., Johnson, M. R., Cognitive Issues in Mathematics Education, in: *Research Ideas for the Classroom: High School Mathematics*, (P. S. Wilson, ed.), Macmillan, New York, (1993).

[18] Stiff, L. V., Introducion, Reaching All Students: a Vision of Learning Mathematics, in: *Reaching All Students with Mathematics*, G. Cuevas and M. Driscoll (Eds.), The National Council of Teachers of Mathematics, Virginia, (1993), 3-16.

[19] Strading, R., Saunders, L., Weston P., *Differentiation in Action: a Whole School Approach for Raising Attainment*, HMSO, London, (1991).

[20] Szász, R., The mathematics teacher trainee as an assistant teacher, *Teaching Mathematics and Computer Science*, 3/2 (2005), 295-306.

[21] Vári, P., *Monitor '95*, Országos Közoktatási Intézet, Budapest, (1997).

[22] Vári, P., *PISA-vizsgálat 2000*, Műszaki Könyvkiadó, Budapest, (2003).

Réka Szász
Alfréd Rényi Institute of Mathematics
P.O. Box 127
H-1053 Budapest
Hungary