ANNALES MATHEMATICAE ET INFORMATICAE borító külső oldala

Contents

Methodological papers

ANNALES MATHEMATICAE ET INFORMATICAE 43. (2014)

ANNALES MATHEMATICAE ET INFORMATICAE

TOMUS 43. (2014)

HUNGARIA, EGER

ANNALES MATHEMATICAE ET INFORMATICAE borító belső oldala

ANNALES MATHEMATICAE ET INFORMATICAE

International journal for mathematics and computer science

Referred by
Zentralblatt für Mathematik
and
Mathematical Reviews

The journal of the Institute of Mathematics and Informatics of Eszterházy Károly College is open for scientific publications in mathematics and computer science, where the field of number theory, group theory, constructive and computer aided geometry as well as theoretical and practical aspects of programming languages receive particular emphasis. Methodological papers are also welcome. Papers submitted to the journal should be written in English. Only new and unpublished material can be accepted.

Authors are kindly asked to write the final form of their manuscript in LaTeX. If you have any problems or questions, please write an e-mail to the managing editor Miklós Hoffmann: `hofi@ektf.hu`

The volumes are available at `http://ami.ektf.hu`

ANNALES MATHEMATICAE ET INFORMATICAE

VOLUME 43. (2014)

INSTITUTE OF MATHEMATICS AND INFORMATICS
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: Czeglédi László
Műszaki szerkesztő: Tómács Tibor
Megjelent: 2014. december Példányszám: 30

Készítette az
Eszterházy Károly Főiskola nyomdája
Felelős vezető: Kérészy László

Annales Mathematicae et Informaticae
43 (2014) pp. 3–12
http://ami.ektf.hu

The log-concavity and log-convexity properties associated to hyperpell and hyperpell-lucas sequences

Moussa Ahmia[ab], Hacène Belbachir[b], Amine Belkhir[b]

[a]UFAS, Dep. of Math., DG-RSDT, Setif 19000, Algeria
ahmiamoussa@gmail.com

[b]USTHB, Fac. of Math., RECITS Laboratory, DG-RSDT,
BP 32, El Alia 16111, Bab Ezzouar, Algiers, Algeria
hacenebelbachir@gmail.com or hbelbachir@usthb.dz
ambelkhir@gmail.com or ambelkhir@usthb.dz

Submitted July 22, 2014 — Accepted December 12, 2014

Abstract

We establish the log-concavity and the log-convexity properties for the hyperpell, hyperpell-lucas and associated sequences. Further, we investigate the q-log-concavity property.

Keywords: hyperpell numbers; hyperpell-lucas numbers; log-concavity; q-log-concavity, log-convexity.

MSC: 11B39; 05A19; 11B37.

1. Introduction

Zheng and Liu [13] discuss the properties of the hyperfibonacci numbers $F_n^{[r]}$ and the hyperlucas numbers $L_n^{[r]}$. They investigate the log-concavity and the log convexity property of hyperfibonacci and hyperlucas numbers. In addition, they extend their work to the generalized hyperfibonacci and hyperlucas numbers.

The *hyperfibonacci* numbers $F_n^{[r]}$ and *hyperlucas* numbers $L_n^{[r]}$, introduced by Dil and Mező [9] are defined as follows. Put

$$F_n^{[r]} = \sum_{k=0}^{n} F_k^{[r-1]}, \quad \text{with} \quad F_n^{[0]} = F_n,$$

$$L_n^{[r]} = \sum_{k=0}^{n} L_k^{[r-1]}, \quad \text{with} \quad L_n^{[0]} = L_n,$$

where r is a positive integer, and F_n and L_n are the Fibonacci and Lucas numbers, respectively.

Belbachir and Belkhir [1] gave a combinatorial interpretation and an explicit formula for hyperfibonacci numbers,

$$F_{n+1}^{[r]} = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n+r-k}{k+r}. \tag{1.1}$$

Let $\{U_n\}_{n\geq 0}$ and $\{V_n\}_{n\geq 0}$ denote the generalized Fibonacci and Lucas sequences given by the recurrence relation

$$W_{n+1} = pW_n + W_{n-1} \ \ (n \geq 1), \ \ \text{with} \ \ U_0 = 0, \ U_1 = 1, \ V_0 = 2, \ V_1 = p. \tag{1.2}$$

The Binet forms of U_n and V_n are

$$U_n = \frac{\tau^n - (-1)^n \tau^{-n}}{\sqrt{\Delta}} \ \ \text{and} \ \ V_n = \tau^n + (-1)^n \tau^{-n}; \tag{1.3}$$

with $\Delta = p^2 + 4$, $\tau = (p + \sqrt{\Delta})/2$, and $p \geq 1$.

The generalized hyperfibonacci and generalized hyperlucas numbers are defined, respectively, by

$$U_n^{[r]} := \sum_{k=0}^{n} U_k^{[r-1]}, \quad \text{with} \quad U_n^{[0]} = U_n,$$

$$V_n^{[r]} := \sum_{k=0}^{n} V_k^{[r-1]}, \quad \text{with} \quad V_n^{[0]} = V_n.$$

The paper of Zheng and Liu [13] allows us to exploit other relevant results. More precisely, we propose some results on log-concavity and log-convexity in the case of $p = 2$ for the hyperpell sequence and the hyperpell-lucas sequence.

Definition 1.1. Hyperpell numbers $P_n^{[r]}$ and hyperpell-lucas numbers $Q_n^{[r]}$ are defined by

$$P_n^{[r]} := \sum_{k=0}^{n} P_k^{[r-1]}, \quad \text{with} \quad P_n^{[0]} = P_n,$$

$$Q_n^{[r]} := \sum_{k=0}^{n} Q_k^{[r-1]}, \quad \text{with} \quad Q_n^{[0]} = Q_n,$$

where r is a positive integer, and $\{P_n\}$ and $\{Q_n\}$ are the Pell and the Pell-Lucas sequences respectively.

Now we recall some formulas for Pell and Pell-Lucas numbers. It is well know that the Binet forms of P_n and Q_n are

$$P_n = \frac{\alpha^n - (-1)^n \alpha^{-n}}{2\sqrt{2}} \text{ and } Q_n = \alpha^n + (-1)^n \alpha^{-n}, \tag{1.4}$$

where $\alpha = (1 + \sqrt{2})$. The integers

$$P(n,k) = 2^{n-2k} \binom{n-k}{k} \text{ and } Q(n,k) = 2^{n-2k} \frac{n}{n-k} \binom{n-k}{k}, \tag{1.5}$$

are linked to the sequences $\{P_n\}$ and $\{Q_n\}$. It is established [2] that for each fixed n these two sequences are log-concave and then unimodal. For the generalized sequence given by (1.2), also the corresponding associated sequences are log-concave and then unimodal, see [3, 4].

The sequences $\{P_n\}$ and $\{Q_n\}$ satisfy the recurrence relation (1.2), for $p = 2$, and for $n \geq 0$ and $n \geq 1$ respectively, we have

$$P_{n+1} = \sum_{k=0}^{\lfloor n/2 \rfloor} 2^{n-2k} \binom{n-k}{k} \text{ and } Q_n = \sum_{k=0}^{\lfloor n/2 \rfloor} 2^{n-2k} \frac{n}{n-k} \binom{n-k}{k}. \tag{1.6}$$

It follows from (1.4) that the following formulas hold

$$P_n^2 - P_{n-1}P_{n+1} = (-1)^{n+1}, \tag{1.7}$$

$$Q_n^2 - Q_{n-1}Q_{n+1} = 8(-1)^n. \tag{1.8}$$

It is easy to see, for example by induction, that for $n \geq 1$

$$P_n \geq n \text{ and } Q_n \geq n. \tag{1.9}$$

Let $\{x_n\}_{n\geq 0}$ be a sequence of nonnegative numbers. The sequence $\{x_n\}_{n\geq 0}$ is *log-concave* (respectively *log-convex*) if $x_j^2 \geq x_{j-1}x_{j+1}$ (respectively $x_j^2 \leq x_{j-1}x_{j+1}$) for all $j > 0$, which is equivalent (see [5]) to $x_i x_j \geq x_{i-1}x_{j+1}$ (respectively $x_i x_j \leq x_{i-1}x_{j+1}$) for $j \geq i \geq 1$.

We say that $\{x_n\}_{n\geq 0}$ is *log-balanced* if $\{x_n\}_{n\geq 0}$ is log-convex and $\{x_n/n!\}_{n\geq 0}$ is log-concave.

Let q be an indeterminate and $\{f_n(q)\}_{n\geq 0}$ be a sequence of polynomials of q. If for each $n \geq 1$, $f_n^2(q) - f_{n-1}(q)f_{n+1}(q)$ has nonnegative coefficients, we say that $\{f_n(q)\}_{n\geq 0}$ is *q-log-concave.*

In section 2, we give the generating functions of hyperpell and hyperpell-lucas sequences. In section 3, we discuss their log-concavity and log-convexity. We investigate also the q-log-concavity of some polynomials related to hyperpell and hyperpell-lucas numbers.

2. The generating functions

The generating function of Pell numbers and Pell-Lucas numbers denoted $G_P(t)$ and $G_Q(t)$, respectively, are

$$G_P(t) := \sum_{n=0}^{+\infty} P_n t^n = \frac{t}{1-2t-t^2}, \tag{2.1}$$

and

$$G_Q(t) := \sum_{n=0}^{+\infty} Q_n t^n = \frac{2-2t}{1-2t-t^2}. \tag{2.2}$$

So, we establish the generating function of hyperpell and hyperpell-lucas numbers using respectively

$$P_n^{[r]} = P_{n-1}^{[r]} + P_n^{[r-1]} \quad \text{and} \quad Q_n^{[r]} = Q_{n-1}^{[r]} + Q_n^{[r-1]}. \tag{2.3}$$

The generating functions of hyperpell numbers and hyperlucas numbers are

$$G_P^{[r]}(t) = \sum_{n=0}^{\infty} P_n^{[r]} t^n = \frac{t}{(1-2t-t^2)(1-t)^r}, \tag{2.4}$$

and

$$G_Q^{[r]}(t) = \sum_{n=0}^{\infty} Q_n^{[r]} t^n = \frac{2-2t}{(1-2t-t^2)(1-t)^r}. \tag{2.5}$$

3. The log-concavity and log-convexity properties

We start the section by some useful lemmas.

Lemma 3.1. [12] *If the sequences $\{x_n\}$ and $\{y_n\}$ are log-concave, then so is their ordinary convolution $z_n = \sum_{k=0}^n x_k y_{n-k}$, $n = 0, 1, \ldots.$*

Lemma 3.2. [12] *If the sequence $\{x_n\}$ is log-concave, then so is the binomial convolution $z_n = \sum_{k=0}^n \binom{n}{k} x_k$, $n = 0, 1, \ldots.$*

Lemma 3.3. [8] *If the sequence $\{x_n\}$ is log-convex, then so is the binomial convolution $z_n = \sum_{k=0}^n \binom{n}{k} x_k$, $n = 0, 1, \ldots.$*

The following result deals with the log-concavity of hyperpell numbers and hyperlucas sequences.

Theorem 3.4. *The sequences $\left\{P_n^{[r]}\right\}_{n\geq 0}$ and $\left\{Q_n^{[r]}\right\}_{n\geq 0}$ are log-concave for $r \geq 1$ and $r \geq 2$ respectively.*

Proof. We have

$$P_n^{[1]} = \frac{1}{4}\left(Q_{n+1} - 2\right) \quad \text{and} \quad Q_n^{[1]} = 2P_{n+1}. \tag{3.1}$$

When $n = 1$, $\left(P_n^{[1]}\right)^2 - P_{n-1}^{[1]}P_{n+1}^{[1]} = 1 > 0$. When $n \geq 2$, it follows from (3.1) and (1.8) that

$$\begin{aligned}\left(P_n^{[1]}\right)^2 - P_{n-1}^{[1]}P_{n+1}^{[1]} &= \frac{1}{16}\left[\left(Q_{n+1} - 2\right)^2 - \left(Q_n - 2\right)\left(Q_{n+2} - 2\right)\right] \\ &= \frac{1}{16}\left(Q_{n+1}^2 - Q_nQ_{n+2} - 4Q_{n+1} + 2Q_n + 2Q_{n+2}\right) \\ &= \frac{1}{4}\left(2(-1)^{n-1} + Q_{n+1}\right) \geq 0.\end{aligned}$$

Then $\left\{P_n^{[1]}\right\}_{n\geq 0}$ is log-concave. By Lemma 3.1, we know that $\left\{P_n^{[r]}\right\}_{n\geq 0}$ $(r \geq 1)$ is log-concave.

It follows from (3.1) and (1.7) that

$$\left(Q_n^{[1]}\right)^2 - Q_{n-1}^{[1]}Q_{n+1}^{[1]} = 4\left(P_{n+1}^2 - P_nP_{n+2}\right) = 4\left(-1\right)^n = \pm 4 \tag{3.2}$$

Hence $\left\{Q_n^{[1]}\right\}_{n\geq 0}$ is not log-concave.

One can verify that

$$Q_n^{[2]} = \frac{1}{2}\left(Q_{n+2} - 2\right) = 2P_{n+1}^{[1]}. \tag{3.3}$$

Then $\left\{Q_n^{[2]}\right\}_{n\geq 0}$ is log-concave. By Lemma 3.1, we know that $\left\{Q_n^{[r]}\right\}_{n\geq 0}$ $(r \geq 2)$ is log-concave. This completes the proof of Theorem 3.4. □

Then we have the following corollary.

Corollary 3.5. *The sequences* $\left\{\sum_{k=0}^{n}\binom{n}{k}P_k^{[r]}\right\}_{n\geq 0}$ *and* $\left\{\sum_{k=0}^{n}\binom{n}{k}Q_k^{[r]}\right\}_{n\geq 0}$ *are log-concave for* $r \geq 1$ *and* $r \geq 2$ *respectively.*

Proof. Use Lemma 3.2.

Now we establish the *log-concavity of order two* of the sequences $\left\{P_n^{[1]}\right\}_{n\geq 0}$ and $\left\{Q_n^{[2]}\right\}_{n\geq 0}$ for some special sub-sequences. □

Theorem 3.6. *Let be for* $n \geq 1$

$$T_n := \left(P_n^{[1]}\right)^2 - P_{n-1}^{[1]}P_{n+1}^{[1]} \quad \textit{and} \quad R_n := \left(Q_n^{[2]}\right)^2 - Q_{n-1}^{[2]}Q_{n+1}^{[2]}.$$

Then $\{T_{2n}\}_{n\geq 1}$, $\{R_{2n+1}\}_{n\geq 0}$ *are log-concave, and* $\{T_{2n+1}\}_{n\geq 0}$, $\{R_{2n}\}_{n\geq 1}$ *are log-convex.*

Proof. Using respectively (3.3) and (1.8) , we get

$$\left(Q_n^{[2]}\right)^2 - Q_{n-1}^{[2]}Q_{n+1}^{[2]} = 2(-1)^n + Q_{n+1},$$

and thus, for $n \geq 1$,

$$T_n = \frac{1}{4}\left(2\left(-1\right)^{n-1} + Q_n\right) \quad \text{and} \quad R_n = 2(-1)^n + Q_{n+1}. \tag{3.4}$$

By applying (3.4) and (1.8), for $n \geq 1$ we get

$$Q_{2n}^2 - Q_{2n-2}Q_{2n+2} = -32 \ \text{ and } \ Q_{2n+1}^2 - Q_{2n-1}Q_{2n+3} = 32. \tag{3.5}$$

Then

$$\begin{aligned} T_{2n}^2 - T_{2(n-1)}T_{2(n+1)} &= \frac{1}{16}\left(Q_{2n}^2 - Q_{2n-2}Q_{2n+2} - 4Q_{2n} + 2Q_{2n-2} + 2Q_{2n+2}\right) \\ &= 4(Q_{2n} - 4) > 0. \end{aligned}$$

and

$$\begin{aligned} R_{2n+1}^2 - R_{2n-1}R_{2n+3} &= \left(Q_{2n+2}^2 - Q_{2n}Q_{2n+2} - 4Q_{2n+2} + 2Q_{2n} + 2Q_{2n+4}\right) \\ &= 64(Q_{2n+2} - 4) > 0. \end{aligned}$$

Then $\{T_{2n}\}_{n\geq 1}$ and $\{R_{2n+1}\}_{n\geq 0}$ are log-concave.

Similarly by applying (3.4) and (3.5), we have

$$T_{2n+1}^2 - T_{2n-1}T_{2n+3} = -\frac{1}{2}Q_{2n+1} < 0,$$

and

$$R_{2n}^2 - R_{2(n-1)}R_{2(n+1)} = -8Q_{2n+1} < 0.$$

Then $\{T_{2n+1}\}_{n\geq 0}$ and $\{R_{2n}\}_{n\geq 1}$ are log-convex. This completes the proof. □

Corollary 3.7. *The sequences* $\left\{\sum_{k=0}^{n}\binom{n}{k}T_{2k}\right\}_{n\geq 0}$ *and* $\left\{\sum_{k=0}^{n}\binom{n}{k}R_{2k+1}\right\}_{n\geq 0}$ *are log-concave.*

Proof. Use Lemma 3.2. □

Corollary 3.8. *The sequences* $\left\{\sum_{k=0}^{n}\binom{n}{k}T_{2k+1}\right\}_{n\geq 1}$ *and* $\left\{\sum_{k=0}^{n}\binom{n}{k}R_{2k}\right\}_{n\geq 1}$ *are log-convex.*

Proof. Use Lemma 3.3. □

Lemma 3.9. *Let* $a_n := \sum_{k=0}^{n}\binom{n}{k}P_{k+1}$, *where* $\{P_n\}_{n\geq 0}$ *is the Pell sequence. Then* $\{a_n\}_{n\geq 0}$ *satisfy the following recurrence relations*

$$a_n = 3a_{n-1} + \sum_{k=0}^{n-2} a_k \quad \text{and} \quad a_n = 4a_{n-1} - 2a_{n-2}.$$

Proof. Let be $b_n := \sum_{k=0}^{n} \binom{n}{k} P_k$, where $\{P_n\}_{n\geq -1}$ is the Pell sequence extended to $P_{-1} = 1$.

Using Pascal formula and the recurrence relation of Pell sequence together into the development $\sum_{k=0}^{n} \binom{n}{k} P_{k+1}$ we get $a_n = 3a_{n-1} + b_{n-1}$, then by $b_n = b_{n-1} + a_{n-1}$. By iterated use of this relation with the precedent one, we get $a_n = 3a_{n-1} + \sum_{k=0}^{n-2} a_k$ (with $b_0 = 0$ and $a_0 = 1$), thus $a_n = 4a_{n-1} - 2a_{n-2}$. □

Theorem 3.10. *The sequences* $\left\{nQ_n^{[1]}\right\}_{n\geq 0}$ *and* $\left\{\sum_{k=0}^{n} \binom{n}{k} Q_k^{[1]}\right\}_{n\geq 0}$ *are log-concave and log-convex, respectively.*

Proof. Let be

$$S_n := n^2 \left(Q_n^{[1]}\right)^2 - (n^2-1) Q_{n-1}^{[1]} Q_{n+1}^{[1]} \quad \text{and} \quad K_n := \sum_{k=0}^{n} \binom{n}{k} Q_k^{[1]},$$

with the convention that $K_{<0} = 0$.

From (3.2), we have

$$\begin{aligned} S_n &= 4(n^2-1)(-1)^n + \left(Q_n^{[1]}\right)^2 \\ &= 4\left[(n^2-1)(-1)^n + P_{n+1}^2\right] \geq 4\left[(n^2-1)(-1)^n + (n+1)^2\right] > 0. \end{aligned}$$

Then $\left\{nQ_n^{[1]}\right\}_{n\geq 0}$ is log-concave.

Using Lemma 3.9, we can verify that

$$K_n = 4K_{n-1} - 2K_{n-2}. \tag{3.6}$$

The associated Binet-formula is

$$K_n = \frac{\left(1+\sqrt{2}\right)\alpha^n - \left(1-\sqrt{2}\right)\beta^n}{\alpha - \beta}, \quad \text{with} \quad \alpha, \beta = 2 \pm \sqrt{2},$$

which provides

$$K_n^2 - K_{n-1}K_{n+1} = -2^{n+1} < 0.$$

Then $\left\{\sum_{k=0}^{n} \binom{n}{k} Q_k^{[1]}\right\}_{n\geq 0}$ is log-convex. □

Remark 3.11. The terms of the sequence $\{K_n\}_n$ satisfy $K_n = 2^{(n+2)/2} P_{n+1}$ if n is even, and $K_n = 2^{(n-1)/2} Q_{n+1}$ if n is odd.

Theorem 3.12. *The sequences* $\left\{n! P_n^{[1]}\right\}_{n\geq 0}$ *and* $\left\{n! Q_n^{[2]}\right\}_{n\geq 0}$ *are log-balanced.*

Proof. By Theorem 3.4, in order to prove the log-balanced property of $\left\{n! P_n^{[1]}\right\}_{n\geq 0}$ and $\left\{n! Q_n^{[2]}\right\}_{n\geq 0}$ we only need to show that they are log-convex. It follows from the proof of Theorem 3.4 that

$$\left(P_n^{[1]}\right)^2 - P_{n-1}^{[1]} P_{n+1}^{[1]} = \frac{1}{4}\left(2(-1)^{n-1} + Q_{n+1}\right), \tag{3.7}$$

and from the proof of Theorem 3.6 that

$$\left(Q_n^{[2]}\right)^2 - Q_{n-1}^{[2]} Q_{n+1}^{[2]} = 2\,(-1)^n + Q_{n+1}. \tag{3.8}$$

Let

$$M_n := n\left(P_n^{[1]}\right)^2 - (n+1) P_{n-1}^{[1]} P_{n+1}^{[1]},$$
$$B_n := n\left(Q_n^{[2]}\right)^2 - (n+1) Q_{n-1}^{[2]} Q_{n+1}^{[2]},$$

from (3.3), (3.7) and (3.8), we get

$$M_n = \frac{(n+1)}{4}\left(2\,(-1)^{n-1} + Q_{n+1}\right) - \frac{1}{4}(Q_{n+1} - 2)^2,$$
$$B_n = (n+1)\left(2\,(-1)^n + Q_{n+1}\right) - \frac{1}{4}(Q_{n+2} - 2)^2.$$

Clearly $B_n \leq 0$ for $n = 0, 1, 2$. We have by induction that for $n \geq 1$, $Q_n \geq n+1$. This gives

$$B_n \leq (Q_{n+1} - 1)\left(2\,(-1)^n + Q_{n+1}\right) - \frac{1}{4}(2Q_{n+1} + Q_n - 2)^2 < 0.$$

Also, $M_n \leq 0$ for $n = 2$ and for $n \geq 3$, $Q_n \geq n+6$. This gives $n+1 \leq Q_{n+1} - 6$, and

$$M_n \leq \frac{1}{4}\left[(Q_{n+1} - 6)\left(2\,(-1)^{n-1} + Q_{n+1}\right) - (Q_{n+1} - 2)^2\right]$$
$$= \frac{1}{4}\left[\left(-2 + 2\,(-1)^{n-1}\right) Q_{n+1} - 4 - 12\,(-1)^{n-1}\right] < 0.$$

Hence $\{n! P_n^{[1]}\}_{n \geq 0}$ and $\{n! Q_n^{[2]}\}_{n \geq 0}$ are log-convex. As the sequences $\{P_n^{[1]}\}_{n \geq 0}$ and $\{Q_n^{[2]}\}_{n \geq 0}$ are log-concave, so the sequences $\{n! P_n^{[1]}\}_{n \geq 0}$ and $\{n! Q_n^{[2]}\}_{n \geq 0}$ are log-balanced. □

Theorem 3.13. *Define, for $r \geq 1$, the polynomials*

$$P_{n,r}(q) := \sum_{k=0}^{n} P_k^{[r]} q^k \quad \text{and} \quad Q_{n,r}(q) := \sum_{k=0}^{n} Q_k^{[r]} q^k.$$

The polynomials $P_{n,r}(q)$ $(r \geq 1)$ and $Q_{n,r}(q)$ $(r \geq 2)$ are q-log-concave.

Proof. When $n \geq 1$, $r \geq 1$,

$$P_{n,r}^2(q) - P_{n-1,r}(q) P_{n+1,r}(q)$$
$$= \left(\sum_{k=0}^{n} P_k^{[r]} q^k\right)^2 - \left(\sum_{k=0}^{n-1} P_k^{[r]} q^k\right)\left(\sum_{k=0}^{n+1} P_k^{[r]} q^k\right)$$

$$= \left(\sum_{k=0}^{n} P_k^{[r]} q^k\right)^2 - \left(\sum_{k=0}^{n} P_k^{[r]} q^k - P_n^{[r]} q^n\right)\left(\sum_{k=0}^{n} P_k^{[r]} q^k + P_{n+1}^{[r]} q^{n+1}\right)$$

$$= \left(P_n^{[r]} q^n - P_{n+1}^{[r]} q^{n+1}\right)\sum_{k=0}^{n} P_k^{[r]} q^k + P_n^{[r]} P_{n+1}^{[r]} q^{2n+1}$$

$$= \sum_{k=1}^{n} \left(P_k^{[r]} P_n^{[r]} - P_{k-1}^{[r]} P_{n+1}^{[r]}\right) q^{k+n}.$$

When $n \geq 1$, $r \geq 2$, through computation, we get

$$Q_{n,r}^2(q) - Q_{n-1,r}(q) Q_{n+1,r}(q) = \sum_{k=1}^{n} \left(Q_k^{[r]} Q_n^{[r]} - Q_{k-1}^{[r]} Q_{n+1}^{[r]}\right) q^{k+n} + Q_n^{[r]} q^n.$$

As $\left\{P_n^{[r]}\right\}$ and $\left\{Q_n^{[r]}\right\}$ $(r \geq 2)$ are log-concave, then the polynomials $P_{n,r}(q)$ $(r \geq 1)$ and $Q_{n,r}(q)$ $(r \geq 2)$ are q-log-concave. □

Acknowledgements. We would like to thank the referee for useful suggestions and several comments witch involve the quality of the paper.

References

[1] Belbachir, H., Belkhir, A., Combinatorial Expressions Involving Fibonacci, Hyperfibonacci, and Incomplete Fibonacci Numbers, *J. Integer Seq.,* Vol. 17 (2014), Article 14.4.3.

[2] Belbachir, H., Bencherif, F., Unimodality of sequences associated to Pell numbers, *Ars Combin.,* 102 (2011), 305–311.

[3] Belbachir, H., Bencherif, F., Szalay, L., Unimodality of certain sequences connected with binomial coefficients, *J. Integer Seq.,* 10 (2007), Article 07. 2. 3.

[4] Belbachir, H., Szalay, L., Unimodal rays in the regular and generalized Pascal triangles, *J. Integer Seq.,* 11 (2008), Article. 08.2.4.

[5] Brenti, F., Unimodal, log-concave and Pólya frequency sequences in combinatorics, *Mem. Amer. Math. Soc.,* no. 413 (1989).

[6] Cao, N. N., Zhao, F. Z, Some Properties of Hyperfibonacci and Hyperlucas Numbers, *J. Integer Seq.,* 13 (8) (2010), Article 10.8.8.

[7] Chen, W. Y. C., Wang, L. X. W., Yang, A. L. B., Schur positivity and the q-log-convexity of the Narayana polynomials, *J. Algebr. Comb.,* 32 (2010), 303–338.

[8] Davenport, H., Pólya, G., On the product of two power series, *Canadian J. Math.,* 1 (1949), 1–5.

[9] Dil, A., Mező, I., A symmetric algorithm for hyperharmonic and Fibonacci numbers, *Appl. Math. Comput.,* 206 (2008), 942–951.

[10] Liu, L., Wang, Y., On the log-convexity of combinatorial sequences, *Advances in Applied Mathematics,* vol. 39, Issue 4, (2007), 453–476.

[11] Sloane, N. J. A., *On-line Encyclopedia of Integer Sequences,* http://oeis.org, (2014).

[12] Wang, Y., Yeh, Y. N., Log-concavity and LC-positivity, *Combin. Theory Ser. A,* 114 (2007), 195–210.

[13] Zheng, L. N., Liu, R., On the Log-Concavity of the Hyperfibonacci Numbers and the Hyperlucas Numbers, *J. Integer Seq.,* Vol. 17 (2014), Article 14.1.4.

Annales Mathematicae et Informaticae
43 (2014) pp. 13–17
http://ami.ektf.hu

On geodesic mappings of Riemannian spaces with cyclic Ricci tensor

Sándor Bácsó[a], Robert Tornai[a], Zoltán Horváth[b]

[a]University of Debrecen, Faculty of Informatics
bacsos@unideb.hu, tornai.robert@inf.unideb.hu

[b]Ferenc Rákóczi II. Transcarpathian Hungarian Institute
zolee27@kmf.uz.ua

Submitted May 4, 2012 — Accepted March 25, 2014

Abstract

An n-dimensional Riemannian space V^n is called a Riemannian space with cyclic Ricci tensor [2, 3], if the Ricci tensor R_{ij} satisfies the following condition

$$R_{ij,k} + R_{jk,i} + R_{ki,j} = 0,$$

where R_{ij} the Ricci tensor of V^n, and the symbol "," denotes the covariant derivation with respect to Levi-Civita connection of V^n.

In this paper we would like to treat some results in the subject of geodesic mappings of Riemannian space with cyclic Ricci tensor.

Let $V^n = (M^n, g_{ij})$ and $\overline{V}^n = (M^n, \overline{g}_{ij})$ be two Riemannian spaces on the underlying manifold M^n. A mapping $V^n \to \overline{V}^n$ is called geodesic, if it maps an arbitrary geodesic curve of V^n to a geodesic curve of $\overline{V}^n$.[4]

At first we investigate the geodesic mappings of a Riemannian space with cyclic Ricci tensor into another Riemannian space with cyclic Ricci tensor.

Finally we show that, the Riemannian - Einstein space with cyclic Ricci tensor admit only trivial geodesic mapping.

Keywords: Riemannian spaces, geodesic mapping.

MSC: 53B40.

1. Introduction

Let an n-dimensional V^n Riemannian space be given with the fundamental tensor $g_{ij}(x)$. V^n has the Riemannian curvature tensor R^i_{jkl} in the following form:

$$R^h_{ijk}(x) = \partial_j \Gamma^h_{ik}(x) + \Gamma^\alpha_{ik}(x)\Gamma^h_{j\alpha}(x) - \partial_k \Gamma^h_{ij}(x) - \Gamma^\alpha_{ij}(x)\Gamma^h_{k\alpha}(x), \tag{1.1}$$

where $\Gamma^i_{jk}(x)$ are the coefficients of Levi-Civita connection of V^n.

The Ricci curvature tensor we obtain from the Riemannian curvature tensor using of the following transvection: $R^\alpha_{jk\alpha}(x) = R_{jk}(x)$[1].

Definition 1.1. [2, 3] A Riemannian space V^n is called a Riemannian space with cyclic Ricci tensor, if the Ricci tensor of V^n satisfies the following equation:

$$R_{ij,k} + R_{jk,i} + R_{ki,j} = 0, \tag{1.2}$$

where the symbol "," means the covariant derivation with respect to Levi-Civita connection of V^n.

Definition 1.2. [4] Let two Riemannian spaces V^n and $\overline{V}^n$ be given on the underlying manifold M_n . The maps: $\gamma : V^n \to \overline{V}^n$ is called geodesic (projective) mappings, if any geodesic curve of V^n coincides with a geodesic curve of $\overline{V}^n$.

It is wellknown, that the the geodesic curve $x^i(t)$ of V^n is a result of the second order ordinary differential equations in a canonical parameter t:

$$\frac{d^2x^i}{dt^2} + \Gamma^i_{\alpha\beta}(x)\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt} = 0. \tag{1.3}$$

We need in the investigations the next:

Theorem 1.3. *[4] The maps: $V^n \to \overline{V}^n$ is geodesic if and only if exits a gradient vector field $\psi_i(x)$, which satisfies the following condition:*

$$\overline{\Gamma}^i_{jk}(x) = \Gamma^i_{jk}(x) + \delta^i_j\psi_k(x) + \delta^i_k\psi_j(x), \tag{1.4}$$

and

Definition 1.4. [1] A Riemannian space V^n is called Einstein space, if exists a $\rho(x)$ scalar function, which satisfies the equation:

$$R_{ij} = \rho(x)g_{ij}(x). \tag{1.5}$$

[1]The Roman and Greek indices run over the range 1, …, n; the Roman indices are free but the Greek indices denote summation.

2. Geodesic mappings of Riemannian spaces with cyclic Ricci tensors

It is easy to get the next equations [4]:

$$\overline{R}_{ij} = R_{ij} + (n-1)\psi_{ij}, \tag{2.1}$$

where $\psi_{ij} = \psi_{i,j} - \psi_i\psi_j$ and

$$\overline{R}_{ij,k} = \frac{\partial \overline{R}_{ij}}{\partial x^k} - \overline{\Gamma}^{\alpha}_{ik}(x)\overline{R}_{\alpha j} - \overline{\Gamma}^{\alpha}_{jk}(x)\overline{R}_{\alpha i}, \tag{2.2}$$

where $\overline{\Gamma}^{\alpha}_{ik}(x)$ are components of Levi-Civita connection if $\overline{V}^n$.

At now we suppose, that $\overline{V}^n$ in a Riemannian space with cyclic Ricci tensor, that is

$$\overline{R}_{ij,k} + \overline{R}_{jk,i} + \overline{R}_{ki,j} = 0. \tag{2.3}$$

Using (2.2) we can rewrite (2.3) in the following form:

$$\begin{aligned}
\frac{\partial \overline{R}_{ij}}{\partial x^k} - \overline{\Gamma}^{\alpha}_{ik}(x)\overline{R}_{\alpha j} - \overline{\Gamma}^{\alpha}_{jk}(x)\overline{R}_{\alpha i} + \\
\frac{\partial \overline{R}_{jk}}{\partial x^i} - \overline{\Gamma}^{\alpha}_{ji}(x)\overline{R}_{\alpha k} - \overline{\Gamma}^{\alpha}_{ki}(x)\overline{R}_{\alpha j} + \\
\frac{\partial \overline{R}_{ki}}{\partial x^j} - \overline{\Gamma}^{\alpha}_{kj}(x)\overline{R}_{\alpha i} - \overline{\Gamma}^{\alpha}_{ij}(x)\overline{R}_{\alpha k} = 0.
\end{aligned} \tag{2.4}$$

From (1.4) and (2.1) we can compute:

$$\begin{gathered}
\frac{\partial (R_{ij} + (n-1)\psi_{ij})}{\partial x^k} - (\Gamma^{\alpha}_{ik}(x) + \psi_i(x)\delta^{\alpha}_k + \psi_k(x)\delta^{\alpha}_i)(R_{\alpha j} + (n-1)\psi_{\alpha j}) - \\
-(\Gamma^{\alpha}_{jk}(x) + \psi_j(x)\delta^{\alpha}_k + \psi_k(x)\delta^{\alpha}_j)(R_{\alpha i} + (n-1)\psi_{\alpha i}) + \\
\frac{\partial (R_{jk} + (n-1)\psi_{jk})}{\partial x^i} - (\Gamma^{\alpha}_{ji}(x) + \psi_j(x)\delta^{\alpha}_i + \psi_i(x)\delta^{\alpha}_j)(R_{\alpha k} + (n-1)\psi_{\alpha k}) - \\
-(\Gamma^{\alpha}_{ki}(x) + \psi_k(x)\delta^{\alpha}_i + \psi_i(x)\delta^{\alpha}_k)(R_{\alpha j} + (n-1)\psi_{\alpha j}) + \\
\frac{\partial (R_{ki} + (n-1)\psi_{ki})}{\partial x^j} - (\Gamma^{\alpha}_{kj}(x) + \psi_k(x)\delta^{\alpha}_j + \psi_j(x)\delta^{\alpha}_k)(R_{\alpha i} + (n-1)\psi_{\alpha i}) - \\
-(\Gamma^{\alpha}_{ij}(x) + \psi_i(x)\delta^{\alpha}_j + \psi_j(x)\delta^{\alpha}_i)(R_{\alpha k} + (n-1)\psi_{\alpha k}) = 0.
\end{gathered}$$

That is

$$\left.\begin{aligned}
\tfrac{\partial R_{ij}}{\partial x^k} - \Gamma^{\alpha}_{ik}(x)R_{\alpha j} - \Gamma^{\alpha}_{jk}(x)R_{\alpha i} + \\
+\tfrac{\partial R_{jk}}{\partial x^i} - \Gamma^{\alpha}_{ji}(x)R_{\alpha k} - \Gamma^{\alpha}_{ki}(x)R_{\alpha j} + \\
+\tfrac{\partial R_{ki}}{\partial x^j} - \Gamma^{\alpha}_{kj}(x)R_{\alpha i} - \Gamma^{\alpha}_{ij}(x)R_{\alpha k} +
\end{aligned}\right\} R_{ij,k} + R_{jk,i} + R_{ki,j}$$

$$+(n-1)\frac{\partial\psi_{ij}}{\partial x^k} - (n-1)\Gamma^{\alpha}_{ik}(x)\psi_{\alpha j} - \psi_i(x)R_{kj} - (n-1)\psi_i(x)\psi_{kj} -$$
$$-\psi_k(x)R_{ij} - (n-1)\psi_k(x)\psi_{ij} - (n-1)\Gamma^{\alpha}_{jk}(x)\psi_{\alpha i} - \psi_j(x)R_{ki} -$$
$$-(n-1)\psi_j(x)\psi_{ki} - \psi_k(x)R_{ji} - (n-1)\psi_k(x)\psi_{ji} +$$
$$+(n-1)\frac{\partial\psi_{jk}}{\partial x^i} - (n-1)\Gamma^{\alpha}_{ji}(x)\psi_{\alpha k} - \psi_j(x)R_{ik} - (n-1)\psi_j(x)\psi_{ik} -$$
$$-\psi_i(x)R_{jk} - (n-1)\psi_i(x)\psi_{jk} - (n-1)\Gamma^{\alpha}_{ki}(x)\psi_{\alpha j} - \psi_k(x)R_{ij} -$$
$$-(n-1)\psi_k(x)\psi_{ij} - \psi_i(x)R_{kj} - (n-1)\psi_i(x)\psi_{kj} +$$
$$+(n-1)\frac{\partial\psi_{ki}}{\partial x^j} - (n-1)\Gamma^{\alpha}_{kj}(x)\psi_{\alpha i} - \psi_k(x)R_{ji} - (n-1)\psi_k(x)\psi_{ji} -$$
$$-\psi_j(x)R_{ki} - (n-1)\psi_j(x)\psi_{ki} - (n-1)\Gamma^{\alpha}_{ij}(x)\psi_{\alpha k} - \psi_i(x)R_{jk} -$$
$$-(n-1)\psi_i(x)\psi_{jk} - \psi_j(x)R_{ik} - (n-1)\psi_j(x)\psi_{ik} = 0.$$

If we suppose, that V^n has cyclic Ricci tensor we have:

$$(n-1)\left(\frac{\partial\psi_{ij}}{\partial x^k} - \Gamma^{\alpha}_{ik}(x)\psi_{\alpha j} - \Gamma^{\alpha}_{jk}(x)\psi_{\alpha i}\right) +$$
$$+(n-1)\left(\frac{\partial\psi_{jk}}{\partial x^i} - \Gamma^{\alpha}_{ji}(x)\psi_{\alpha k} - \Gamma^{\alpha}_{ki}(x)\psi_{\alpha j}\right) +$$
$$+(n-1)\left(\frac{\partial\psi_{ki}}{\partial x^j} - \Gamma^{\alpha}_{kj}(x)\psi_{\alpha i} - \Gamma^{\alpha}_{ij}(x)\psi_{\alpha k}\right) +$$
$$-4\psi_i(x)R_{jk} - 4\psi_j(x)R_{ki} - 4\psi_k(x)R_{ij} -$$
$$-(n-1)(4\psi_i(x)\psi_{jk} + 4\psi_j(x)\psi_{ki} + 4\psi_k(x)\psi_{ij}) = 0.$$

That is

$$\begin{aligned} (n-1)(\psi_{ij,k} + \psi_{jk,i} + \psi_{ki,j}) - \\ -4(\psi_i(x)R_{jk} + \psi_j(x)R_{ki} + \psi_k(x)R_{ij}) - \\ -4(n-1)(\psi_i(x)\psi_{jk} + \psi_j(x)\psi_{ki} + \psi_k(x)\psi_{ij}) = 0. \end{aligned} \tag{2.5}$$

Theorem 2.1. *V^n and $\overline{V}^n$ Riemannian spaces with cyclic Ricci tensors have common geodesics, that is V^n and $\overline{V}^n$ have a geodesic mapping if and only if exists a $\psi_i(x)$ gradient vector, which satisfies the condition:*

$$(n-1)(\psi_{ij,k} + \psi_{jk,i} + \psi_{ki,j}) -$$
$$-4(\psi_i(x)R_{jk} + \psi_j(x)R_{ki} + \psi_k(x)R_{ij}) -$$
$$-4(n-1)(\psi_i(x)\psi_{jk} + \psi_j(x)\psi_{ki} + \psi_k(x)\psi_{ij}) = 0.$$

3. Consequences

A) If $\psi_{ij} = 0$, then $\overline{R}_{ij} = R_{ij}$, and $\psi_{i,j} = \psi_i \psi_j$, so we obtain:

$$\psi_i(x)R_{jk} + \psi_j(x)R_{ki} + \psi_k(x)R_{ij} = 0. \tag{3.1}$$

B) If the V^n is a Riemannian space with cyclic Ricci tensor and at the same time is a Einstein space, then we get

$$\rho\psi_i(x)g_{jk} + \rho\psi_j(x)g_{ki} + \rho\psi_k(x)g_{ij} = 0$$

that is

$$n\psi_i(x) + 2\psi_i(x) = 0, \tag{3.2}$$

so

$$(n+2)\psi_i(x) = 0. \tag{3.3}$$

It means

Theorem 3.1. *A Riemannian-Einstein space V^n with cyclic Ricci tensor admits into $\overline{V}^n$ with cyclic Ricci tensor only trivial (affin) geodesic mapping.*

References

[1] A. L. Besse, Einstein manifolds, *Springer-Verlag*, (1987)

[2] T. Q. Binh, On weakly symmetric Riemannian spaces, *Publ. Math. Debrecen*, 42/1-2 (1993), 103–107.

[3] M. C. Chaki - U. C. De, On pseudo symmetric spaces, *Acta Math. Hung.*, 54 (1989), 185–190.

[4] N. Sz. Szinjukov, Geodezicseszkije otrobrazsenyija Rimanovih prosztransztv, *Moscow, Nauka*, (1979)

Annales Mathematicae et Informaticae
43 (2014) pp. 19–27
http://ami.ektf.hu

A symmetric algorithm for hyper-Fibonacci and hyper-Lucas numbers

Mustafa Bahşi[a], **István Mező**[b*], **Süleyman Solak**[c]

[a]Aksaray University, Education Faculty, Aksaray, Turkey
mhvbahsi@yahoo.com

[b]Nanjing University of Information Science and Technology, Nanjing, P. R. China
istvanmezo81@gmail.com

[c]N. E. University, A.K Education Faculty, 42090, Meram, Konya, Turkey
ssolak42@yahoo.com

Submitted July 22, 2014 — Accepted November 14, 2014

Abstract

In this work we study some combinatorial properties of hyper-Fibonacci, hyper-Lucas numbers and their generalizations by using a symmetric algorithm obtained by the recurrence relation $a_n^k = ua_n^{k-1} + va_{n-1}^k$. We point out that this algorithm can be applied to hyper-Fibonacci, hyper-Lucas and hyper-Horadam numbers.

Keywords: Hyper-Fibonacci numbers; hyper-Lucas numbers

MSC: 11B37; 11B39; 11B65

1. Introduction

The sequence of Fibonacci numbers is one of the most well known sequence, and it has many applications in mathematics, statistics, and physics. The Fibonacci numbers are defined by the second order linear recurrence relation: $F_{n+1} = F_n + F_{n-1}$ $(n \geq 1)$ with the initial conditions $F_0 = 0$ and $F_1 = 1$. Similarly, the Lucas

*The research of István Mező was supported by the Scientific Research Foundation of Nanjing University of Information Science & Technology, and The Startup Foundation for Introducing Talent of NUIST. Project no.: S8113062001

numbers are defined by $L_{n+1} = L_n + L_{n-1}$ $(n \geq 1)$ with the initial conditions $L_0 = 2$ and $L_1 = 1$. There are some elementary identities for F_n and L_n. Two of them are $F_s + L_s = 2F_{s+1}$ and $F_s - L_s = 2F_{s-1}$. These will be generalized in section 2 (see Theorem 2.5).

The Fibonacci sequence can be generalized to the second order linear recurrence $W_n(a, b; p, q)$, or briefly W_n, defined by

$$W_{n+1} = pW_n + qW_{n-1},$$

where $n \geq 1$, $W_0 = a$ and $W_1 = b$. This sequence was introduced by Horadam [7]. Some of the special cases are:

i) The Fibonacci number $F_n = W_n(0, 1; 1, 1)$,

ii) The Lucas number $L_n = W_n(2, 1; 1, 1)$,

iii) The Pell number $P_n = W_n(0, 1; 2, 1)$.

In [4], Dil and Mező introduced the "hyper-Fibonacci" numbers $F_n^{(r)}$ and "hyper-Lucas" numbers $L_n^{(r)}$. These are defined as

$$F_n^{(r)} = \sum_{k=0}^{n} F_k^{(r-1)} \quad \text{with} \quad F_n^{(0)} = F_n,\ F_0^{(r)} = 0,\ F_1^{(r)} = 1,$$

$$L_n^{(r)} = \sum_{k=0}^{n} L_k^{(r-1)} \quad \text{with} \quad L_n^{(0)} = L_n,\ L_0^{(r)} = 2,\ L_1^{(r)} = 2r + 1,$$

where r is a positive integer, moreover F_n and L_n are the ordinary Fibonacci and Lucas numbers, respectively. The generating functions of hyper-Fibonacci and hyper-Lucas numbers are [4]:

$$\sum_{n=0}^{\infty} F_n^{(r)} t^n = \frac{t}{(1 - t - t^2)(1 - t)^r}, \quad \sum_{n=0}^{\infty} L_n^{(r)} t^n = \frac{2 - t}{(1 - t - t^2)(1 - t)^r}.$$

Also, the hyper-Fibonacci and hyper-Lucas numbers have the recurrence relations $F_n^{(r)} = F_{n-1}^{(r)} + F_n^{(r-1)}$ and $L_n^{(r)} = L_{n-1}^{(r)} + L_n^{(r-1)}$, respectively. The first few values of $F_n^{(r)}$ and $L_n^{(r)}$ are as follows [2]:

$$F_n^{(1)}: 0, 1, 2, 4, 7, 12, 20, 33, 54, \ldots, \qquad F_n^{(2)}: 0, 1, 3, 7, 14, 26, 46, 79, \ldots$$

$$L_n^{(1)}: 2, 3, 6, 10, 17, 28, 46, 75, \ldots, \qquad L_n^{(2)}: 2, 5, 11, 21, 38, 66, 112, \ldots.$$

Now we introduce the hyper-Horadam numbers $W_n^{(r)}$ defined by

$$W_n^{(r)} = W_{n-1}^{(r)} + W_n^{(r-1)} \quad \text{with} \quad W_n^{(0)} = W_n,\ W_0^{(n)} = W_0 = a$$

where W_n is the nth Horadam number. Some of the special cases of hyper-Horadam number $W_n^{(r)}$ are as follows:

i) If $W_n^{(0)} = F_n = W_n(0,1;1,1)$ and $W_0^{(n)} = W_0 = F_0 = 0$, then $W_n^{(r)}$ is the hyper-Fibonacci number, that is, $W_n^{(r)} = F_n^{(r)}$.

ii) If $W_n^{(0)} = L_n = W_n(2,1;1,1)$ and $W_0^{(n)} = W_0 = L_0 = 2$, then $W_n^{(r)}$ is the hyper-Lucas number, that is, $W_n^{(r)} = L_n^{(r)}$.

iii) If $W_n^{(0)} = P_n = W_n(0,1;2,1)$ and $W_0^{(n)} = W_0 = P_0 = 0$, then $W_n^{(r)}$ is the hyper-Pell number, that is, $W_n^{(r)} = P_n^{(r)}$.

The paper is organized as follows: In Section 2 we give some combinatorial properties of the hyper-Fibonacci and hyper-Lucas numbers by using a symmetric algorithm. In Section 3 we generalize the symmetric algorithm introduced in section 2 and, in addition, we generalize the hyper-Horadam numbers as well.

2. A symmetric algorithm

The Euler–Seidel algorithm and its analogues are useful in the study of recurrence relations of some numbers and polynomials [2, 3, 4, 5]. Let (a_n) and (a^n) be two real initial sequences. Then the infinite matrix, which is called symmetric infinite matrix in [4], with entries a_n^k corresponding to these sequences is determined recursively by the formulas

$$\begin{aligned} a_n^0 &= a_n, \quad a_0^n = a^n \quad (n \geq 0)\,, \\ a_n^k &= a_n^{k-1} + a_{n-1}^k \quad (n \geq 1, k \geq 1)\,, \end{aligned}$$

i.e., in matrix form

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & a_n^{k-1} & \cdot & \cdot & \cdot \\ & & & & \downarrow & & & \\ \cdot & \cdot & \cdot & a_{n-1}^k \rightarrow & a_n^k & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

The entries a_n^k (where k is the row index, n is the column index) have the following symmetric relation [4]:

$$a_n^k = \sum_{i=1}^{k} \binom{n+k-i-1}{n-1} a_0^i + \sum_{s=1}^{n} \binom{n+k-s-1}{k-1} a_s^0. \tag{2.1}$$

Dil and Mező [4], by using the relation (2.1), obtained an explicit formula for hyperharmonic numbers, general generating functions of the Fibonacci and Lucas numbers. By using relation (2.1) and the following well known identity [6, p. 160]

$$\sum_{t=a}^{c} \binom{t}{a} = \binom{c+1}{a+1}, \tag{2.2}$$

we have some new findings contained in the following theorems.

Theorem 2.1. *If $n \geq 1$, $r \geq 1$ and $m \geq 0$, then*

$$F_n^{(m+r)} = \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} F_s^{(m)}.$$

Proof. Let $a_n^0 = F_{n+1}^{(m)}$ and $a_0^n = F_1^{(m+n)} = 1$ be given for $n \geq 1$. If we calculate the elements of the corresponding infinite matrix by using the recursive formula (2.1), it turns out that they equal to

$$\begin{pmatrix} F_1^{(m)} & F_2^{(m)} & F_3^{(m)} & F_4^{(m)} & \cdots \\ F_1^{(m+1)} & F_2^{(m+1)} & F_3^{(m+1)} & F_4^{(m+1)} & \cdots \\ F_1^{(m+2)} & F_2^{(m+2)} & F_3^{(m+2)} & F_4^{(m+2)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{2.3}$$

From relation (2.1) it follows that

$$\begin{aligned} a_{n+1}^{r+1} &= \sum_{i=1}^{r+1} \binom{n+r-i+1}{n} + \sum_{s=1}^{n+1} \binom{n+r-s+1}{r} F_{s+1}^{(m)} \\ &= \sum_{i=0}^{r} \binom{n+r-i}{n} + \sum_{s=0}^{n} \binom{n+r-s}{r} F_{s+2}^{(m)} \\ &= \sum_{k=0}^{r} \binom{n+k}{n} + \sum_{b=0}^{n} \binom{r+b}{r} F_{n-b+2}^{(m)}, \end{aligned}$$

where $k = r - i$ and $b = n - s$. From (2.2), we have

$$a_{n+1}^{r+1} = \binom{n+r+1}{n+1} + \sum_{b=0}^{n} \binom{r+b}{r} F_{n-b+2}^{(m)} = \sum_{b=0}^{n+1} \binom{r+b}{r} F_{n-b+2}^{(m)}.$$

Then the matrix (2.3) yields

$$a_{n-1}^{r} = F_n^{(m+r)} = \sum_{b=0}^{n-1} \binom{r+b-1}{r-1} F_{n-b}^{(m)} = \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} F_s^{(m)}.$$

Thus the proof is completed. □

We then can easily deduce an expression for the hyper-Fibonacci numbers which contains the ordinary Fibonacci numbers.

Corollary 2.2. *If $n \geq 1$ and $r \geq 1$, then*

$$F_n^{(r)} = \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} F_s$$

where F_s is the sth Fibonacci number.

The corresponding theorem for the hyper-Lucas numbers is as follows.

Theorem 2.3. *If $n \geq 1$, $r \geq 1$ and $m \geq 0$, then*

$$L_n^{(m+r)} = \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} L_s^{(m)}.$$

Proof. Let $a_n^0 = L_n^{(m)}$ and $a_0^n = L_0^{(m+n)} = 2$ be given for $n \geq 1$. This special case gives the following infinite matrix:

$$\begin{pmatrix} L_0^{(m)} & L_1^{(m)} & L_2^{(m)} & L_3^{(m)} & \cdots \\ L_0^{(m+1)} & L_1^{(m+1)} & L_2^{(m+1)} & L_3^{(m+1)} & \cdots \\ L_0^{(m+2)} & L_1^{(m+2)} & L_2^{(m+2)} & L_3^{(m+2)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{2.4}$$

From the relation (2.1) we get that

$$\begin{aligned} a_n^r &= \sum_{i=1}^{r} \binom{n+r-i-1}{n-1} 2 + \sum_{s=1}^{n} \binom{n+r-s-1}{r-1} L_s^{(m)} \\ &= 2\sum_{i=0}^{r-1} \binom{n+r-i-2}{n-1} + \sum_{s=0}^{n-1} \binom{n+r-s-2}{r-1} L_{s+1}^{(m)} \\ &= 2\sum_{k=0}^{r-1} \binom{n+k-1}{n-1} + \sum_{b=0}^{n-1} \binom{r+b-1}{r-1} L_{n-b}^{(m)}, \end{aligned}$$

where $k = r-i-1$ and $b = n-s-1$. From (2.2), we have

$$a_n^r = 2\binom{n+r-1}{n} + \sum_{b=0}^{n-1} \binom{r+b-1}{r-1} L_{n-b}^{(m)} = \sum_{b=0}^{n} \binom{r+b-1}{r-1} L_{n-b}^{(m)}.$$

Then the matrix (2.4) yields

$$a_n^r = L_n^{(m+r)} = \sum_{b=0}^{n} \binom{r+b-1}{r-1} L_{n-b}^{(m)} = \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} L_s^{(m)},$$

this completes the proof. □

Corollary 2.4. *If $n \geq 1$ and $r \geq 1$, then*

$$L_n^{(r)} = \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} L_s,$$

where L_n is the nth Lucas number.

Theorem 2.5. *If $n \geq 1$ and $r \geq 1$, then*

i) $F_n^{(r)} + L_n^{(r)} = 2F_{n+1}^{(r)}$,

ii) $F_n^{(r)} - L_n^{(r)} = 2F_{n+1}^{(r-1)}$.

Proof. From Corollaries 2.2 and 2.4, we have

$$\begin{aligned} F_n^{(r)} + L_n^{(r)} &= \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} (F_s + L_s) \\ &= \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} (2F_{s+1}) = 2F_{n+1}^{(r)} \end{aligned}$$

and

$$\begin{aligned} F_n^{(r)} - L_n^{(r)} &= \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} (F_s - L_s) \\ &= \sum_{s=0}^{n} \binom{n+r-s-1}{r-1} (2F_{s-1}) = 2F_{n+1}^{(r-1)}. \end{aligned}$$

□

Theorem 2.6. *If $n \geq 1$ and $r \geq 1$, then*

$$\sum_{s=0}^{r} F_n^{(s)} = F_{n+1}^{(r)} - F_{n-1}.$$

Proof. From Corollary 2.2, we have

$$\sum_{s=1}^{r} F_n^{(s)} = \sum_{s=1}^{r} \left(\sum_{t=0}^{n} \binom{n+s-t-1}{s-1} F_t \right) = \sum_{t=0}^{n} \left(F_t \sum_{s=1}^{r} \binom{n+s-t-1}{s-1} \right).$$

From (2.2), we obtain

$$\sum_{s=1}^{r} F_n^{(s)} = \sum_{t=0}^{n} \binom{n+r-t}{r-1} F_t = \sum_{t=0}^{n+1} \binom{n+r-t}{r-1} F_t - F_{n+1} = F_{n+1}^{(r)} - F_{n+1}.$$

Thus

$$\sum_{s=0}^{r} F_n^{(s)} = F_{n+1}^{(r)} - F_{n-1}.$$

□

Theorem 2.7. *If* $n \geq 1$ *and* $r \geq 1$, *then*

$$\sum_{s=0}^{r} L_n^{(s)} = L_{n+1}^{(r)} - L_{n-1}.$$

Proof. The proof is similar to the proof of Theorem 2.6. □

3. A generalized symmetric algorithm

In this section we generalize the algorithm for determining a_n^k in the symmetric infinite matrix. To this end we fix two arbitrary, nonzero real numbers u and v. Then our new algorithm reads as

$$\begin{aligned} a_n^0 &= a_n, \quad a_0^n = a^n \quad (n \geq 0)\,, \\ a_n^k &= u a_n^{k-1} + v a_{n-1}^k \quad (n \geq 1, k \geq 1)\,. \end{aligned}$$

That is, the symmetric infinite matrix now can be constructed in the following way:

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & u a_n^{k-1} & \cdot & \cdot & \cdot \\ & & & & \downarrow & & & \\ \cdot & \cdot & \cdot & v a_{n-1}^k \rightarrow & a_n^k & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

It can easily be seen that (2.1) generalizes to

$$a_n^k = \sum_{i=1}^{k} v^n u^{k-i} \binom{n+k-i-1}{n-1} a_0^i + \sum_{s=1}^{n} v^{n-s} u^k \binom{n+k-s-1}{k-1} a_s^0. \tag{3.1}$$

As an application, we can generalize the hyper-Horadam number as

$$W_n^{(r)}(u,v) = u W_n^{(r-1)} + v W_{n-1}^{(r)}$$

where u and v are two nonzero real parameters and the initial conditions are $W_n^{(0)}(u,v) = W_n(a,b;p,q) = W_n$ and $W_0^{(n)}(u,v) = W_0(a,b;p,q) = a$. Some special cases of the hyper-Horadam numbers $W_n^{(r)}(u,v)$ are:

i) If $W_n^{(0)}(u,v) = F_n^{(0)}(u,v) = F_n$ and $W_0^{(n)}(u,v) = F_0^{(n)}(u,v) = 0$, then we have the generalized hyper-Fibonacci numbers defined as

$$F_n^{(r)}(u,v) = u F_n^{(r-1)} + v F_{n-1}^{(r)},$$

ii) If $W_n^{(0)}(u,v) = L_n^{(0)}(u,v) = L_n$ and $W_0^{(n)}(u,v) = L_0^{(n)}(u,v) = 2$, we have the generalized hyper-Lucas number defined as

$$L_n^{(r)}(u,v) = uL_n^{(r-1)} + vL_{n-1}^{(r)},$$

iii) If $W_n^{(0)}(u,v) = P_n^{(0)}(u,v) = P_n$ and $W_0^{(n)}(u,v) = P_0^{(n)}(u,v) = 0$, we have the generalized hyper-Pell number defined as

$$P_n^{(r)}(u,v) = uP_n^{(r-1)} + vP_{n-1}^{(r)}.$$

By using (3.1), Theorem 2.1 generalizes to the following Theorem.

Theorem 3.1. *If $n \geq 1$, $r \geq 1$ and $m \geq 0$, then*

$$\begin{aligned} W_n^{(m+r)}(u,v) = a\left(\frac{v}{1-u}\right)^n &\left[1 - rB_u(r,n)\binom{n+r-1}{n-1}\right] \\ &+ u^r \sum_{s=1}^{n} v^{n-s} \binom{n+r-s-1}{r-1} W_s^{(m)}(u,v). \end{aligned}$$

where $B_u(r,n)$ is the incomplete beta function [1].

Proof. The incomplete beta function $B_u(r,n)$ appears when we would like to evaluate the sum

$$\sum_{k=0}^{r-1} \binom{n+k-1}{k} u^k.$$

This sum equals to

$$\frac{1}{(1-u)^n}\left[1 - rB_u(r,n)\binom{n+r-1}{n-1}\right].$$

This is the most compact form we could find. The other parts of the proof are the same as the proof of Theorem 2.1, if we use relation (3.1) and assume that $a_n^0 = W_n^{(m)}(u,v)$ and $a_0^n = W_0^{(m+n)} = a$. $\square$

Corollary 3.2. *If $n \geq 1$ and $r \geq 1$, then*

$$\begin{aligned} W_n^{(r)}(u,v) = a\left(\frac{v}{1-u}\right)^n &\left[1 - rB_u(r,n)\binom{n+r-1}{n-1}\right] \\ &+ u^r \sum_{s=1}^{n} v^{n-s} \binom{n+r-s-1}{r-1} W_s. \end{aligned}$$

From these results we have some particular results for the hyper-Fibonacci, hyper-Lucas, hyper-Pell numbers and their generalizations such as

$$F_n^{(r)}(u,v) = u^r \sum_{s=1}^{n} v^{n-s} \binom{n+r-s-1}{r-1} F_s,$$

$$L_n^{(r)}(u,v) = 2\left(\frac{v}{1-u}\right)^n \left[1 - rB_u(r,n)\binom{n+r-1}{n-1}\right] + u^r \sum_{s=1}^{n} v^{n-s}\binom{n+r-s-1}{r-1} L_s,$$

$$P_n^{(r)}(u,v) = u^r \sum_{s=1}^{n} v^{n-s}\binom{n+r-s-1}{r-1} P_s,$$

$$P_n^{(r)} = \sum_{s=1}^{n} \binom{n+r-s-1}{r-1} P_s,$$

where F_s, L_s and P_s is the s^{th} Fibonacci, Lucas and Pell number, respectively.

References

[1] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, Dover, 1965.

[2] N.-N. Cao, F.-Z. Zhao, Some Properties of Hyperfibonacci and Hyperlucas Numbers, Journal of Integer Sequences, 2010; 13: Article 10.8.8.

[3] A. Dil, V. Kurt, M. Cenkci, Algorithms for Bernoulli and allied polynomials, Journal of Integer Sequences, (2007); 10: Article 07.5.4.

[4] A. Dil and I. Mező, A symmetric algorithm hyperharmonic and Fibonacci numbers, Applied Mathematics and Computation 2008; 206: 942–951.

[5] D. Dumont, Matrices d'Euler–Seidel, Seminaire Lotharingien de Combinatorie, 1981, B05c. Available online at
http://www.emis.de/journals/SLC/opapers/s05dumont.pdf

[6] R. L. Graham, D. E. Knuth, O. Patashnik, Concrete Mathematics, Addison Wesley, 1993.

[7] A. F. Horadam, Basic properties of a certain generalized sequence of numbers, The Fibonacci Quarterly 1965; 3: 161–176.

Annales Mathematicae et Informaticae
43 (2014) pp. 29–41
http://ami.ektf.hu

An unusual cubic representation problem

Andrew Bremner[a], Allan Macleod[b]

[a]School of Mathematics and Mathematical Statistics, Arizona State University
bremner@asu.edu

[b]Mathematics Group, University of the West of Scotland
Allan.MacLeod@uws.ac.uk

Submitted September 24, 2014 — Accepted December 12, 2014

Abstract

For a non-zero integer N, we consider the problem of finding 3 integers (a, b, c) such that
$$N = \frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b}.$$
We show that the existence of solutions is related to points of infinite order on a family of elliptic curves. We discuss strictly positive solutions and prove the surprising fact that such solutions do not exist for N odd, even though there may exist solutions with one of a, b, c negative. We also show that, where a strictly positive solution does exist, it can be of enormous size (trillions of digits, even in the range we consider).

Keywords: cubic representation, elliptic curve, rational points.

MSC: Primary 11D25 11G05, Secondary 11Y50

1. Introduction

Several authors have considered the problem of representing integers N (and in particular, positive integers N) by a homogeneous cubic form in three variables. See, for example, the papers of Bremner & Guy [1], Bremner, Guy, and Nowakowski [2], Brueggeman [3]. Analysis for cubic forms is made possible by the fact the the resulting equation is that of a cubic curve, and hence in general is of genus one.

In this note, we shall study the representation problem

$$N = \frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b}, \tag{1.1}$$

where a, b, c are rationals. Equivalently, by homogeneity, we may consider $a, b, c \in \mathbb{Z}$.

Studying numerical data, it was observed that, when N is odd, there seem never to exist “positive” solutions of (1.1), that is, solutions with $a, b, c > 0$, even though there may indeed exist solutions with one of a, b, c negative. This fact precludes a simple congruence argument to show the non-existence of positive solutions. In contrast, when N is even, there may or may not be positive solutions. The proof we give of non-existence of positive solutions, for N odd, depends on local considerations at judiciously chosen primes.

In investigating the existence of solutions to (1.1), and more specifically, existence of positive solutions, we discovered that on occasion solutions exist, but the smallest positive solution may be rather large. For example, when $N = 896$, the smallest positive solution has a, b, c with several trillion digits (we do not list it explicitly).

2. The cubic curve

We consider the following problem, that of representing integers N in the form

$$N = \frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b}$$

for rationals (or, by homogeneity, integers) a, b, c.

For fixed N, the homogenization is a cubic curve

$$C_N : \; N(a+b)(b+c)(c+a) = a(a+b)(c+a) + b(b+c)(a+b) + c(c+a)(b+c)$$

in projective 2-dimensional space which has a rational point, for example, $(a, b, c) = (1, -1, 0)$. The curve is therefore elliptic, and a cubic model is readily computed in the form

$$E_N : \; y^2 = x^3 + (4N^2 + 12N - 3)x^2 + 32(N+3)x.$$

Setting $s = a + b + c$, maps are given by

$$\frac{a}{s} = \frac{8(N+3) - x + y}{2(4-x)(N+3)}, \quad \frac{b}{s} = \frac{8(N+3) - x - y}{2(4-x)(N+3)}, \tag{2.1}$$

and

$$\frac{c}{s} = \frac{-4(N+3) - (N+2)x}{(4-x)(N+3)},$$

with inverse

$$x = \frac{-4(a+b+2c)(N+3)}{(2a+2b-c) + (a+b)N}, \quad y = \frac{4(a-b)(N+3)(2N+5)}{(2a+2b-c) + (a+b)N}.$$

The curve has discriminant

$$\Delta(E_N) = 2^{14}(N+3)^2(2N-3)(2N+5)^3$$

so $\Delta > 0$ for all integers N except $-3, -2, -1, 0, 1$. Thus, other than for these five values, the defining cubic has three real roots, and the elliptic curve has two components. There is an unbounded component with $x \geq 0$, and a bounded component with $x < 0$ (frequently referred to as the "egg").

Lemma 2.1. *The torsion subgroup of E_N is isomorphic to $\mathbb{Z}/6\mathbb{Z}$, except when $N = 2$, when it is isomorphic to $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$.*

Proof. The point $(0, 0)$ is clearly of order 2.

For there to be three rational points of order 2, necessarily there must be rational roots of

$$x^2 + (4N^2 + 12N - 3)x + 32(N + 3) = 0,$$

implying $(2N - 3)(2N + 5) = (2N + 1)^2 - 16 = \Box$, with the only integer possibility $N = 2$.

Points of order 3 are points of inflexion of the curve E_N, and it is a standard exercise in calculus that

$$(4, \pm 4(2N + 5))$$

is such a point.

Points of order 2 and of order 3 imply a point (x, y) of order 6, which by the duplication formula, must satisfy

$$\frac{(x^2 - 32(N + 3))^2}{4(x^3 + (4N^2 + 12N - 3)x^2 + 32(N + 3)x)} = 4$$

giving the points $\pm T_0$ of order 6, where

$$T_0 = (\,8(N + 3), 8(N + 3)(2N + 5)\,).$$

Note: the corresponding torsion point in $C_N(\mathbb{Q})$ is the point $(-1, 1, 1)$.

Further, there can be no point of order 12. For such can arise only when T_0 is divisible by 2, implying $8(N + 3) = \Box$. Then from the duplication formula, the following equation

$$(U^2 - 32(N + 3))^2 = 32(N + 3)(U^3 + (4N^2 + 12N - 3)U^2 + 32(N + 3)U)$$

must have a rational root for U. Substituting $N = 2K^2 - 3$,

$$(U^2 + 8K(1 - 4K - 4K^2)U + 64K^2)(U^2 + 8K(-1 - 4K + 4K^2)U + 64K^2) = 0,$$

which demands

$$K(2K - 1)(2K + 1)(2K + 3)(2K - 3) = 0,$$

leading to singular curves.

Thus the torsion group is cyclic of order 6 when $N \neq 2$, and is isomorphic to $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$ when $N = 2$. $\Box$

Remark 2.2. The torsion points themselves lead to singular solutions to the original problem, so we need points of infinite order for a finite non-trivial solution. Thus the rank of E_N must be at least 1. The first example, for positive N, is $N = 4$ with a generator for $E_4(\mathbb{Q})$ given by $G = (-4, 28)$. The formulae above give the integer solution $a = 11$, $b = 4$ and $c = -1$. We have $(-4, 28) + (0, 0) = (-56, -392)$ which gives $a = -5$, $b = 9$ and $c = 11$. Adding the other four torsion points gives cyclic permutations of these basic solutions. The point $9G$ is the smallest multiple of G that corresponds to a positive solution (in which $a, b, c \sim 10^{80}$).

Remark 2.3. The torsion points for $N > 0$ all lie on the unbounded component of E_N.

Remark 2.4. In the group $C_N(\mathbb{Q})$, the inverse of the point (a, b, c) is the point (b, a, c). Further, adding the torsion generator $(-1, 1, 1)$ to (a, b, c) gives rise to the order six automorphism ϕ of C given by

$$\begin{aligned}\phi(a,b,c) = \big(&a^2 + ab - ac + bc - 2c^2 - (b^2 - c^2)N,\\ &a^2 + 3ab + 2b^2 + 3ac + bc + 2c^2 - (b + c)(2a + b + c)N,\\ &(a + b)(a - 2b + c) + (b^2 - c^2)N\big).\end{aligned}$$

Then $\phi^2(a, b, c) = (b, c, a)$, $\phi^4(a, b, c) = (c, a, b)$.

Remark 2.5. The torsion group of $E_N(\mathbb{Q})$ is cyclic of order 6, and so there exist isogenies of E_N of degrees 2, 3, 6, which are readily computed from the formulae in Vélu [8] and which we record here in the following Lemma.

Lemma 2.6. *For $i = 2, 3, 6$, there are the following isogenies $\phi_i \colon E_N \to E_N^{(i)}$ of degree i.*

1. $E_N^{(2)} : \ Y^2 = X^3 - 2(4N^2 + 12N - 3)X^2 + (2N - 3)(2N + 5)^3X,$
   $\phi_2(x, y) = (y^2/x^2, (x^2 - 32(N + 3))y/x^2);$
2. $E_N^{(3)} : \ Y^2 = X^3 + (4N^2 + 60N + 117)X^2 + 128(N + 3)^3X,$
   $$\begin{aligned}\phi_3(x, y) = (&x(x - 8(N + 3))^2/(x - 4)^2,\\ &(x - 8(N + 3))(x^2 + 4(2N + 3)x + 32(N + 3))y/(x - 4)^3);\end{aligned}$$
3. $E_N^{(6)} : \ Y^2 = X^3 - 2(4N^2 + 60N + 117)X^2 + (2N - 3)^3(2N + 5)X,$
   $$\begin{aligned}\phi_6(x, y) = (&(x^2 + 4(2N + 3)x + 32(N + 3))^2y^2/(x(x - 4)(x - 8(N + 3)))^2,\\ &p_1(x)p_2(x)p_3(x)y/(x^2(x - 4)^3(x - 8(N + 3))^3)),\end{aligned}$$

where

$$p_1(x) = x^2 - 32(N + 3), \quad p_2(x) = x^2 + 4(2N + 3)x + 32(N + 3),$$

$$\begin{aligned}p_3(x) = x^4 &- 32(N + 3)x^3 - 32(N + 3)(4N^2 + 12N - 1)x^2 - 1024(N + 3)^2x\\ &+ 1024(N + 3)^2.\end{aligned}$$

3. Rational solutions for $N > 0$

We computed the rank of E_N in the range $1 \le N \le 1000$, and in the case of positive rank, attempted to compute a set of generators. The existence of 2-, 3-, and 6-isogenies was particularly helpful when treating the curves with generators of large height, in that we could focus on the curve where the estimated size of a generator was minimal.

In the range of N we consider, the curve with generator of largest height is E_{616}, where the rank is one, and a generator has height ~ 672.28. This was discovered by finding a point of height ~ 224.09 on a 3-isogenous curve. Most of these rank computations were feasible using programs written in Pari-GP; the very largest points were found with the help of Magma [5]. The rank results are summarized in the following table.

# rank 0	# rank 1	# rank 2	# rank 3
436	485	76	3

Table 1: Rank distribution for $1 \le N \le 1000$

Rank one examples occur for $N = 4, 6, 10, 12, \ldots$, rank two examples for $N = 34, 94, 98, 111, \ldots$, and rank three examples for $N = 424, 680, 975$.

4. Positivity

Henceforth, we assume that $N > 0$. A natural question is do positive solutions a, b, c of the original equation exist? In particular, how do we recognise points $(x, y) \in E_N(\mathbb{Q})$ that correspond to positive solutions of (1.1)?

Theorem 4.1. *Suppose $(a, b, c) \in C_N$ corresponds to $(x, y) \in E_N(\mathbb{Q})$. Then $a, b, c > 0$ if and only if either*

$$\tfrac{(3-12N-4N^2-(2N+5)\sqrt{4N^2+4N-15}}{2} < x < -2(N+3)(N+\sqrt{N^2-4}), \tag{4.1}$$

or

$$-2(N+3)(N-\sqrt{N^2-4}) < x < -4\left(\frac{N+3}{N+2}\right). \tag{4.2}$$

Proof. Suppose $a, b, c > 0$. From (2.1),

$$0 < \frac{ab}{s^2} = \tfrac{(4-x)\left(x^2+4N(N+3)x+16(N+3)^2\right)}{4(N+3)^2(4-x)^2}, \tag{4.3}$$

so necessarily $x < 4$; and then $c/s > 0$ implies

$$x < -4\left(\frac{N+3}{N+2}\right)$$

(and, in particular, the point (x, y) lies on the egg). By symmetry in a, b, we may suppose $y > 0$. From (2.1),

$$\frac{a}{s} = \frac{8(N+3) - x + y}{2(4-x)(N+3)} > 0.$$

It remains to ensure that $b/s = \frac{8(N+3)-x-y}{2(4-x)(N+3)} > 0$. But from (4.3), $b/s > 0$ precisely when

$$x^2 + 4N(N+3)x + 16(N+3)^2 > 0;$$

and this latter happens when

$$\text{either} \quad x < -2(N+3)(N + \sqrt{N^2-4}), \quad \text{or} \quad x > -2(N+3)(N - \sqrt{N^2-4}).$$

Putting these results together, necessary conditions for a, b, c to be positive are the following:

$$\frac{1}{2}(3 - 12N - 4N^2 - (2N+5)\sqrt{4N^2+4N-15}) < x < -2(N+3)(N + \sqrt{N^2-4}),$$

where the left inequality is automatic, arising from $y^2 > 0$, or

$$-2(N+3)(N - \sqrt{N^2-4}) < x < -4\left(\frac{N+3}{N+2}\right).$$

It is straightforward to see that these conditions on x, y are now also sufficient for the positivity of a, b, c. $\square$

It follows that positive solutions can only come from rational points on the egg component of the curve.

5. *N* odd

Analyzing solutions found from computation, it was observed that when N is odd (in contrast to the case N even) there seem never to be points on the curve E_N with $x < 0$. We show that this is indeed the case.

Theorem 5.1. *Suppose* $N \equiv 1 \bmod 2$. *Then* $(x, y) \in E_N(\mathbb{Q})$ *implies* $x \geq 0$.

Proof. Set $N + 3 = 2M$, $M \geq 2$, so that the curve E_N takes the form

$$E_M : \; y^2 = x(x^2 + (16M^2 - 24M - 3)x + 64M).$$

A point $(x, y) \in E_M(\mathbb{Q})$ satisfies $x = dr^2/s^2$, for $d, r, s \in \mathbb{Z}$, $(r, s) = 1$, with $d \mid 64M$, and without loss of generality, d squarefree. Then

$$dr^4 + (16M^2 - 24M - 3)r^2 s^2 + \frac{64M}{d}s^4 = \square.$$

The claim is that this quartic can have no points r, s when $d < 0$.

On completing the square

$$\bigl(2dr^2 + (16M^2 - 24M - 3)s^2\bigr)^2 - (4M-1)^3(4M-9)s^4 = 4d\,\Box, \qquad d \mid 2M.$$

Case I: $d < 0$, d odd.
Let $d = -u$, $u > 0$, u odd, with $M = um$. We now have

$$-ur^4 + (16M^2 - 24M - 3)r^2s^2 - 64ms^4 = \Box,$$

equivalently,

$$(-2ur^2 + (16M^2 - 24M - 3)s^2)^2 - (4M-1)^3(4M-9)s^4 = -4u\Box.$$

Note that the Jacobi symbol

$$\begin{aligned}\left(\frac{-u}{4M-1}\right) = \left(\frac{-1}{4M-1}\right)\left(\frac{u}{4M-1}\right) &= -\left(\frac{4M-1}{u}\right)(-1)^{(u-1)/2}\\ &= -\left(\frac{-1}{u}\right)(-1)^{(u-1)/2} = -1.\end{aligned}$$

However, if every prime p dividing $4M-1$ satisfies $\left(\frac{-u}{p}\right) = +1$, then the Jacobi symbol $\left(\frac{-u}{4M-1}\right) = +1$ by multiplicativity of the symbol. In consequence, there exists a prime p dividing $4M-1$ satisfying $\left(\frac{-u}{p}\right) = -1$. Then for such a prime p,

$$\begin{aligned}-2ur^2 + (16M^2 - 24M - 3)s^2 &\equiv 0 \bmod p,\\ -2ur^2 - 8s^2 &\equiv 0 \bmod p,\\ 4s^2 &\equiv -ur^2 \bmod p,\end{aligned}$$

forcing $r \equiv s \equiv 0 \bmod p$, contradiction.

Case II: $d < 0$, d even.
Let $d = -2u$, $u > 0$, u odd, with $M = um$. We now have

$$-2ur^4 + (16M^2 - 24M - 3)r^2s^2 - 32ms^4 = \Box,$$

equivalently

$$(-4ur^2 + (16M^2 - 24M - 3)s^2)^2 - (4M-1)^3(4M-9)s^4 = -8u\,\Box.$$

Subcase (i): M even.
Now

$$\left(\frac{-2u}{4M-1}\right) = \left(\frac{-2}{4M-1}\right)\left(\frac{4M-1}{u}\right)(-1)^{(u-1)/2} = -\left(\frac{-1}{u}\right)(-1)^{(u-1)/2} = -1,$$

and arguing as above, there exists a prime p dividing $4M-1$ with $\left(\frac{-2u}{p}\right) = -1$. Then

$$\begin{aligned} -4ur^2 + (16M^2 - 24M - 3)s^2 &\equiv 0 \bmod p, \\ -4ur^2 - 8s^2 &\equiv 0 \bmod p, \\ 4s^2 &\equiv -2ur^2 \bmod p, \end{aligned}$$

forcing $r \equiv s \equiv 0 \bmod p$, contradiction.

Subcase (ii): M odd (so in particular, m odd).
In this case, the equation is 2-adically unsolvable, as follows. We have

$$-2ur^4 + (16M^2 - 24M - 3)r^2s^2 - 32ms^4 = \square$$

implying s is odd. Modulo 4, r cannot be odd, and thus r is even. Then

$$-3(r/2)^2 \equiv \square \bmod 8$$

so that $r/2$ is even; and now

$$-3(r/4)^2 - 2m \equiv \square \bmod 4,$$

$$-3(r/4)^2 - 2 \equiv \square \bmod 4,$$

impossible. $\square$

Corollary 5.2. *If N is odd and E_N is of positive rank, then generators for $E_N(\mathbb{Q})$ lie on the unbounded component of E_N.*

Consequently, in the situation of Corollary 5.2, there are no rational points on the egg, so no positive solutions of (1.1) exist. This happens when the rank is one for $N = 19, 21, 23, 29, \ldots$, when the rank is two, for $N = 111, 131, 229, 263, \ldots$, and when the rank is three, for $n = 975$. It can also occur that when N is even, all generators for $E_N(\mathbb{Q})$ lie on the unbounded component of E_N, so that there are no rational points on the egg. This situation occurs for rank one examples $N = 40, 44, 50, 68, \ldots$, rank two examples $N = 260, 324, 520, 722, \ldots$, and the rank three example $N = 680$.
Hence there exist even N, namely $N = 40, 44, 50, 68, \ldots$ where there exist solutions to (1.1), but there do not exist positive solutions. In contrast, we have the following result.

Theorem 5.3. *There exist infinitely many positive even integers N such that (1.1) has positive solutions.*

Proof. The proof is immediate, using the parameterization $N = t^2 + t + 4$ with the point on the egg of E_N given by

$$(x, y) = \left(-4(t^2+t+1)^2,\ 4(2t+1)(t^2+t+1)(3t^2+3t+7)\right). \qquad \square$$

Remark 5.4. It is straightforward to show that this point corresponds to

$$a = (t^2+1)(3t^3+8t^2+14t+11), \quad b = -(t^2+2t+2)(3t^3+t^2+7t-2),$$

$$c = t^6+3t^5+11t^4+17t^3+20t^2+12t-1;$$

with no (real) value of t making $a, b, c > 0$; so some multiple of the point will be needed to obtain a positive solution.

6. Size of positive solutions

A positive solution of (1.1) demands the existence of a point in $E_N(\mathbb{Q})$ that lies on the egg; and in particular not all generators for $E_N(\mathbb{Q})$ can lie on the unbounded branch of the curve. For a positive solution, therefore, a generator in $E_N(\mathbb{Q})$ must lie on the egg.

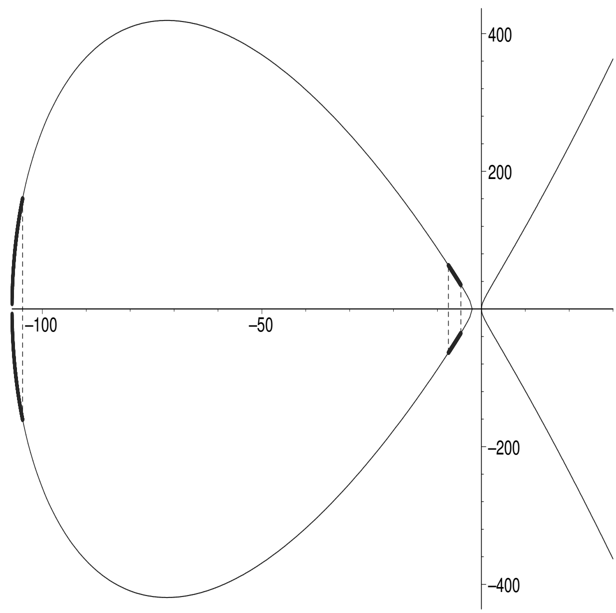


Figure 1: Region for $a, b, c > 0$ on $E_4(\mathbb{Q})$

Such a point may not satisfy the inequalities (4.1), (4.2), of course, but a result of Hurwitz [4] implies that rational points on E_N are now dense on both components of E_N, so that there will indeed exist points in $E_N(\mathbb{Q})$ that satisfy (4.1), (4.2). The width of the interval at (4.1) tends to 1 as $N \to \infty$, and the width of the interval at (4.2) tends to 0. The width of the egg however is $O(N^2)$. Thus if the rank of the curve E_N is equal to one, with a generator P on the egg regarded as lying essentially at random on the egg, then the smallest integer m such that mP satisfies (4.1), (4.2), may be very large. If we assume equidistribution of random points on the egg, then a crude estimate of arc length shows that there is probability $O(\frac{1}{N})$ that a random point of the egg lies within the region defined by (4.1), (4.2). In Figure 1 we sketch the graph for $N = 4$, indicating the region corresponding to positive a, b, c, representing the intervals $-106.9046 < x < -104.4974$, $-7.5026 < x < -4.6667$. We consider curves E_N of rank one in the range $1 \leq N \leq 1000$, where there is a generator P of $E_N(\mathbb{Q})$ lying on the egg. For these curves, we computed the smallest integer m such that one of the points $mP + T$, $T \in \mathrm{Tor}(E_N(\mathbb{Q}))$, satisfies (4.1), (4.2).

We then computed the maximum number of digits in a, b, c. The results for $1 \leq N \leq 200$ are given in Table 2.

N	m	# digits	N	m	# digits	N	m	# digits
4	9	81	48	311	418086	136	65	26942
6	11	134	58	221	244860	146	307	259164
10	13	190	60	61	9188	156	353	12046628
12	35	2707	66	107	215532	158	1211	15097279
14	47	1876	76	65	23662	162	457	1265063
16	11	414	82	157	85465	178	2945	398605460
18	49	10323	92	321	252817	182	853	2828781
24	107	33644	102	423	625533	184	851	20770896
28	121	81853	112	223	935970	186	643	5442988
32	65	14836	116	101	112519	196	701	11323026
38	659	1584369	126	75	196670	198	121	726373
42	419	886344	130	707	8572242	200	2957	71225279
46	201	198771	132	461	3607937			

Table 2: The maximum number of digits in a, b, c in the range $1 \leq N \leq 200$

For comparison, the twenty volume second edition of the Oxford English Dictionary is estimated to contain 350 million printed characters (see [6]), a little less than the number of digits in each of a, b, c when $n = 178$.

It is not practical to compute points on elliptic curves with heights that begin to exceed those of the previous table. For example, when $N = 896$, the curve E_{896} has rank one, and the smallest multiple of the generator P (which itself has height 128.76) such that mP corresponds to a positive solution at (1.1), is given by $m = 161477$.

Remark 6.1. Such computations were performed using high-precision real arithmetic. Computing a multiple mP takes $O(\log_2(m))$ operations, so in the computed range where $m \leq 161477$, precision is not a problem. For safety however, and because the computation took only slightly longer, we worked with 10^6 digits of precision.

7. Size bounds on positive solutions

We proceed to determine a crude lower bound for the number of digits in the positive solution a, b, c from a knowledge of the canonical height of the corresponding point on E_N.

Suppose that $P(x, y)$, where $x < 0$, is a point on E_N giving rise to a positive a, b, c. So one of the inequalities (4.1), (4.2) holds, and in particular x is negative (in fact, $x < -4$).

Theorem 7.1. *Let $(a, b, c) \in C_N(\mathbb{Q})$ correspond to $P(x, y) \in E_N(\mathbb{Q})$, and suppose that $a, b, c > 0$. Then*

$$\max(\log(a), \log(b)) > \frac{3}{2}h(P) - 6\log(N) - 10,$$

where h denotes the canonical height function on E_N.

Proof. The mapping $E_N \to C_N$ is given by

$$a : b : c = -x + y + 8(N+3) : -x - y + 8(N+3) : -2x(N+2) - 8(N+3).$$

Write $x = -u/w^2$, $y = v/w^3$, where $u > 0$, $w > 0$, and $(u, w) = (v, w) = 1$. Since $x < -1$, we have $u > w^2$, and the naive height $H(P)$ of P is equal to u.

Either inequality (4.1), (4.2), implies

$$u/w^2 < \frac{1}{2}(-3 + 12N + 4N^2 + (2N+5)(2N+1)) = 4N^2 + 12N + 1 < (2N+3)^2.$$

Write

$$\begin{aligned} ah &= uw + v + 8(N+3)w^3, \\ bh &= uw - v + 8(N+3)w^3, \\ ch &= 2uw(N+2) - 8(N+3)w^3, \end{aligned}$$

where h is the greatest common divisor of the three expressions on the right.

Now $ah + bh = 2(u + 8(N+3)w^2)w$ so that

$$\begin{aligned} ah + bh = 2(u + 8(N+3)w^2)w &> 2\left(1 + \frac{8(N+3)}{(2N+3)^2}\right)uw \\ &= \frac{2(4N^2 + 20N + 33)}{(2N+3)^2}uw \\ &> \frac{2(4N^2 + 20N + 33)}{(2N+3)^3}u^{3/2} \\ &> \frac{1}{N}H(P)^{3/2}. \end{aligned}$$

Thus necessarily either ah or bh (and by choice of sign of y, we may assume this is ah) is at least equal to $\frac{1}{2N}H(P)^{3/2}$.

We now estimate h. We have $(a-b)h = 2v$, $(a+b)h = -2uw + 16(N+3)w^3$, $(a+b+2c)h = -2uw(2N+5)$, so that $((N+2)(a+b)-c)h = 8(N+3)(2N+5)w^3$. Thus $h \mid 8(N+3)(2N+5)w^3$. Now if p is a prime dividing (h,w), necessarily $p \mid 2v$, so that $p=2$, since $(v,w)=1$. But w even implies v odd, so that $2\|h$. Moreover, in the case that w is odd, then $(h,w)=1$. It follows that $h \mid 8(N+3)(2N+5)$, and in particular, $h \le 8(N+3)(2N+5)$. This bound is best possible, in that in our range of computation, there are several instances where $h = 8(N+3)(2N+5)$. Consequently, a is at least equal to $H(P)^{3/2}/(16N(N+3)(2N+5))$.

There are known bounds on the difference between the canonical height and the logarithm of the naive height, in the form

$$c_1 \le \log H(P) - h(P) \le c_2,$$

for constants c_1, c_2. The following estimate for c_1 is taken from Silverman [7, Theorem 1.1], where $\Delta(E_N)$ and $j(E_N)$ denote the discriminant and j-invariant of E_N, respectively.

$$\begin{aligned} c_1 &= -\frac{1}{12}\log|\Delta(E_N)j(E_N)| - \frac{1}{2}\log|\frac{4N^2+12N-3}{3}| - \frac{1}{2}\log(2) - 1.07 \\ &= -\frac{1}{4}\log|\frac{(2N+3)(8N^3+36N^2+6N-93)(4N^2+12N-3)^2}{9}| - \frac{3}{2}\log(2) - 1.07 \\ &> -\frac{1}{4}\log(226N^8) - \frac{3}{2}\log(2) - 1.07 \quad (\text{for } N \ge 4) \\ &> -2\log(N) - 3.47. \end{aligned}$$

Thus,

$$\begin{aligned} \log a &> \frac{3}{2}\log H(P) - \log(16N(N+3)(2N+5)) \\ &> \frac{3}{2}h(P) + \frac{3}{2}c_1 - \log(16N(N+3)(2N+5)) \\ &> \frac{3}{2}h(P) - 3\log(N) - 5.20 - \log(92N^3) \quad (\text{for } N \ge 4) \\ &> \frac{3}{2}h(P) - 6\log(N) - 10. \end{aligned}$$ □

Remark 7.2. In the case where (4.2) holds, the above bound may be improved to $\log a > \frac{3}{2}h(P) - 4\log(N) - 9$.

When $N = 896$, with multiple $m = 161477$, then $x(161477P) \sim -4.0133512$, so that (4.2) holds. Now $h(P) = 3357394890723.0389$ and the above estimate gives $\log a > 5036092336048.36658$. That is, a has in excess of 2.187 trillion digits (which amounts to about 6250 OED units).

Remark 7.3. This estimate is very crude. For example, when $N = 178$, with multiple 2945, then $h(P) = 265736973.117$ and the above estimate gives $\log(a) > 398605418.5847$, that is, a has in excess of 173112134 digits. From the table in section 4, we see that the actual number of digits is equal to 398605460.

8. Rational solutions for $N < 0$

The motivation has been to study the case $N > 0$ because of our interest in positive solutions. But in the course of the investigation we also computed ranks and generators for all $-1 \geq N \geq -1000$ ($N \neq -3$). The rank results are summarized in the following table.

# rank 0	# rank 1	# rank 2	# rank 3
393	471	126	9

Table 3: Rank distribution for $-1000 \leq N \leq -1$, $N \neq -3$

Rank one examples occur for $N = -5, -8, -9, \ldots$, rank two examples for $N = -17, -29, -38, \ldots$, and rank three examples for $N = -181, -365, -369, \ldots$. The generator of greatest height occurs for $N = -994$, where the height is ~ 690.84.

References

[1] Bremner, A., Guy, R.K., Two more representation problems, *Proc. Edin. Math. Soc.* Vol. 40 (1997), 1–17.

[2] Bremner,A., Guy, R.K., Nowakowski, R.J., Which integers are representable as the product of the sum of three integers with the sum of their reciprocals, *Math. Comp.* Vol. 61 (1993), no. 203, 117–130.

[3] Brueggeman, S., Integers representable by $(x + y + z)^3/xyz$, *Internat. J. Math. Math. Sci.* Vol. 21 (1998), no. 1, 107—116.

[4] Hurwitz, A., Über ternäre diophantische Gleichungen dritten Grades, *Vierteljahrschrift d. Naturforsch. Ges. Zürich* Vol. 62 (1917), 207–229.

[5] Bosma, W., Cannon, J., Playoust, C. The Magma algebra system. I. The user language, *J. Symbolic Comput.* Vol. 24 (1997), 235–265.

[6] `http://public.oed.com/history-of-the-oed/dictionary-facts/`

[7] Silverman, J.H., The difference between the Weil height and the canonical height on elliptic curves, *Math. Comp.* Vol. 55 (1990), no. 192, 723—743.

[8] Vélu, J., Isogénies entre courbes elliptiques, *C.R. Acad. Sci. Paris (A)*, Vol. 273, (1971), 238-241.

Annales Mathematicae et Informaticae
43 (2014) pp. 43–54
http://ami.ektf.hu

Counting permutations by cyclic peaks and valleys

Chak-On Chow[a], Shi-Mei Ma[b], Toufik Mansour[c*]
Mark Shattuck[d]

[a]Division of Science and Technology, BNU-HKBU United International College, Zhuhai 519085, P.R. China
cchow@alum.mit.edu

[b]School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Hebei 066004, P.R. China
shimeimapapers@gmail.com

[c]Department of Mathematics, University of Haifa, 3498838 Haifa, Israel
tmansour@univ.haifa.ac.il

[d]Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA
shattuck@math.utk.edu

Submitted June 29, 2014 — Accepted November 14, 2014

Abstract

In this paper, we study the generating functions for the number of permutations having a prescribed number of cyclic peaks or valleys. We derive closed form expressions for these functions by use of various algebraic methods. When restricted to the set of derangements (i.e., fixed point free permutations), the evaluation at -1 of the generating function for the number of cyclic valleys gives the Pell number. We provide a bijective proof of this result, which can be extended to the entire symmetric group.

Keywords: Derangements; Involutions; Pell numbers; Cyclic valleys

MSC: 05A05; 05A15

*Corresponding author

1. Introduction

Let $\mathfrak{S}_n$ denote the set of all permutations of $[n]$, where $[n] = \{1, 2, \ldots, n\}$. Let $\pi = \pi(1)\pi(2)\cdots\pi(n) \in \mathfrak{S}_n$. A *peak* in π is defined to be an index $i \in [n]$ such that $\pi(i-1) < \pi(i) > \pi(i+1)$, where we take $\pi(0) = \pi(n+1) = 0$. Let $\mathrm{pk}\,(\pi)$ denote the number of peaks in π. A *left peak* (resp. an *interior peak*) in π is an index $i \in [n-1]$ (resp. $i \in [2, n-1] = \{2, 3, \ldots, n-1\}$) such that $\pi(i-1) < \pi(i) > \pi(i+1)$, where we take $\pi(0) = 0$. Let $\mathrm{lpk}\,(\pi)$ (resp. $\mathrm{ipk}\,(\pi)$) denote the number of left peaks (resp. interior peaks) in π.

Combinatorial statistics on cycle notation have been extensively studied in recent years from several points of view (see, e.g., [7, 8, 9, 10, 14]). We say that π is a *circular permutation* if it has only one cycle. The notions of cyclic peak and cyclic valley are defined on circular permutations as follows. Let $A = \{x_1, x_2, \ldots, x_k\}$ be a finite set of positive integers with $k \geq 1$, and let $\mathcal{C}_A$ be the set of all circular permutations of A. We will write a circular permutation $w \in \mathcal{C}_A$ by using the *canonical presentation* $w = y_1 y_2 \cdots y_k$, where $y_1 = \min A$, $y_i = w^{i-1}(y_1)$ for $2 \leq i \leq k$ and $y_1 = w^k(y_1)$. The number $\mathrm{cpk}\,(w)$ of *cyclic peaks* (resp. $\mathrm{cval}\,(w)$ of *cyclic valleys*) of w is defined to be the number of indices $i \in [2, k-1]$ such that $y_{i-1} < y_i > y_{i+1}$ (resp. $y_{i-1} > y_i < y_{i+1}$).

The organization of this paper is as follows. In Section 2, we study the generating functions for the number of cyclic peaks or valleys, providing explicit expressions in both cases. In Section 3, a new combinatorial interpretation for the Pell numbers is obtained by considering the sign-balance of the cyclic valley statistic on the set of derangements. Our argument may be extended to yield a simple sign-balance formula for the entire symmetric group.

We now recall some prior results which we will need in our derivation of the generating function formulas for cyclic peaks and valleys. Define

$$\begin{aligned} P(\mathfrak{S}_n; q) &= \sum_{\pi \in \mathfrak{S}_n} q^{\mathrm{pk}\,(\pi)}, \\ IP(\mathfrak{S}_n; q) &= \sum_{\pi \in \mathfrak{S}_n} q^{\mathrm{ipk}\,(\pi)}, \\ LP(\mathfrak{S}_n; q) &= \sum_{\pi \in \mathfrak{S}_n} q^{\mathrm{lpk}\,(\pi)}. \end{aligned}$$

For convenience, set $P(\mathfrak{S}_0; q) = IP(\mathfrak{S}_0; q) = LP(\mathfrak{S}_0; q) = 1$. It is well known (see [4, ex. 3.3.46] and [11, A008303, A008971]) that

$$\begin{aligned} IP(\mathcal{S}; q, z) &= \sum_{n \geq 1} IP(\mathfrak{S}_n; q) \frac{z^n}{n!} = \frac{\sin(z\sqrt{q-1})}{\sqrt{q-1}\cos(z\sqrt{q-1}) - \sin(z\sqrt{q-1})}, \\ LP(\mathcal{S}; q, z) &= \sum_{n \geq 0} LP(\mathfrak{S}_n; q) \frac{z^n}{n!} = \frac{\sqrt{q-1}}{\sqrt{q-1}\cos(z\sqrt{q-1}) - \sin(z\sqrt{q-1})}. \end{aligned}$$

See also the related generating function formula found earlier by Entringer [3].

The complement map $\pi \mapsto \pi^c$, defined by $\pi^c(i) = n+1-\pi(i)$, shows that $\operatorname{pk}(\pi^c) = 1 + \operatorname{ipk}(\pi)$, upon considering cases as to whether there is one more, one less, or the same number of interior peaks as valleys in the permutation π. Thus

$$P(\mathfrak{S}_n; q) = qIP(\mathfrak{S}_n; q)$$

for $n \geq 1$. From the preceding, we conclude that

$$P(\mathcal{S}, q, z) = \sum_{n\geq 0} P(\mathfrak{S}_n; q)\frac{z^n}{n!} = 1 + \frac{q\sin(z\sqrt{q-1})}{\sqrt{q-1}\cos(z\sqrt{q-1}) - \sin(z\sqrt{q-1})}.$$

2. Generating functions

Let $\mathcal{C}_n = \mathcal{C}_{[n]}$. For each $w \in \mathcal{C}_n$, we define w' by the mapping

$$\Phi : w = 1y_2\cdots y_n \mapsto w' = (y_2-1)(y_3-1)\cdots(y_n-1).$$

One can verify the following result.

Lemma 2.1. *For $n \geq 2$, the mapping Φ is a bijection of $\mathcal{C}_n$ onto $\mathfrak{S}_{n-1}$ having the properties:*

$$\operatorname{cpk}(w) = \operatorname{lpk}(w'),\ \operatorname{cval}(w) = \operatorname{pk}(w') - 1.$$

Define

$$CP(\mathcal{C}_n; q) = \sum_{w\in\mathcal{C}_n} q^{\operatorname{cpk}(w)};\ CP(\mathcal{C}; q, z) = \sum_{n\geq 1} CP(\mathcal{C}_n; q)\frac{z^n}{n!};$$

$$CV(\mathcal{C}_n; q) = \sum_{w\in\mathcal{C}_n} q^{\operatorname{cval}(w)};\ CV(\mathcal{C}; q, z) = \sum_{n\geq 1} CV(\mathcal{C}_n; q)\frac{z^n}{n!}.$$

By Lemma 2.1, we have

$$\begin{aligned} CP(\mathcal{C}; q, z) &= z + \sum_{n\geq 2} CP(\mathcal{C}_n; q)\frac{z^n}{n!} \\ &= z + \sum_{n\geq 2} LP(\mathfrak{S}_{n-1}; q)\frac{z^n}{n!} \\ &= \sum_{n\geq 0} LP(\mathfrak{S}_n; q)\frac{z^{n+1}}{(n+1)!} \\ &= \int_0^z LP(\mathcal{S}; q, z)dz. \end{aligned}$$

Along the same lines, we obtain

$$qCV(\mathcal{C}; q, z) = qz + q\sum_{n\geq 2} CV(\mathcal{C}_n; q)\frac{z^n}{n!}$$

$$= qz + \sum_{n\geq 2} P(\mathfrak{S}_{n-1};q)\frac{z^n}{n!}$$
$$= \int_0^z P(\mathcal{S};q,z)dz + z(q-1).$$

With the aid of Maple, we can find expressions for the following two antiderivatives:

$$\int \frac{a}{a\cos(az)-\sin(az)}dz = \frac{1}{2\sqrt{1+a^2}}\ln\frac{\sqrt{1+a^2}+\cos(az)+a\sin(az)}{\sqrt{1+a^2}-\cos(az)-a\sin(az)} + C,$$
$$\int \frac{\sin(az)}{a\cos(az)-\sin(az)}dz = \frac{1}{1+a^2}\left(-z+\ln\frac{a}{a\cos(az)-\sin(az)}\right) + C.$$

Therefore, we get

$$CP(\mathcal{C};q,z) = \frac{1}{2\sqrt{q}}\ln\left[\left(\frac{\sqrt{q}-1}{\sqrt{q}+1}\right)\left(\frac{\sqrt{q}+\cos(z\sqrt{q-1})+\sqrt{q-1}\sin(z\sqrt{q-1})}{\sqrt{q}-\cos(z\sqrt{q-1})-\sqrt{q-1}\sin(z\sqrt{q-1})}\right)\right],$$
$$CV(\mathcal{C};q,z) = \frac{1}{q}\ln\frac{\sqrt{q-1}}{\sqrt{q-1}\cos(z\sqrt{q-1})-\sin(z\sqrt{q-1})} + z\left(1-\frac{1}{q}\right).$$

We write $\pi \in \mathfrak{S}_n$ as the product of disjoint cycles: $\pi = w_1w_2\cdots w_k$. When each of these cycles is expressed in canonical form, we define

$$\operatorname{cpk}(\pi) := \operatorname{cpk}(w_1) + \operatorname{cpk}(w_2) + \cdots + \operatorname{cpk}(w_k),$$
$$\operatorname{cval}(\pi) := \operatorname{cval}(w_1) + \operatorname{cpk}(w_2) + \cdots + \operatorname{cval}(w_k).$$

Both $\mu_1\colon \pi \mapsto q^{\operatorname{cpk}(\pi)}$ and $\mu_2\colon \pi \mapsto q^{\operatorname{cval}(\pi)}$ are multiplicative, in the sense that

$$\mu_i(\pi) = \mu_i(w_1)\mu_i(w_2)\cdots\mu_i(w_k),\ i = 1, 2.$$

Using the *exponential formula* [12, Corollary 5.5.5], we have

$$\sum_{n\geq 0}\frac{z^n}{n!}\sum_{\pi\in\mathfrak{S}_n}\mu_i(\pi) = \exp\left(\sum_{n\geq 1}\frac{z^n}{n!}\sum_{w\in\mathcal{C}_n}\mu_i(w)\right).$$

Define

$$CP(\mathfrak{S}_n;q) = \sum_{w\in\mathfrak{S}_n} q^{\operatorname{cpk}(w)};$$
$$CV(\mathfrak{S}_n;q) = \sum_{w\in\mathfrak{S}_n} q^{\operatorname{cval}(w)}.$$

Accordingly,

$$CP(\mathcal{S};q,z) = \sum_{n\geq 0} CP(\mathfrak{S}_n;q)\frac{z^n}{n!} = \exp(CP(\mathcal{C};q,z)),$$

$$CV(\mathcal{S};q,z)=\sum_{n\geq 0} CV(\mathfrak{S}_n;q)\frac{z^n}{n!}=\exp(CV(\mathcal{C};q,z)).$$

Combining the prior observations yields the following result.

Theorem 2.2. *We have*

$$\begin{aligned} &CP(\mathcal{S};q,z)\\ &=\left[\left(\frac{\sqrt{q}-1}{\sqrt{q}+1}\right)\left(\frac{\sqrt{q}+\cos(z\sqrt{q-1})+\sqrt{q-1}\sin(z\sqrt{q-1})}{\sqrt{q}-\cos(z\sqrt{q-1})-\sqrt{q-1}\sin(z\sqrt{q-1})}\right)\right]^{\frac{1}{2\sqrt{q}}}, \qquad (2.1)\\ &CV(\mathcal{S};q,z)=e^{z(1-1/q)}\left(\frac{\sqrt{q-1}}{\sqrt{q-1}\cos(z\sqrt{q-1})-\sin(z\sqrt{q-1})}\right)^{\frac{1}{q}}. \qquad (2.2)\end{aligned}$$

We say that $\pi\in\mathfrak{S}_n$ changes direction at position i if either $\pi(i-1)<\pi(i)>\pi(i+1)$ or $\pi(i-1)>\pi(i)<\pi(i+1)$, where $i\in[2,n-1]$. We say that π has k *alternating runs* if there are $k-1$ indices i such that π changes direction at these positions. Let $R(n,k)$ denote the number of permutations in $\mathfrak{S}_n$ with k alternating runs. It is well known that the numbers $R(n,k)$ satisfy the recurrence relation

$$R(n,k)=kR(n-1,k)+2R(n-1,k-1)+(n-k)R(n-1,k-2)$$

for $n,k\geqslant 1$, where $R(1,0)=1$ and $R(1,k)=0$ for $k\geqslant 1$ (see [11, A059427]). Let $R_n(q)=\sum_{k\geqslant 1}R(n,k)q^k$. There is an extensive literature devoted to the polynomials $R_n(q)$. The reader is referred to [5, 6, 13] for recent progress on this subject.

In [1], Carlitz proved that

$$H(q,z)=\sum_{n=0}^{\infty}\frac{z^n}{n!}\sum_{k=0}^{n}R(n+1,k)q^{n-k}=\left(\frac{1-q}{1+q}\right)\left(\frac{\sqrt{1-q^2}+\sin(z\sqrt{1-q^2})}{q-\cos(z\sqrt{1-q^2})}\right)^2.$$

Consider

$$R(q,z)=\sum_{n\geqslant 0}R_{n+1}(q)\frac{z^n}{n!}.$$

It is clear that $R(q,z)=H(\frac{1}{q},qz)$. Hence

$$R(q,z)=\left(\frac{q-1}{q+1}\right)\left(\frac{\sqrt{q^2-1}+q\sin(z\sqrt{q^2-1})}{1-q\cos(z\sqrt{q^2-1})}\right)^2.$$

There is an equivalent expression for $R(q,z)$ (see [2, eq. (20)]):

$$R(q,z)=\left(\frac{q-1}{q+1}\right)\left(\frac{q+\cos(z\sqrt{q^2-1})+\sqrt{q^2-1}\sin(z\sqrt{q^2-1})}{q-\cos(z\sqrt{q^2-1})-\sqrt{q^2-1}\sin(z\sqrt{q^2-1})}\right). \qquad (2.3)$$

By Theorem 2.2, we can now conclude the following result.

Corollary 2.3. *We have*

$$CP(\mathcal{S};q,z) = R(\sqrt{q},z)^{\frac{1}{2\sqrt{q}}},$$
$$CV(\mathcal{S};q,z) = e^{z(1-1/q)}LP(\mathcal{S};q,z)^{\frac{1}{q}}.$$

Corollary 2.4. *For $n \geqslant 2$, the total number of cyclic peaks in all the members of $\mathfrak{S}_n$ is given by*

$$\frac{n!(4n+1-6H_n)}{12},$$

and the total number of cyclic valleys in all the members of $\mathfrak{S}_n$ is given by

$$\frac{n!(2n+5-6H_n)}{6},$$

where $H_n = \sum_{j=1}^{n} \frac{1}{j}$ is the n-th harmonic number.

Proof. It follows from Theorem 2.2 that

$$\begin{aligned}
\frac{d}{dq}CP(\mathcal{S};q,z)\mid_{q=1} &= \frac{(z^3-3z^2+6z+6(1-z)\ln(1-z))}{12(1-z)^2} \\
&= \frac{z^3-3z^2+6z}{12(1-z)^2} + \frac{\ln(1-z)}{2(1-z)} \\
&= \sum_{n\geqslant 2}(4n+1-6H_n)\frac{z^n}{12}, \\
\frac{d}{dq}CV(\mathcal{S};q,z)\mid_{q=1} &= \frac{(-z^3-3z^2+6z+6(1-z)\ln(1-z))}{6(1-z)^2} \\
&= \frac{-z^3-3z^2+6z}{6(1-z)^2} + \frac{\ln(1-z)}{1-z} \\
&= \sum_{n\geqslant 2}(2n+5-6H_n)\frac{z^n}{6},
\end{aligned}$$

as required. $\square$

Note that
$$CP(\mathfrak{S}_n;0) = \#\{\pi \in \mathfrak{S}_n : \mathrm{cpk}\,(\pi) = 0\}.$$

Consider a permutation

$$\pi = (\pi(i_1),\ldots)(\pi(i_2),\ldots)\cdots(\pi(i_j),\ldots)$$

counted by $CP(\mathfrak{S}_n;0)$. Replacing the parentheses enclosing cycles with brackets, we get a partition of $[n]$ with j blocks. Therefore, we obtain

$$CP(\mathfrak{S}_n;0) = B_n, \tag{2.4}$$

where B_n is the nth *Bell number* [11, A000110], i.e., the number of partitions of $[n]$ into non-empty blocks.

We present now a generating function proof of (2.4). Note that

$$\sum_{n\geqslant 0} CP(\mathfrak{S}_n;0)\frac{z^n}{n!}$$
$$= \lim_{q\to 0}\left[\left(\frac{\sqrt{q}-1}{\sqrt{q}+1}\right)\left(\frac{\sqrt{q}+\cos(z\sqrt{q-1})+\sqrt{q-1}\sin(z\sqrt{q-1})}{\sqrt{q}-\cos(z\sqrt{q-1})-\sqrt{q-1}\sin(z\sqrt{q-1})}\right)\right]^{\frac{1}{2\sqrt{q}}}.$$

Denote the limit on the right by L. It is easy to see that L is of the indeterminate form 1^∞. So, by l'Hôpital's rule, we have

$$\ln L = \lim_{q\to 0}\frac{1}{2\sqrt{q}}\ln\left[\left(\frac{\sqrt{q}-1}{\sqrt{q}+1}\right)\left(\frac{\sqrt{q}+\cos(z\sqrt{q-1})+\sqrt{q-1}\sin(z\sqrt{q-1})}{\sqrt{q}-\cos(z\sqrt{q-1})-\sqrt{q-1}\sin(z\sqrt{q-1})}\right)\right]$$
$$= \cosh z+\sinh z-1 = e^z-1.$$

Consequently,

$$\sum_{n\geqslant 0} CP(\mathfrak{S}_n;0)\frac{z^n}{n!} = e^{e^z-1},$$

the right-hand side being the exponential generating function of B_n, thus proving (2.4).

3. A new combinatorial interpretation for the Pell numbers

Recall that the *Pell numbers* P_n are defined by the recurrence relation $P_n = 2P_{n-1}+P_{n-2}$ for $n \geqslant 2$, with initial values $P_0 = 0$ and $P_1 = 1$ (see [11, A000129]). The Pell numbers are also given, equivalently, by the Binet formula

$$P_n = \frac{(1+\sqrt{2})^n-(1-\sqrt{2})^n}{2\sqrt{2}}, \quad n \geqslant 0,$$

which implies

$$P_n = \sum_{0\leqslant r\leqslant \lfloor\frac{n-1}{2}\rfloor}\binom{n}{2r+1}2^r, \quad n \geqslant 1.$$

By a *fixed point* of a permutation π, we mean an $i \in [n]$ such that $\pi(i) = i$. A fixed point free permutation is called a *derangement*. Let $\mathcal{D}_n$ denote the set of derangements of $[n]$.

Our next result reveals a somewhat unexpected connection between derangements and Pell numbers.

Theorem 3.1. *For $n \geqslant 1$, we have*

$$\sum_{\sigma\in\mathcal{D}_n}(-1)^{\mathrm{cval}\,(\sigma)} = P_{n-1}. \tag{3.1}$$

Combinatorial proof of Theorem 3.1

Let us assume $n \geqslant 3$ and let $\mathcal{D}_n^+$ and $\mathcal{D}_n^-$ denote the subsets of $\mathcal{D}_n$ whose members contain an even or odd number of cyclic valleys, respectively. To show (3.1), we will define a cval-parity changing involution of $\mathcal{D}_n$ whose survivors belong to $\mathcal{D}_n^+$ and have cardinality P_{n-1}. We will say that a permutation is in *standard form* if the smallest element is first within each cycle, with cycles arranged in increasing order of smallest elements. Let $\mathcal{D}_n^*$ consist of those members $\pi = C_1C_2\cdots C_r$ of $\mathcal{D}_n$ in standard form whose cycles C_i satisfy the following two conditions for $1 \leqslant i \leqslant r$:

(a) C_i consists of a set of consecutive integers, and

(b) C_i is either increasing or contains exactly one cyclic peak but no cyclic valleys.

Note that $\mathcal{D}_n^* \subseteq \mathcal{D}_n^+$ and in the lemma that follows, it is shown that $|\mathcal{D}_n^*| = P_{n-1}$.

We now proceed to define an involution of $\mathcal{D}_n - \mathcal{D}_n^*$. Given $\pi = C_1C_2\cdots C_r \in \mathcal{D}_n - \mathcal{D}_n^*$ in standard form, let j_0 denote the smallest index j such that cycle C_j violates condition (a) or (b) (possibly both). Let us assume for now that $j_0 = 1$. Then let i_0 be the smallest index i such that either

(I) i is the middle letter of some cyclic valley of C_1, or

(II) i fails to belong to C_1 with at least one member of $[i+1,n]$ belonging to C_1.

Observe that if (I) occurs, then C_1 may be decomposed as

$$C_1 = 1\alpha\gamma\delta\beta,$$

where α is a subset of $[2, i_0-1]$ and is increasing, β is a subset of $[2, i_0-1]$ and is decreasing, the union of α and β is $[2, i_0-1]$ with α or β possibly empty, γ consists of letters in $[i_0+1, n]$, and δ starts with the letter i_0. Note in this case that i_0 being the middle letter of some cyclic valley implies γ is non-empty and δ has length at least two. Next observe that if (II) occurs, then $C_1 = 1\alpha\rho\beta$, where α and β are as before and ρ is non-empty. Note that the second cycle C_2 must start with i_0 in this case.

We define an involution by splitting the cycle C_1 into two cycles $L_1 = 1\alpha\gamma\beta$, $L_2 = \delta$ if (I) occurs, and by merging cycles C_1 and C_2 such that the letters of C_2 go between ρ and β if (II) occurs. Note that the former operation removes exactly one cyclic valley (namely, the one involving i_0) since all of the letters of γ are greater than those of β with β decreasing, while the latter operation is seen to add exactly one cyclic valley. Furthermore, the standard ordering of the cycles is preserved by the former operation, by the minimality of i_0.

For $j_0 \geqslant 1$ in general, perform the operations defined above using the cycle C_{j_0} and its successor, treating the letters contained therein as those in $[\ell]$ for some ℓ and leaving the cycles $C_1, C_2, \ldots, C_{j_0-1}$ undisturbed. Let π' denote the resulting derangement. Then it may be verified that the mapping $\pi \mapsto \pi'$ is an involution of $\mathcal{D}_n - \mathcal{D}_n^*$ such that π and π' have opposite cval parity for all π.

For example, if $n = 20$ and $\pi = \mathcal{D}_{20} - \mathcal{D}^*_{20}$ is given by

$$\pi = (1,3,5,4,2),\ (6,7),\ (8,10,11,18,13,15,9),\ (12,17,14,20),\ (16,19),$$

then $j_0 = 3$ and

$$\pi' = (1,3,5,4,2),\ (6,7),\ (8,10,11,18,13,15,12,17,14,20,9),\ (16,19).$$

Lemma 3.2. *If $n \geqslant 1$, then $|\mathcal{D}^*_n| = P_{n-1}$.*

Proof. Recall that P_m counts the tilings of length $m-1$ consisting of squares and dominos such that squares may be colored black or white (called *Pell tilings*). To complete the proof, we define a bijection f between $\mathcal{D}^*_n$ and the set of Pell tilings of length $n-2$, where $n \geqslant 3$. Suppose $\pi = C_1C_2\cdots C_r \in \mathcal{D}^*_n$. If $1 \leqslant i < r$, then we convert the cycle C_i into a Pell subtiling as follows. First assume $i = 1$ and let t denote the largest letter of cycle C_1. If $j \in [2, t-1]$ and occurs to the left (resp. right) of t in C_1, then let the $(j-1)$-st piece of $f(\pi)$ be a white (resp. black) square. To the resulting sequence of $t-2$ squares, we append a domino. Thus C_1 has been converted to a Pell subtiling of the same length ending in a domino. Repeat for the cycles $C_2, C_3, \ldots, C_{r-1}$, at each step appending the subtiling that results to the current tiling. For cycle C_r, we perform the same procedure, but this time no domino is added at the end. Let $f(\pi)$ denote the resulting Pell tiling of length $n-2$. It may be verified that the mapping f is a bijection. Note that $f(\pi)$ ends in a domino if and only if cycle C_r has length two and that the number of dominos of $f(\pi)$ is one less than the number of cycles of π. □

In the remainder of this section, we present a comparable sign-balance result for $\mathfrak{S}_n$. Let $i = \sqrt{-1}$. Note that $\cosh(x) = \cos(ix)$ and $\sinh(x) = -i\sin(ix)$. Setting $q = -1$ in (2.2), we obtain

$$\begin{aligned}\sum_{n\geqslant 0}\frac{z^n}{n!}\sum_{\pi\in\mathfrak{S}_n}(-1)^{\mathrm{cval}\,(\pi)} &= \sum_{n\geqslant 0} CV(\mathfrak{S}_n;-1)\frac{z^n}{n!}\\ &= \frac{1}{\sqrt{2}}e^{2z}(\sqrt{2}\cosh(\sqrt{2}z) - \sinh(\sqrt{2}z)).\end{aligned}$$

Equating coefficients yields the following result.

Theorem 3.3. *For $n \geq 1$, we have*

$$\begin{aligned}\sum_{\pi\in\mathfrak{S}_n}(-1)^{\mathrm{cval}\,(\pi)} &= \frac{1}{2}((2+\sqrt{2})^{n-1} + (2-\sqrt{2})^{n-1})\\ &= \sum_{0\leqslant r\leqslant\lfloor\frac{n-1}{2}\rfloor}\binom{n-1}{2r}2^{n-1-r}. \qquad (3.2)\end{aligned}$$

Combinatorial proof of Theorem 3.3

Let $\mathfrak{S}_n^+$ and $\mathfrak{S}_n^-$ denote the subsets of $\mathfrak{S}_n$ whose members contain an even or odd number of cyclic valleys. To show (3.2), we seek an involution of $\mathfrak{S}_n$ which changes the parity of cval. Indeed, we define a certain extension of the mapping used in the proof of (3.1). Let $\mathfrak{S}_n^*$ consist of those members $\pi = C_1C_2\cdots C_r$ of $\mathfrak{S}_n$ in standard form all of whose cycles satisfy the following two properties:

(a) C_i is either a singleton or if it is not a singleton, it comprises a set of consecutive integers when taken together with all singleton cycles between it and the next non-singleton cycle (if there is one), and

(b) C_i is either increasing or contains exactly one cyclic peak but no cyclic valleys.

Note that $\mathfrak{S}_n^* \subseteq \mathfrak{S}_n^+$ and below it is shown that $|\mathfrak{S}_n^*| = \sum_r \binom{n-1}{2r} 2^{n-1-r}$.

We now define an involution of $\mathfrak{S}_n - \mathfrak{S}_n^*$. Given $\pi = C_1C_2\cdots C_r \in \mathfrak{S}_n - \mathfrak{S}_n^*$ in standard form, let j_0 denote the smallest index j such that cycle C_j violates condition (a) or (b) (possibly both). Let us assume for now that $j_0 = 1$, the general case being done in a similar manner as will be apparent. Let i_0 denote the smallest index i satisfying conditions (I) or (II) in the proof above for (3.1), where in (II) we must now add the assumption that i belongs to a non-singleton cycle. The involution $\pi \mapsto \pi'$ is then defined in an analogous manner as it was in the proof of (3.1) above except now, in the merging operation, a non-singleton cycle is moved to the first non-singleton cycle which precedes it (with possibly some singletons separating the two).

For example, if $n = 20$ and $\pi = \mathfrak{S}_{20} - \mathfrak{S}_{20}^*$ is given by

$$\pi = (1,3,5,4),\ (2),\ (6,7),\ (8,18,13,15,11),\ (9),\ (10),\ (12,17,14),\ (16,20),(19),$$

then $j_0 = 4$ and

$$\pi' = (1,3,5,4),\ (2),\ (6,7),\ (8,18,13,15,12,17,14,11),\ (9),\ (10),(16,20),\ (19).$$

We now seek the cardinality of $\mathfrak{S}_n^*$. To do so, we will first define a bijection between $\mathfrak{S}_n^*$ and the set $\mathcal{A}_{n-1}$ consisting of sequences $s_1s_2\cdots s_{n-1}$ in $[4]$ such that $s_1 = 1$ or 2, with the strings 13 and 24 forbidden. To define it, first observe that members of $\mathfrak{S}_n^*$, $n \geqslant 1$, may be formed recursively from members of $\mathfrak{S}_{n-1}^*$ (on the alphabet $[2,n]$) by performing one of the following operations:

(i) adding 1 as (1),

(ii) either replacing the 1-cycle (2), if it occurs, with $(1,2)$ or replacing the cycle $(2c_1c_2\cdots)$ with the two cycles $(1c_1c_2\cdots)$, (2),

(iii) replacing the cycle $(2c_1c_2\cdots c_s)$, if it occurs where $s \geqslant 1$, with $(12c_1c_2\cdots c_s)$, or

(iv) replacing the cycle $(2c_1c_2\cdots c_s)$, if it occurs where $s \geqslant 1$, with $(1c_1c_2\cdots c_s2)$.

Note that (iii) or (iv) cannot be performed on a member of $\mathfrak{S}_{n-1}^*$ if 2 occurs as a 1-cycle, that is, if (i) has been performed in the previous step. Let $\mathcal{B}_{n-1}$ denote the set of sequences in [4] of length $n-1$ having first letter 1 or 2, with the strings 13 and 14 forbidden. Thus, adding 1 to a member of $\mathfrak{S}_{n-1}^*$ as described to obtain a member of $\mathfrak{S}_n^*$ may be viewed as writing the final letter of some member of $\mathcal{B}_{n-1}$. From this, we see that members of $\mathcal{B}_{n-1}$ serve as encodings for creating members of $\mathfrak{S}_n^*$, starting with the letter n and working downward. For example, the sequence $w = 21123412243 \in \mathcal{B}_{11}$ would correspond to $\pi = (1,2,5,3),(4),(6,8,9,7),(10),(11,12) \in \mathfrak{S}_{12}^*$. Note that replacing any occurrence of the string 24 within a member of $\mathcal{B}_{n-1}$ with the string 14 is seen to define a bijection with the set $\mathcal{A}_{n-1}$.

Taking the composition of the maps described from $\mathfrak{S}_n^*$ to $\mathcal{B}_{n-1}$ and from $\mathcal{B}_{n-1}$ to $\mathcal{A}_{n-1}$ yields the desired bijection from $\mathfrak{S}_n^*$ to $\mathcal{A}_{n-1}$.

The following lemma will imply $|\mathfrak{S}_n^*|$ is given by the right-hand side of (3.2) and complete the proof.

Lemma 3.4. *If $m \geqslant 1$, then $|\mathcal{A}_m| = \sum_{r=0}^{\lfloor \frac{m}{2} \rfloor} \binom{m}{2r} 2^{m-r}$.*

Proof. The $r=0$ term of the sum clearly counts all of the binary members of $\mathcal{A}_m$, so we need to show that the cardinality of all $\pi \in \mathcal{A}_m$ containing at least one 3 or 4 is given by $\sum_{r=1}^{\lfloor \frac{m}{2} \rfloor} \binom{m}{2r} 2^{m-r}$.

Note that π may be decomposed as

$$\pi = S_1 S_2 \cdots S_\ell, \qquad \ell \geqslant 2, \tag{3.3}$$

where the odd-indexed S_i are maximal substrings containing only letters in $\{1,2\}$ and the even-indexed S_i are maximal substrings containing only letters in $\{3,4\}$. If $\ell = 2r$ is even in (3.3), then choose a sequence of length $2r-1$ in $[2,m]$, which we will denote by $i_2 < i_3 < \cdots < i_{2r}$ for convenience with $i_1 = 1$. We wish to create members $\pi = \pi_1\pi_2\cdots\pi_m \in \mathcal{A}_m$ such that the initial letter of the block S_j is in position i_j for $1 \leqslant j \leqslant 2r$. To do so, we first fill in the positions of π whose indices correspond to elements of $[i_{2j-1}, i_{2j}-1]$ with letters from $\{1,2\}$ for each $j \in [r]$. Next, we fill the positions of π in $[i_{2j}, i_{2j+1}-1]$ for $j \in [r-1]$, along with the positions in $[i_{2r}, n]$, with letters from $\{3,4\}$. Note that the letters in positions i_{2j}, $j \in [r]$, are determined by the choice of last letter for the block S_{2j-1}, since the 13 and 24 strings are forbidden. Thus, there are $\binom{m-1}{2r-1} 2^{m-r}$ members of $\mathcal{A}_m$ such that $\ell = 2r$ in (3.3). By similar reasoning, there are $\binom{m-1}{2r} 2^{m-r}$ members of $\mathcal{A}_m$ such that $\ell = 2r+1$ in (3.3). Combining these two cases, it follows that there are $\left(\binom{m-1}{2r-1} + \binom{m-1}{2r}\right) 2^{m-r} = \binom{m}{2r} 2^{m-r}$ members of $\mathcal{A}_m$ for which $\ell = 2r$ or $\ell = 2r+1$ in (3.3). Summing over $r \geqslant 1$ gives the cardinality of all members of $\mathcal{A}_m$ containing at least one 3 or 4 and completes the proof. $\square$

References

[1] Carlitz, L., Enumeration of permutations by sequences, *Fibonacci Quart.* 16(3) (1978), 259–268.

[2] Chow, C.-O. and Ma, S.-M., Counting signed permutations by their alternating runs, *Discrete Math.* 323 (2014), 49–57.

[3] Entringer, R., Enumeration of permutations of $(1,\ldots,n)$ by number of maxima, *Duke Math. J.* 36 (1969), 575–579.

[4] Goulden, I.P. and Jackson, D.M., Combinatorial Enumeration, John Wiley and Sons, N.Y., 1983.

[5] Ma, S.-M., An explicit formula for the number of permutations with a given number of alternating runs, *J. Combin. Theory Ser. A* 119 (2012), 1660–1664.

[6] Ma, S.-M., Enumeration of permutations by number of alternating runs, *Discrete Math.* 313 (2013), 1816–1822.

[7] Mansour, T. and Shattuck, M., Pattern avoidance in flattened permutations, *Pure Math. Appl.* 22:1 (2011), 75–86.

[8] Mansour, T., Shattuck, M. and Wang, D., Counting subwords in flattened permutations, *J. Comb.* 4:3 (2013), 327–356.

[9] Parviainen, R., Permutations, cycles, and the pattern 2-13, *Electron. J. Combin.* 13 (2006), #R111.

[10] Shin, H. and Zeng, J., The symmetric and unimodal expansion of Eulerian polynomials via continued fractions, *European J. Combin.* 33 (2012), 111–127.

[11] Sloane, N.J.A., The On-Line Encyclopedia of Integer Sequences, published electronically at http://oeis.org, 2010.

[12] Stanley, R.P., Enumerative Combinatorics, Vol. 2, Cambridge Univ. Press, Cambridge, UK, 1999.

[13] Stanley, R.P., Longest alternating subsequences of permutations, *Michigan Math. J.* 57 (2008), 675–687.

[14] Zhao, A.F.Y., Excedance numbers for the permutations of type B, *Electron. J. Combin.* 20(2) (2013), #P28.

Annales Mathematicae et Informaticae
43 (2014) pp. 55–65
http://ami.ektf.hu

On the existence of the generalized Gauss composition of means*

Peter Csiba[a], Ferdinánd Filip[a], Attila Komzsík[b], János T. Tóth[a]

[a]Department of Mathematics and Informatics, J. Selye University, Komárno, Slovakia
csibap@ujs.sk, filipf@ujs.sk, tothj@ujs.sk

[b]Institute for Teacher Training, Constantine the Philosopher University in Nitra, Slovakia
akomzsik@ukf.sk

Submitted April 1, 2014 — Accepted July 16, 2014

Abstract

The paper deals with the generalized Gauss composition of arbitrary means. We give sufficient conditions for the existence of this generalized Gauss composition. Finally, we show that these conditions cannot be improved or changed.

Keywords: means, power means, Gauss composition of means, Archimedean composition of means

MSC: 26E60

1. Introduction

In this part we recall some basic definitions. Denote by $\mathbb{N}$ and $\mathbb{R}^+$ the set of all positive integers and positive real numbers, respectively.

Let $I \subset \mathbb{R}$ be a non-empty open interval. A function $M : I^2 \to I$ is called a *mean* on I if for all $x, y \in I$

$$\min\{x, y\} \leq M(x, y) \leq \max\{x, y\}.$$

*Supported by VEGA Grant no. 1/1022/12.

It is obvious that $M(x,x) = x$ for all $x \in I$.

The mean $M : I^2 \to I$ is called *symmetric* if

$$M(x,y) = M(y,x)$$

for all $x, y \in I$.

The mean $M : I^2 \to I$ is called a *strict* mean on I if it is continuous on I^2 and for all $x, y \in I$ with $x \neq y$

$$\min\{x,y\} < M(x,y) < \max\{x,y\}.$$

The mean $M : (\mathbb{R}^+)^2 \to \mathbb{R}^+$ is called *homogeneous* if

$$M(zx, zy) = zM(x,y),$$

for all $x, y, z \in \mathbb{R}^+$.

Classical examples for two-variable strict means on $\mathbb{R}^+$ are:

– The arithmetic, the geometric and the harmonic mean

$$A(x,y) := \frac{x+y}{2}, \quad G(x,y) := \sqrt{xy}, \quad H(x,y) := \frac{2xy}{x+y}\,.$$

– The power means, also called Hölder means, of exponent p

$$M_p(x,y) := \begin{cases} \left(\frac{x^p+y^p}{2}\right)^{\frac{1}{p}} & \text{if} \quad p \neq 0 \\ \lim\limits_{p\to 0} \left(\frac{x^p+y^p}{2}\right)^{\frac{1}{p}} & \text{if} \quad p = 0\,. \end{cases}$$

The case $p = 1$ corresponds to the arithmetic mean, $p = 0$ to the geometric mean, and $p = -1$ to the harmonic mean. It is well known that

$$\lim_{p\to -\infty} M_p(x,y) = \lim_{p\to -\infty} \left(\frac{x^p+y^p}{2}\right)^{\frac{1}{p}} = \min\{x,y\}$$

and

$$\lim_{p\to \infty} M_p(x,y) = \lim_{p\to \infty} \left(\frac{x^p+y^p}{2}\right)^{\frac{1}{p}} = \max\{x,y\}\,.$$

These means are called the minimum and maximum mean, respectively.

– The logarithmic mean

$$L(x,y) := \begin{cases} \frac{y-x}{\ln y - \ln x} & \text{if} \quad x \neq y \\ x & \text{if} \quad x = y\,. \end{cases}$$

This area has been studied by many mathematicians. For this paper we were inspired by [2, 3, 4, 5, 7].

Let $M, N : I^2 \to I$ be means on I and $a, b \in I$. Consider the sequences (a_n) and (b_n) defined by the Gauss iteration in the following way:

$$\begin{aligned} a_1 &:= a, & b_1 &:= b, \\ a_{n+1} &:= M(a_n, b_n), & b_{n+1} &:= N(a_n, b_n) \quad (n \in \mathbb{N}). \end{aligned} \tag{1.1}$$

If the limits $\lim_{n\to\infty} a_n$, $\lim_{n\to\infty} b_n$ exist and

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n,$$

then this common limit is called the *Gauss composition* of the means M and N for the numbers a and b, and is denoted by $M \otimes N(a, b)$. We say that the means M and N are composable in the sense of Gauss (or G-composable). For some applications of Gauss composition see for example [7] or [8].

We can similarly define the Archimedean composition mean of the means M and N (see [11, pp. 77–78]): consider the sequences (a_n) and (b_n) defined by

$$\begin{aligned} a_1 &:= a, & b_1 &:= b, \\ a_{n+1} &:= M(a_n, b_n), & b_{n+1} &:= N(a_{n+1}, b_n) \quad (n \in \mathbb{N}). \end{aligned} \tag{1.2}$$

If the limits $\lim_{n\to\infty} a_n$, $\lim_{n\to\infty} b_n$ exist and

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n,$$

then this common limit is called the *Archimedean composition* mean of the means M and N for the numbers a and b, and is denoted by $M \boxtimes N(a, b)$. We say that the means M and N are composable in the sense of Archimedes (or A-composable).

There is a known relation between Gauss composition of means and Archimedean composition mean of means (see in [11], p. 79):

$$M \boxtimes N(a, b) = M \otimes N(M, \Pi_2)(a, b), \tag{1.3}$$

where $\Pi_2(a, b) = b$ and $N(M, \Pi_2)(a, b) = N(M(a, b), \Pi_2(a, b))$.

It is known (see [1], [6]) that if M, N are strict means on I, then $M \otimes N(a, b)$ exists for every $a, b \in I$.

In this paper we generalise this result. We will show the following: if the means M_1, M_2 (not necessarily continuous) may be bounded "from one direction" by strict means then their Gauss composition exists. Finally, a counter-example will show that the continuity of the bounding mean cannot be omitted.

2. Results

Theorem 2.1. *Let M, N be means on I and let L_1, L_2 be continuous means on I such that for each $x, y \in I$ with $x \neq y$*

$$L_1(x, y) > \min\{x, y\} \tag{2.1}$$

and

$$L_2(x, y) < \max\{x, y\} . \tag{2.2}$$

If any of the following conditions is fulfilled,

a) *for each pair of real numbers* $x, y \in I$*:* $L_1(x, y) \le M(x, y)$ *and* $L_1(x, y) \le N(x, y)$*,*

b) *for each pair of real numbers* $x, y \in I$*:* $L_2(x, y) \ge M(x, y)$ *and* $L_2(x, y) \ge N(x, y)$*,*

c) *for each pair of real numbers* $x, y \in I$*:* $L_1(x, y) \le M(x, y) \le L_2(x, y)$*,*

then the means M *and* N *are G-composable, i.e. the mean* $M \otimes N(a, b)$ *exists.*

Proof. Let us define the sequences (a_n) and (b_n) by (1.1) and the sequences (c_n) and (d_n) by

$$c_n = \min\{a_n, b_n\} \quad \text{and} \quad d_n = \max\{a_n, b_n\} .$$

Then, evidently, the limits

$$\lim_{n\to\infty} c_n = c \quad \text{and} \quad \lim_{n\to\infty} d_n = d$$

exist, and $c \le d$. It is sufficient to prove that $c = d$.

All three cases will be proved by contradiction. Hence assume

$$c < d . \tag{2.3}$$

a) From the definitions of (c_n) and (d_n) it follows that at least one of the following two statements is true.

I. The sequence (c_n) has a subsequence $(c_{n_k})_{k=1}^{\infty}$ such that $c_{n_k} = a_{n_k}$ for each $k \in \mathbb{N}$.

II. The sequence (c_n) has a subsequence $(c_{n_k})_{k=1}^{\infty}$ such that $c_{n_k} = b_{n_k}$ for each $k \in \mathbb{N}$.

In case I., from (2.1), (2.3) and from the continuity of L_1, we get the inequality

$$c < L_1(c, d) = \lim_{n\to\infty} L_1(c_n, d_n) = \lim_{k\to\infty} L_1(c_{n_k}, d_{n_k}) = \lim_{k\to\infty} L_1(a_{n_k}, b_{n_k}). \tag{2.4}$$

On the other hand, from condition a) and the definition of the sequence (c_n), we get the inequality

$$L_1(a_{n_k}, b_{n_k}) \le \min\{M(a_{n_k}, b_{n_k}), N(a_{n_k}, b_{n_k})\} = \min\{a_{n_k+1}, b_{n_k+1}\} = c_{n_k+1}.$$

Substituting this back to (2.4) we get the inequality

$$c < \lim_{k\to\infty} c_{n_k+1} = c$$

and this is a contradiction.

In case II., we similarly get the inequality

$$c < L_1(d,c) = \lim_{n\to\infty} L_1(d_n, c_n) = \lim_{k\to\infty} L_1(d_{n_k}, c_{n_k}) = \lim_{k\to\infty} L_1(a_{n_k}, b_{n_k}) \tag{2.5}$$

and

$$L_1(a_{n_k}, b_{n_k}) \le \min\{M(a_{n_k}, b_{n_k}), N(a_{n_k}, b_{n_k})\} = \min\{a_{n_k+1}, b_{n_k+1}\} = c_{n_k+1}\,.$$

Substituting this back to the (2.5) we get the contradiction

$$c < \lim_{k\to\infty} c_{n_k+1} = c\,.$$

b) The proof is analogous to the proof of case a).

c) From the definitions of (c_n) and (d_n) it follows that at least one of the following three statements is true: the sequence (c_n) has a subsequence $(c_{n_k})_{k=1}^{\infty}$, where for each $k \in \mathbb{N}$,

I.

$$c_{n_k} = a_{n_k} \quad \text{and} \quad c_{n_k+1} = a_{n_k+1}$$

II.

$$c_{n_k} = a_{n_k} \quad \text{and} \quad c_{n_k+1} = b_{n_k+1}$$

III.

$$c_{n_k} = b_{n_k} \quad \text{and} \quad c_{n_k+1} = b_{n_k+1}.$$

In case I., from (2.1), (2.3), continuity of L_1 and the condition c), we obtain

$$\begin{aligned} c < L_1(c,d) &= \lim_{n\to\infty} L_1(c_n, d_n) = \lim_{k\to\infty} L_1(c_{n_k}, d_{n_k}) = \\ &= \lim_{k\to\infty} L_1(a_{n_k}, b_{n_k}) \le \lim_{k\to\infty} M(a_{n_k}, b_{n_k}) = \\ &= \lim_{k\to\infty} a_{n_k+1} = \lim_{k\to\infty} c_{n_k+1} = c\,, \end{aligned}$$

which is a contradiction.

In case II., from (2.2), (2.3), continuity of L_2 and the condition c), we obtain

$$\begin{aligned} d > L_2(c,d) &= \lim_{n\to\infty} L_2(c_n, d_n) = \lim_{k\to\infty} L_2(c_{n_k}, d_{n_k}) = \\ &= \lim_{k\to\infty} L_2(a_{n_k}, b_{n_k}) \ge \lim_{k\to\infty} M(a_{n_k}, b_{n_k}) = \\ &= \lim_{k\to\infty} a_{n_k+1} = \lim_{k\to\infty} d_{n_k+1} = d\,, \end{aligned}$$

which is a contradiction, too.

Finally, in case III., from (2.2), (2.3) and the continuity of L_2 and the condition c), we have

$$
\begin{aligned}
d > L_2(d,c) &= \lim_{n\to\infty} L_2(d_n, c_n) = \lim_{k\to\infty} L_2(d_{n_k}, c_{n_k}) = \\
&= \lim_{k\to\infty} L_2(a_{n_k}, b_{n_k}) \geq \lim_{k\to\infty} M(a_{n_k}, b_{n_k}) = \\
&= \lim_{k\to\infty} a_{n_k+1} = \lim_{k\to\infty} d_{n_k+1} = d\,,
\end{aligned}
$$

a contradiction. □

From the relation (1.3) we obtain a similar result for the Archimedean composition.

Corollary 2.2. *If the conditions of Theorem 2.1 hold, then the means M and N are A-composable.*

As a consequence of Theorem 2.1 we immediately get the following result (see also [6] and [10]).

Corollary 2.3. *Let M be a strict mean and N an arbitrary mean defined on the interval I. Then the means M and N are G-composable.*

Corollary 2.4. *Let M be an arbitrary power mean or the logarithmic mean, and N an arbitrary mean defined on the interval $\mathbb{R}^+$. Then the means M and N are G-composable.*

Proof. The power means and the logarithmic mean are strict means, hence our statement immediately follows from the previous corollary. □

Remark, that the composition of means defined by non-continuous means may exist if one of them can be bounded by a strict mean.

Corollary 2.5. *Let f be a bounded function on $(\mathbb{R}^+)^2$. Let*

$$
M(x,y) = M_{f(x,y)}(x,y) = \begin{cases} \left(\dfrac{x^{f(x,y)}+y^{f(x,y)}}{2}\right)^{\frac{1}{f(x,y)}} & \text{if} \quad f(x,y) \neq 0 \\ \sqrt{xy} & \text{if} \quad f(x,y) = 0 \end{cases},
$$

and N be an arbitrary mean defined on $\mathbb{R}^+$. Then there $M \otimes N(a,b)$ exists for each pair of real numbers $a, b \in \mathbb{R}^+$.

If one of the means is bounded by a strict mean, and the other is the maximum-mean (minimum-mean), then from the fact of convergence we can obtain the limit value as well:

Corollary 2.6. *Let L be a continuous mean defined on I, such that for each pair of numbers $x, y \in I$, where $x \neq y$,*

$$
L(x,y) > \min\{x,y\}\,,
$$

moreover, let M be an arbitrary mean on I, such that for each pair of numbers $x, y \in I$: $L(x,y) \le M(x,y)$. For every $a, b \in I$, where $a < b$, define the sequence $(a_n^*)_{n=1}^{\infty}$ as follows: $a_1^* = a$ and for each $n \in \mathbb{N}$, $a_{n+1}^* = M(a_n^*, b)$. Then

$$\lim_{n\to\infty} a_n^* = b\,.$$

Proof. The assertion immediately follows from case a) of Theorem 2.1 for the means M and N, where $N(a,b) = \max\{a,b\}$. □

We will show that the continuity condition in Theorem 2.1 cannot be omitted. Apart from trivial means (minimum- and maximum means) there exist other means that are not G-composable.

It is not difficult to construct non-continuous means M and N which are not G-composable.

For $a \in (0,1)$ and $b \in (2,3)$ define

$$M(a,b) = \frac{a+1}{2} \quad \text{and} \quad N(a,b) = \frac{2+b}{2}.$$

Then, for the sequences (a_n) and (b_n) defined by Gauss' iteration, we have $a_n \in (0,1)$ and $b_n \in (2,3)$ for $n = 1, 2, 3, \ldots$. So, $M \otimes N(a,b)$ does not exist.

The means M, N constructed above are not homogeneous.

On the other hand, we have:

Theorem 2.7. *There exist symmetric, homogeneous means H, K defined on $\mathbb{R}^+$ such that for each pair of real numbers $x, y \in \mathbb{R}^+$ with $x \neq y$*

$$\min\{x,y\} < H(x,y) \le K(x,y) < \max\{x,y\},$$

however, $H \otimes K(a_1, b_1)$ does not exist for any pair of real numbers $a_1, b_1 \in \mathbb{R}^+$, where $a_1 \neq b_1$.

Proof. Define the function $f : \mathbb{R}^+ \to \mathbb{R}^+$ as follows:

$$f(x) = \begin{cases} \frac{\ln 4}{\ln \frac{x}{2}} & \text{if} \quad x \in (2,\infty) \\ \frac{\ln 4}{\ln \frac{(k+1)x}{k+2}} & \text{if} \quad x \in (\frac{k+2}{k+1}, \frac{k+1}{k}] \quad \text{for all } k \in \mathbb{N} \\ 1 & \text{if} \quad x = 1 \\ f\left(\frac{1}{x}\right) & \text{if} \quad x \in (0,1) \end{cases}$$

For each pair of positive real numbers x, y put

$$K(x,y) = M_{f(\frac{x}{y})}(x,y) = \begin{cases} \left(\frac{x^{f(\frac{x}{y})} + y^{f(\frac{x}{y})}}{2}\right)^{\frac{1}{f(\frac{x}{y})}} & \text{if} \quad f(\frac{x}{y}) \neq 0 \\ \sqrt{xy} & \text{if} \quad f(\frac{x}{y}) = 0 \end{cases},$$

and

$$H(x,y) = M_{-f(\frac{x}{y})}(x,y) = \begin{cases} \left(\frac{x^{-f(\frac{x}{y})}+y^{-f(\frac{x}{y})}}{2}\right)^{-\frac{1}{f(\frac{x}{y})}} & \text{if} \quad f(\frac{x}{y}) \neq 0 \\ \sqrt{xy} & \text{if} \quad f(\frac{x}{y}) = 0 \end{cases} .$$

Using the fact that the power mean is symmetric and homogeneous along with $f(\frac{x}{y}) = f(\frac{y}{x})$ we get that the means H and K are symmetric and homogeneous, too.

Now, let a_1, b_1 be arbitrary positive real numbers. Without loss of generality we may assume

$$a_1 < b_1 \quad \text{and} \quad a_1 b_1 = 1 . \tag{2.6}$$

Contruct the sequences (a_n), (b_n) by

$$a_{n+1} = H(a_n, b_n) \quad \text{and} \quad b_{n+1} = K(a_n, b_n) .$$

Evidently the sequence(a_n) is strictly increasing and bounded and the sequence (b_n) is strictly decreasing and bounded. Due to these facts the limits

$$\lim_{n\to\infty} a_n = a \quad \text{and} \quad \lim_{n\to\infty} b_n = b$$

exist and

$$a_1 < a \le b < b_1 . \tag{2.7}$$

Denote $\alpha_n = f\left(\frac{a_n}{b_n}\right)$. Then

$$\begin{aligned} a_{n+1}b_{n+1} &= \left(\frac{a_n^{-\alpha_n} + b_n^{-\alpha_n}}{2}\right)^{\frac{1}{-\alpha_n}} \left(\frac{a_n^{\alpha_n} + b_n^{\alpha_n}}{2}\right)^{\frac{1}{\alpha_n}} \\ &= \left(\frac{a_n^{\alpha_n} + b_n^{\alpha_n}}{\frac{1}{a_n^{\alpha_n}} + \frac{1}{b_n^{\alpha_n}}}\right)^{\frac{1}{\alpha_n}} = a_n b_n . \end{aligned}$$

We immediately obtain that for each positive integer n

$$a_n b_n = a_1 b_1 = 1 \tag{2.8}$$

and hence

$$ab = a_1 b_1 = 1 . \tag{2.9}$$

It follows that $H \otimes K(a_1, b_1)$ exists if and only if $a = b = 1$.

Consider the function

$$g(x) = \left(\frac{\left(\frac{1}{x}\right)^{f(x^2)} + x^{f(x^2)}}{2}\right)^{\frac{1}{f(x^2)}} .$$

From (2.8) and (2.9) it follows that

$$\begin{aligned} g(b_n) &= \left(\frac{\left(\frac{1}{b_n}\right)^{f\left(b_n^2\right)} + b_n^{f\left(b_n^2\right)}}{2} \right)^{\frac{1}{f\left(b_n^2\right)}} \\ &= \left(\frac{a_n^{f\left(\frac{b_n}{a_n}\right)} + b_n^{f\left(\frac{b_n}{a_n}\right)}}{2} \right)^{\frac{1}{f\left(\frac{b_n}{a_n}\right)}} = b_{n+1}\,. \end{aligned} \tag{2.10}$$

Let $I_1 = (\sqrt{2}, \infty)$. For each positive integer $k \geq 2$, define $I_k = (\sqrt{\frac{k+1}{k}}, \sqrt{\frac{k}{k-1}}]$. Then evidently

$$\bigcup_{k=1}^{\infty} I_k = (1, \infty) \tag{2.11}$$

and

$$I_k \cap I_l = \emptyset \quad \text{if} \quad k \neq l\,. \tag{2.12}$$

Now we will prove the following implication:

$$\text{if} \quad x \in I_k \quad \text{then} \quad g(x) \in I_k\,. \tag{2.13}$$

Let k be an arbitrary positive integer, and x a real number such that $x \in I_k$. So

$$2 = \frac{k+1}{k} < x^2 \quad \text{if} \quad k = 1\,,$$

or

$$\frac{k+1}{k} < x^2 \leq \frac{k}{k-1} \quad \text{if} \quad k \geq 2\,.$$

From the definition of the function f we obtain in both cases that

$$f\left(x^2\right) = \frac{\ln 4}{\ln \frac{x^2 k}{k+1}}\,.$$

Consequently,

$$\begin{aligned} f\left(x^2\right) \ln \frac{x^2 k}{(k+1)} &= \ln 4 \\ \left(\frac{x^2 k}{k+1}\right)^{f\left(x^2\right)} &= 4 \\ x^{2f\left(x^2\right)} &= 4\left(\frac{k+1}{k}\right)^{f\left(x^2\right)} \\ x^{f\left(x^2\right)} &= 2\left(\sqrt{\frac{k+1}{k}}\right)^{f\left(x^2\right)} \end{aligned}$$

Thus,

$$g(x) = \left(\frac{\left(\frac{1}{x}\right)^{f\left(x^2\right)} + x^{f\left(x^2\right)}}{2} \right)^{\frac{1}{f\left(x^2\right)}} > \left(\frac{x^{f\left(x^2\right)}}{2} \right)^{\frac{1}{f\left(x^2\right)}}$$

$$= \left(\frac{2\left(\sqrt{\frac{k+1}{k}}\right)^{f\left(x^2\right)}}{2} \right)^{\frac{1}{f\left(x^2\right)}} = \sqrt{\frac{k+1}{k}}\,.$$

On the other hand, from $\frac{1}{x} < x$ and the fact that $g(x)$ is the power mean of the numbers x and $\frac{1}{x}$, we obtain $g(x) < x$. Thus, $g(x) \in I_k$.

Finally, we will show that $H \otimes K(a_1, b_1)$ does not exist. According to (2.9) it is sufficient to show that $b > 1$.

In view of (2.6) and (2.11), there exists a well defined positive integer k such that $b_1 \in I_k$. However, by (2.10) and (2.13),

$$b_n \in I_k$$

for each positive integer n; thus,

$$b_n > \sqrt{\frac{k+1}{k}}\,.$$

It follows that

$$b = \lim_{n\to\infty} b_n \geq \sqrt{\frac{k+1}{k}} > 1\,,$$

which concludes the proof. □

References

[1] Borwein, J.M., Borwein, P.B., *Pi and the AGM–a Study in Analytic Number Theory and Computational Complexity*, John Wiley & Sons, New York, 1986.

[2] Carlson, B. C., Algorithms involving arithmetic and geometric means, *Amer. Math. Monthly*, Vol. 78 (1971), 496–505.

[3] Carlson, B. C., The logarihmic mean, *Amer. Math. Monthly*, Vol. 79 (1972), 615–618.

[4] Csiba, P., Filip, F., Tóth, T.J., Distribution of terms of a logarithmic sequence, *Annales Mathematicae et Informaticae*, Vol. 34, (2007), 33–45.

[5] Daróczy, Z., Maksa, Gy., Páles. Zs., Functional equations involving means and their Gauss composition *Proc. Amer. Math. Soc.*, Vol. 134 Number 2, (2006), 521–530.

[6] Daróczy, Z., Páles Zs., Gauss-composition of means and the solution of the Matkowski-Sutô problem, *Publ. Math. Debrecen*, 61 (2002), no. 1–2, 157–218.

[7] Gauss, C. F., Nachlass. *Aritmetisch-geometrisches Mittel*, (1800), Werke, Bd. 3, Königlichen Gesell. Wiss., Göttingen, 1876, 361–403.

[8] Šustek, J., Compound Means and Fast Computation of Radicals, *Applied Mathematics*, 5 (2014), 2493–2517.

[9] Toader, G., Complementary means and double sequences, *Annales Academiae Paedagogicae Cracoviensis Studia Mathematica*, Vol 6 (2007), 7–18.

[10] Toader, G., Some remarks on means, *Anal. Numér. Théor. Approx.*, 20 (1991), 97–109.

[11] Toader, G., Toader, S., *Greek Means and the Arithmetic-Geometric Mean*, RGMIA Mono-graphs, Victoria University, 2005. `http://www.staff.vu.edu.au/rgmia/monographs.asp`

Annales Mathematicae et Informaticae
43 (2014) pp. 67–91
http://ami.ektf.hu

Divisible and cancellable subsets of groupoids*

Tamás Glavosits[a], **Árpád Száz**[b]

[a]Department of Applied Mathematics, University of Miskolc, Miskolc-Egyetemváros, Hungary
matgt@uni-miskolc.hu

[b]Department of Mathematics, University of Debrecen, Debrecen, Hungary
szaz@science.unideb.hu

Submitted April 22, 2014 — Accepted October 30, 2014

Abstract

In this paper, after listing some basic facts on groupoids, we establish several simple consequences and equivalents of the following basic definitions and their obvious counterparts.

For some $n \in \mathbb{N}$, a subset U of a groupoid X is called

(1) n-cancellable if $nx = ny$ implies $x = y$ for all $x, y \in U$,

(2) n-divisible if for each $x \in U$ there exists $y \in U$ such that $x = ny$.

Moreover, for some $A \subset \mathbb{N}$, the set U is called A-divisible (A-cancellable) if it is n-divisible (n-cancellable) for all $n \in A$.

Our main tools here are the sets $n^{-1}x = \{y \in X : x = ny\}$ satisfying $n(n^{-1}x) \subset \{x\} \subset n^{-1}(nx)$ for all $n \in \mathbb{N}$ and $x \in X$. They can be used to briefly reformulate properties (1) and (2), and naturally turn a uniquely $\mathbb{N}$-divisible commutative group into a vector space over $\mathbb{Q}$.

Keywords: Groupoids, divisible and cancellable sets.

MSC: 20L05, 20A05.

---

*The work of the authors has been supported by the Hungarian Scientific Research Fund (OTKA) Grant NK-81402.

1. A few basic facts on groupoids

Definition 1.1. If X is a set and $+$ is a function of X^2 to X, then the function $+$ is called a *binary operation* on X, and the ordered pair $X(+) = (X, +)$ is called a *groupoid.*

Remark 1.2. In this case, we may simply write $x + y$ in place of $+(x, y)$ for all $x, y \in X$. Moreover, we may also simply write X in place of $X(+)$.

Instead of groupoids, it is more customary to consider only *semigroups* (associative grupoids) or even *monoids* (semigroups with zero). However, several definitions on semigroups can be naturally extended to groupoids.

Definition 1.3. If X is a groupoid, then for any $x \in X$ and $n \in \mathbb{N}$, we define

$$nx = x \quad \text{if} \quad n = 1 \qquad \text{and} \qquad nx = (n-1)x + x \quad \text{if} \quad n > 1.$$

Now, by induction, we can easily prove the following two basic theorems.

Theorem 1.4. *If X is a semigroup, then for any $x \in X$ and $m, n \in \mathbb{N}$ we have*
(1) $(m + n)x = mx + nx$,
(2) $(nm)x = n(mx)$.

Proof. To prove (2), note that if $(nm)x = n(mx)$ holds for some $n \in \mathbb{N}$, then by (1) we also have

$$((n+1)m)x = (nm + m)x = (nm)x + mx = n(mx) + mx = (n+1)(mx). \qquad \square$$

Theorem 1.5. *If X is a semigroup, then for any $m, n \in \mathbb{N}$ and $x, y \in X$, with $x + y = y + x$, we have*
(1) $mx + ny = ny + mx$,
(2) $n(x + y) = nx + ny$.

Proof. To prove (1), note that if $x + ny = ny + x$ holds for some $n \in \mathbb{N}$, then we also have

$$x + (n+1)y = x + ny + y = ny + x + y = ny + y + x = (n+1)y + x.$$

While, to prove (2), note that if $n(x+y) = nx+ny$ holds for some $n \in \mathbb{N}$, then by (1) we also have

$$\begin{aligned}(n+1)(x+y) &= n(x+y) + x + y = nx + ny + x + y = \\ &= nx + x + ny + y = (n+1)x + (n+1)y. \qquad \square\end{aligned}$$

Definition 1.6. If in particular X is a groupoid with zero, then we also define $0x = 0$ for all $x \in X$.

Moreover, if more specially X is a *group*, then we also define $(-n)x = n(-x)$ for all $x \in X$ and $n \in \mathbb{N}$.

Lemma 1.7. *If X is a group, then for any $x \in X$ and $n \in \mathbb{N}$ we also have $(-n)x = -(nx)$.*

Proof. By using $-x + x = 0 = x + (-x)$ and Theorem 1.5, we can at once see that $n(-x) + nx = n(-x + x) = n0 = 0$. Therefore, $n(-x) = -(nx)$, and thus the required equality is also true. □

Now, we can also easily prove the following counterparts of Theorems 1.4 and 1.5.

Theorem 1.8. *If X is a group, then for any $x \in X$ and $k, l \in \mathbb{Z}$ we have*
(1) $(kl)x = k(lx)$,
(2) $(k + l)x = kx + lx$.

Theorem 1.9. *If X is a group, then for any $k, l \in \mathbb{Z}$ and $x, y \in X$, with $x + y = y + x$, we have*
(1) $kx + ly = ly + kx$,
(2) $k(x + y) = kx + ky$.

Proof. To prove (2), note that by Lemma 1.7, Theorem 1.5 and assertion (1) we have

$$\begin{aligned}(-n)(x + y) &= -\big(n(x + y)\big) = -(nx + ny)\\ &= -(ny) + \big(-(nx)\big) = (-n)y + (-n)x = (-n)x + (-n)y\end{aligned}$$

for all $n \in \mathbb{N}$. Moreover, $0(x + y) = 0 = 0x + 0y$ also holds. □

Remark 1.10. The latter two theorems show that a commutative group X is already a *module* over the ring $\mathbb{Z}$ of all integers.

2. Operations with subsets of groupoids

Definition 2.1. If X is a groupoid with zero, then for any $U \subset X$ we define

$$U_0 = U \cup \{0\} \quad \text{if} \quad 0 \notin U \qquad \text{and} \qquad U_0 = U \setminus \{0\} \quad \text{if} \quad 0 \in U.$$

Remark 2.2. In the sequel, this particular unary operation will mainly be applied to the subsets $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ of the additive group $\mathbb{R}$ of all real numbers.

Definition 2.3. If X is a groupoid, then for any $A \subset \mathbb{N}$, and $U, V \subset X$ we define

$$AU = \big\{nu : n \in A, u \in U\big\} \quad \text{and} \quad U + V = \big\{u + v : u \in U, v \in V\big\}.$$

Remark 2.4. Now, by identifying singletons with their elements, we may simply write $nU = \{n\}U$, $Au = A\{u\}$, $u + V = \{u\} + V$, and $U + v = U + \{u\}$ for all $n \in \mathbb{N}$ and $u, v \in X$.

The notation nU may cause some confusions since in general we only have $nU \subset (n - 1)U + U$ for all $n > 1$. However, assertions 1.4(1),(2) and 1.5(1) can be generalized to sets.

Remark 2.5. If in particular, X is a group, then we may quite similarly define AU for all $A \subset \mathbb{Z}$ and $U \subset X$.

Moreover, we may also naturally define $-U = (-1)U$ and $U - V = U + (-V)$ for all $V \subset X$. However, thus we have $U - U = \{0\}$ if and only if U is a singleton.

Remark 2.6. Moreover, if more specially if X is a *vector space* over K, then we may also quite similarly define AU for all $A \subset K$ and $U \subset X$.

Thus, only two axioms of a vector space may fail to hold for $\mathcal{P}(X)$. Namely, in general, we only have $(\lambda + \mu)U \subset \lambda U + \mu U$ for all $\lambda, \mu \in K$.

The corresponding elementwise operations with subsets of various algebraic structures allow of some more concise treatments of several basic theorems on substructures of these structures.

Remark 2.7. For instance, a subset U of a groupoid X is called a *subgroupoid* of X if U is itself a groupoid with respect to the restriction of the addition on X to $U \times U$.

Thus, U is a subgroupoid of X if and only if U is *superadditive* in the sense $U + U \subset U$. Moreover, if U is a subgroupoid of X, then U is in particular $\mathbb{N}$-*superhomogeneous* in the sense that $\mathbb{N}U \subset U$.

Concerning subgroups, we can prove some more interesting theorems.

Theorem 2.8. *If X is a group, then for a nonvoid subset U of X the following assertions are equivalent:*

(1) *U is a subgroup of X,*

(2) *$-U \subset U$ and $U + U \subset U$,*

(3) *$U - U \subset U$.*

Remark 2.9. Note that if U is a subset of a group X such that $-U \subset U$, then U is already *symmetric* in the sense that $-U = U$.

While, if U is a subset of a groupoid X with zero such that $U + U \subset U$ and $0 \in U$, then U is already *idempotent* in the sense that $U + U = U$.

Therefore, as an immediate consequence of Theorem 2.8, we can also state

Corollary 2.10. *A nonvoid subset U of a group X is a subgroup of X if and only if it is symmetric and idempotent.*

In addition to Theorem 2.8, we can also easily prove the following

Theorem 2.11. *If X is a group, then for any two symmetric subsets U and V of X the following assertions are equivalent:*

(1) *$U + V = V + U$,*

(2) *$U + V$ is symmetric.*

Proof. If (1) holds, then $-(U + V) = -V + (-U) = V + U = U + V$, and thus (2) also holds.

While, if (2) holds, then $U + V = -(U + V) = -V + (-U) = V + U$, and thus (1) also holds. $\square$

Remark 2.12. If U and V are idempotent subsets of a semigroup X such that (1) holds, then

$$U + V + U + V = U + V + V + U = U + V + U = U + U + V = U + V,$$

and thus $U + V$ is also an idempotent subset of X.

Therefore, as an immediate consequence of Theorem 2.11 and Corollary 2.10, we can also state

Theorem 2.13. *If X is a group, then for any two subgroups U and V of X the following assertions are equivalent:*

(1) *$U + V = V + U$,*

(2) *$U + V$ is a subgroup of X.*

Hence, it is clear that in particular we also have the following

Corollary 2.14. *If U and V are commuting subgroups of a group X, then $U + V$ is the smallest subgroup of X containing both U and V.*

Remark 2.15. In the standard textbooks, Theorem 2.13, or its corollary, is usually proved directly without using Theorems 2.8 and 2.11. (See, for instance, Sott [13, p. 18] and Burton [4, p. 118].)

3. Direct sums of subsets of groupoids

Analogously to Fuchs [6, p. 3.15], we may naturally introduce the following

Definition 3.1. If U, V and W are subsets of a groupoid X such that for every $x \in W$ there exists a unique pair $(u_x, v_x) \in U \times V$ such that

$$x = u_x + v_x,$$

then we say that W is the *direct sum* of U and V, and we write $W = U \oplus V$.

Remark 3.2. Thus, in particular we have $W = U + V$. Hence, if in addition X has a zero such that $0 \in V$, we can infer that $U \subset W$.

Moreover, in this particular case for any $x \in U$ we have $x = x + 0$. Hence, by using the unicity of u_x and v_x, we can infer that $u_x = x$ and $v_x = 0$.

Remark 3.3. Therefore, if $W = U \oplus V$ and in particular X has a zero such that $0 \in U \cap V$, then in addition to $W = U + V$ we can also state that $U \cup V \subset W$ and $U \cap V = \{0\}$.

Namely, by Remark 3.2 and its dual, we have $U \subset W$ and $V \subset W$, and thus $U \cup V \subset W$. Moreover, if $x \in U \cap V$, i.e., $x \in U$ and $x \in V$, then we have $v_x = 0$ and $u_x = 0$, and thus $x = u_x + v_x = 0$.

In this respect, we can also easily prove the following

Theorem 3.4. *If U and V are subgroups of a monoid X, with $0 \in U \cap V$, then the following assertions are equivalent:*

(1) $X = U \oplus V$;

(2) $X = U + V$ and $U \cap V = \{0\}$.

Proof. If $x \in X$ such that $x = u_1 + v_1$ and $x = u_2 + v_2$ for some $u_1, u_2 \in U$ and $v_1, v_2 \in V$, then $u_1 + v_1 = u_2 + v_2$, and thus $-u_2 + u_1 = v_2 - v_1$. Moreover, we also have $-u_2 + u_1 \in U$ and $v_2 - v_1 \in V$. Hence, if the second part of (2) holds, we can infer that $-u_2 + u_1 = 0$ and $v_2 - v_1 = 0$. Therefore, $u_1 = u_2$, and $v_1 = v_2$ also hold. □

Remark 3.5. Note that if U and V are subgroups of a monoid X, with $0 \in U \cap V$, such that $X = U + V$, then for any $x \in X$ there exist $u \in U$ and $v \in V$ such that $x = u + v$. Hence, by taking $y = -v - u$, we can see that $x + y = 0$ and $y + x = 0$. Therefore, $-x = y$, and thus X is also a group.

Remark 3.6. Note that if G is a group, then the Descartes product $X = G \times G$, with the coordinatewise addition, is also a group. Moreover,

$$U = \big\{(x, 0) : x \in G\big\} \quad \text{and} \quad V = \big\{(0, y) : y \in G\big\}$$

are subgroups of X such that $X = U + V$ and $U \cap V = \{(0, 0)\}$. Therefore, by Theorem 3.4, we also have $X = U \oplus V$.

Furthermore, it is also worth noticing that the sets U and V are elementwise commuting in the sense that $u + v = v + u$ for all $u \in U$ and $v \in V$.

The importance of elementwise commuting sets is apparent from the following

Theorem 3.7. *If U and V are elementwise commuting subgroupoids of a semigroup X such that $X = U \oplus V$, then the mappings*

$$x \mapsto u_x \quad \textit{and} \quad x \mapsto v_x,$$

where $x \in X$, are additive. Thus, in particular, they are $\mathbb{N}$-homogeneous.

Proof. If $x, y \in X$, then by the assumed associativity and commutativity properties of the addition in X we have

$$x + y = (u_x + v_x) + (u_y + v_y) = (u_x + u_y) + (v_x + v_y).$$

Therefore, since $u_x + u_y \in U$ and $v_x + v_y \in V$, the equalities

$$u_{x+y} = u_x + u_y \quad \text{and} \quad v_{x+y} = v_x + v_y$$

are also true.

Moreover, by induction, it can be easily seen that if f is an additive function of one groupoid X to another Y, then $f(nx) = nf(x)$ for all $n \in \mathbb{N}$ and $x \in X$. □

Remark 3.8. Note that if in particular X has a zero such that $0 \in V$, then by Remark 3.2 the mapping $x \mapsto u_x$, where $x \in X$, is idempotent. Moreover, if $0 \in U$ also holds, then $u_0 = 0$. Thus, the above mapping is also zero-homogeneous.

Remark 3.9. In this respect, it is also worth noticing that if in particular X is a monoid, and U and V are subgroups of X, with $0 \in U \cap V$, then by Remark 3.5 X is also a group, and thus the mappings considered in Theorem 3.7 are actually $\mathbb{Z}$-homogeneous.

Remark 3.10. If in particular X is a vector space, then by using Zorn's lemma [14, p. 38] it can be shown that for each subspace U of X there exists a subspace V of X such that $X = U \oplus V$.

In the standard textbooks, this fundamental decomposition theorem is usually proved with the help of Hamel bases. (See, for instance, Cotlar and Cignoli [5, p. 15] and Taylor and Lay [14, p. 43].)

4. Some further results on elementwise commuting sets

The importance of elementwise commuting sets is also apparent from the following

Theorem 4.1. *If U and V are elementwise commuting, comutative subsets of a semigroup X, then $U + V$ is also commutative.*

Proof. Namely, if $x, y \in U + V$, then there exist $u, \omega \in U$ and $v, w \in V$ such that $x = u + v$ and $y = \omega + w$. Hence, we can already see that

$$x + y = u + v + \omega + w = u + \omega + v + w = \omega + u + w + v = \omega + w + u + v = y + x.$$

Therefore, the required assertion is also true. □

Remark 4.2. Conversely, we can also easily note that if U and V are subsets of a groupoid X such that $U + V$ is commutative and $U \cup V \subset U + V$, then U and V are commutative and elementwise commuting.

Therefore, as an immediate consequence of Theorem 4.1, we can also state

Corollary 4.3. *If U and V are subsets of monoid X such that $0 \in U \cap V$, then the following assertions are equivalent:*

(1) *$U + V$ is commutative,*

(2) *U and V are commutative and elementwise commuting.*

Remark 4.4. Note that if U and V are elementwise commuting subsets of a groupoid X, then we have not only $U + V = V + U$, but also $u + V = V + u$ and $U + v = v + U$ for all $u \in U$ and $v \in V$.

Therefore, it is of some interest to note that we also have the following

Theorem 4.5. *If U and V are subsets of a groupoid X such that $U+V=U\oplus V$, then the following assertions are equivalent:*

(1) *U and V are elementwise commuting,*

(2) *$u+V=V+u$ and $v+U=U+v$ for all $u\in U$ and $v\in V$,*

(3) *$u+V\subset V+u$ and $v+U\subset U+v$ for all $u\in U$ and $v\in V$,*

(4) *$V+u\subset u+V$ and $U+v\subset v+U$ for all $u\in U$ and $v\in V$.*

Proof. Namely, if for instance (3) holds, then for any $u\in U$ and $v\in V$ we have $u+v\in u+V\subset V+u$. Therefore, there exists $w\in V$ such that $u+v=w+u$. Moreover, again by (3), we can see that $w+u\in w+U\subset U+w$. Therefore, there exists $\omega\in U$ such that $w+u=\omega+w$. Thus, we also have $u+v=\omega+w$. Hence, by using that $U+V=U\oplus V$, we can infer that $u=\omega$ and $v=w$. Therefore, $u+v=v+u$, and thus (1) is also true. □

Remark 4.6. In this respect, it is also worth noticing that if U is a subset and V is a subgroup of a monoid X, then the following assertions are also equivalent:

(1) $U+v=v+U$ for all $v\in V$,

(2) $U+v\subset v+U$ for all $v\in V$,

(3) $v+U\subset U+v$ for all $v\in V$.

Namely, if for instance (2) holds, then we have

$$v+U=v+U+0=v+U+(-v)+v\subset v+(-v)+U+v=0+U+v=U+v$$

for all $v\in V$, and thus (1) also holds.

Concerning elementwise commuting sets, by Theorems 1.5 and 1.9, we can at once state the following two theorems.

Theorem 4.7. *If U and V are elementwise commuting sets of a semigroup X, then the sets $\mathbb{N}U$ and $\mathbb{N}V$ are also also elementwise commuting.*

Theorem 4.8. *If U and V are elementwise commuting subsets of a group X, then the sets $\mathbb{Z}U$ and $\mathbb{Z}V$ are also also elementwise commuting.*

Moreover, concerning elementwise commuting sets, we can also easily prove

Theorem 4.9. *If U and V are elementwise commuting subsets of a semigroup X such that U is commutative, then U and $U+V$ are also elementwise commuting.*

Proof. Suppose that $x\in U$ and $y\in U+V$. Then, there exist $u\in U$ and $v\in V$ such that $y=u+v$. Moreover, by the assumed commutativity properties of U and V, we have

$$x+y=x+u+v=u+x+v=u+v+x=y+x.$$

Therefore, the required assertion is also true. □

Remark 4.10. The importance of elementwise commuting subsets will also be well shown by the forthcoming theorems of Section 10.

5. Divisible and cancellable subsets of groupoids

Analogously to Hall [10, p. 197], Fuchs [6, p. 58] and Scott [13, p. 95], we may naturally introduce the following

Definition 5.1. A subset U of a groupoid X is called *n-divisible*, for some $n \in \mathbb{N}$, if $U \subset nU$.

Now, the subset U may also be naturally called *A-divisible*, for some $A \subset \mathbb{N}$, if it is n-divisible for all $n \in A$.

Remark 5.2. Thus, U is n-divisible if and only if it is *n-subhomogeneous.* That is, for each $x \in U$ there exists $y \in U$ such that $x = ny$.

Therefore, the set U may be naturally called *uniquely n-divisible* if for each $x \in U$ there exists a unique $y \in U$ such that $x = ny$.

Moreover, the subset U may also be naturally called *uniquely A-divisible* if it is uniquely n-divisible for all $n \in A$.

Now, in addition to Definition 5.1, we may also naturally introduce the following definition which has also been utilized in [8].

Definition 5.3. A subset U of a groupoid X is called *n-cancellable*, for some $n \in \mathbb{N}$, if $nx = ny$ implies $x = y$ for all $x, y \in U$.

Now, the set U may also be naturally called *A-cancellable*, for some $A \subset \mathbb{N}$, if it is n-cancellable for all $n \in A$.

Remark 5.4. Thus, if U is both n-divisible and n-cancellable, then U is already uniquely n-divisible.

Namely, if $x \in U$ such that $x = ny_1$ and $x = ny_2$ for some $y_1, y_2 \in U$, then we also have $ny_1 = ny_2$, and hence $y_1 = y_2$.

Remark 5.5. Moreover, by using some obvious analogues of Definitions 5.1 and 5.3, we can also see that if U is a both k-divisible and k-cancellable subset of a group X, for some $k \in \mathbb{Z}$, then U is already uniquely k-divisible.

In this respect, it is worth noticing that the following two theorems are also true.

Theorem 5.6. *If U is an n-superhomogeneous subset of a groupoid X, for some $n \in \mathbb{N}$, then the following assertions are equivalent:*

(1) *U is uniquely n-divisible,*

(2) *U is both n-divisible and n-cancellable.*

Proof. Namely, if (1) holds and $x, y \in U$ such that $nx = ny$, then because of $nx \in U$ and (1) we also have $x = y$. Therefore, U is n-cancellable, and thus (2) also holds. The converse implication (2) $\Longrightarrow$ (1) has been proved in Remark 5.4. □

Theorem 5.7. *If U is a k-superhomogeneous subset of a group X, for some $k \in \mathbb{Z}$, then following assertions are equivalent:*

(1) *U is uniquely k-divisible,*

(2) *U is both k-divisible and k-cancellable.*

By using the corresponding definitions and Theorems 1.4 and 1.8, we can easily prove the following two theorems.

Theorem 5.8. *If U is an n-divisible subset of a semigroup X, for some $n \in N$, and $p, q \in \mathbb{N}$ such that $n = pq$ and U is q-superhomogeneous, then U is also p-divisible.*

Proof. If $x \in U$, then by the n-divisibility of U there exists $y \in U$ such that $x = ny$. Now, by using Theorem 1.4, we can see that $x = ny = (pq)y = p(qy)$. Hence, because of $qy \in U$, it is clear that U is also p-divisible. □

Theorem 5.9. *If U is an k-divisible subset of a semigroup X, for some $k \in Z$, and $p, q \in \mathbb{Z}$ such that $k = pq$ and U is q-superhomogeneous, then U is also p-divisible.*

In addition to the latter two theorems, it is also worth proving the following

Theorem 5.10. *For a subset U of a monoid X, the following assertions are equivalent:*

(1) $U \subset \{0\}$,

(2) U *is* 0-*divisible,*

(3) U *is* $\mathbb{N}_0$-*divisible.*

By using the corresponding definitions and Theorems 1.4 and 1.8, we can also easily prove the following counterparts of Theorems 5.8, 5.9 and 5.10.

Theorem 5.11. *If U is an m-superhomogeneous, both n- and m-cancellable subset of a semigroup X, for some $m, n \in \mathbb{N}$, then U is also nm-cancellable.*

Proof. If $x, y \in U$ such that $(nm)x = (nm)y$, then by Theorem 1.4 we also have $n(mx) = n(my)$. Hence, by using the n-cancelability of U, and the fact that $mx, my \in U$, we can infer that $mx = my$. Now, by the m-cancelability of U, we can see that $x = y$. Therefore, U is also nm-cancellable. □

Theorem 5.12. *If U is an l-superhomogeneous, both k- and l-cancellable subset of a group X, for some $k, l \in \mathbb{N}$, then U is also kl-cancellable.*

Theorem 5.13. *For a subset U of a monoid X, the following assertions are equivalent:*

(1) $\operatorname{card}(U) \leq 1$,

(2) U *is* 0-*cancellable,*

(3) U *is* $\mathbb{N}_0$-*cancellable.*

In addition to Theorems 5.8 and 5.9, we can also prove the following two theorems.

Theorem 5.14. *If U is a uniquely n-divisible, n-superhomogeneous subset of a semigroup X for some $n \in N$, and $p, q \in \mathbb{N}$ such that $n = pq$ and U is q-superhomogeneous, then U is also uniquely p-divisible.*

Proof. By Theorem 5.8 and Remark 5.4, we need only show that now U is also p-cancellable.

For this, note that if $x, y \in U$ such that $px = py$, then by Theorem 1.4 we also have $nx = (qp)x = q(px) = q(py) = (qp)x = ny$. Moreover, by Theorem 5.6, U is now n-cancellable. Therefore, we necessarily have $x = y$. □

Theorem 5.15. *If U is a uniquely k-divisible, k-superhomogeneous subset of a group X, for some $k \in Z$, and $p, q \in \mathbb{Z}$ such that $n = pq$ and U is q-superhomogeneous, then U is also uniquely p-divisible.*

Remark 5.16. Note that in assertion (3) of Theorem 5.10 we may also write "uniquely $\mathbb{N}_0$-divisible" instead of "$\mathbb{N}_0$-divisible".

6. Some further results on divisible and cancellable sets

Theorem 6.1. *If U is a k-divisible, symmetric subset of a group X, for some $k \in \mathbb{Z}$, then U is also $-k$-divisible.*

Proof. If $x \in U$, then by the k-divisibility of U there exists $y \in U$ such that $x = ky$. Now, by using Theorem 1.8, we can see that

$$x = ky = \big((-k)(-1)\big)y = (-k)\big((-1)y\big) = (-k)(-y).$$

Hence, since now we also have $-y \in -U = U$, it is clear that U is also $-k$-divisible. □

From this theorem, it is clear that in particular we also have

Corollary 6.2. *If U is an $\mathbb{N}$-divisible, symmetric subset of a group X, then U is $\mathbb{Z}_0$-divisible.*

Analogously to Theorem 6.1, we can also easily prove the following

Theorem 6.3. *If U is a k-cancellable subset of a group X, for some $k \in \mathbb{Z}$, then U is also $-k$-cancellable.*

Proof. If $x, y \in U$ such that $(-k)x = (-k)y$, then by Theorem 1.8 we also have

$$kx = \big((-1)(-k)\big)x = (-1)\big((-k)x\big) = (-1)\big((-k)y\big) = \big((-1)(-k)\big)y = ky.$$

Hence, by the assumption, it follows that $x = y$, and thus the required assertion is also true. □

From this theorem, it is clear that in particular we also have

Corollary 6.4. *If U is an $\mathbb{N}$-cancellable subset of a group X, then U is also $\mathbb{Z}_0$-cancellable.*

Now, as an immediate consequence of Theorems 6.1 and 6.3 and Remark 5.5, we can also state

Theorem 6.5. *If U is a uniquely k-divisible, symmetric subset of a group X, for some $k \in \mathbb{Z}$, then U is also uniquely $-k$-divisible.*

Hence, it is clear that in particular we also have

Corollary 6.6. *If U is a uniquely $\mathbb{N}$-divisible, symmetric subset of a group X, then U is also uniquely $\mathbb{Z}_0$-divisible.*

Remark 6.7. By using some obvious analogues of Definition 5.1 and Remark 5.2, we can also easily see that a subset U of a vector space X over K is k-divisible (uniquely k-divisible), for some $k \in K_0$, if and only if $k^{-1}x \in U$ for all $x \in U$. That is, $k^{-1}U \subset U$.

Remark 6.8. If U is an n-cancellable subset of a groupoid X with zero, for some $n \in \mathbb{N}$, such that $0 \in U$, then $nx = 0$ implies $x = 0$ for all $x \in U$.

Namely, if $x \in U$ such that $nx = 0$, then by the corresponding definitions we also have $nx = n0$, and hence $x = 0$.

Remark 6.9. Quite similarly, we can also see that if U is a k-cancellable subset of a group X, for some $k \in \mathbb{Z}$, such that $0 \in U$, then $kx = 0$ implies $x = 0$ for all $x \in U$.

Now, by using the letter observation and Corollary 6.4, we can also easily prove

Theorem 6.10. *If U is an $\mathbb{N}$-cancellable subset of a group X such that $0 \in U$, then $kx = lx$ implies $k = l$ for all $k, l \in \mathbb{Z}$ and $x \in U_0$.*

Proof. Assume on the contrary that there exist $k, l \in \mathbb{Z}$ and $x \in U_0$ such that $kx = lx$, but $k \neq l$. Then, by using Theorem 1.8, we can see that

$$(k-l)x = \big(k + (-l)\big)x = kx + (-l)x = lx + (-l)x = \big(l + (-l)\big)x = 0x = 0.$$

Hence, by using Corollary 6.4 and Remark 6.9, we can infer that $x = 0$. This contradiction proves the theorem. □

From the above theorem, by taking $l = 0$, we can immediately derive

Corollary 6.11. *If U is an $\mathbb{N}$-cancellable subset of group X such that $0 \in U$, then $kx = 0$ implies $k = 0$ for all $k \in \mathbb{Z}$ and $x \in U_0$.*

In addition to Remark 6.9, we can also easily prove the following

Theorem 6.12. *If X is a commutative group, then for each $k \in \mathbb{Z}$ the following assertions are equivalent:*

(1) *X is k-cancellable;*

(2) *$kx = 0$ implies $x = 0$ for all $x \in X$.*

Proof. From Remark 6.9, we can see that (1) $\Longrightarrow$ (2) even if the group X is not assumed to be commutative.

Moreover, if $x, y \in X$ such that $kx = ky$, then by using Theorem 1.9 we can see that

$$k(x-y) = k(x+(-y)) = kx + k(-y) = ky + k(-y) = k(y+(-y)) = k0 = 0.$$

Hence, if (2) holds, then we can already infer that $x - y = 0$, and thus $x = y$. Therefore, (1) also holds. $\square$

From this theorem, by using Corollary 6.4, we can immediately derive

Corollary 6.13. *If X is a commutative group such that $nx = 0$ implies $x = 0$ for all $n \in \mathbb{N}$ and $x \in X$, then X is $\mathbb{Z}_0$-cancellable.*

Remark 6.14. By using an obvious analogue of Definition 5.3, we can also easily see that every subset U of a vector space X over K is K_0-cancellable. Moreover, $kx = lx$ implies $k = l$ for all $k, \in K$ and $x \in X_0$.

7. Characterizations of divisible and cancellable sets

Definition 7.1. If X is a groupoid, then for any $x \in X$ and $n \in \mathbb{N}$ we define

$$n^{-1}x = \{y \in X : x = ny\}.$$

Remark 7.2. Now, having in mind the definition of the image of a set under a relation, for any $U \subset X$, we may also naturally define $n^{-1}U = \bigcup_{x \in U} n^{-1}x$.

Thus, we can easily see that $n^{-1}U = \{y \in X : ny \in U\}$. Namely, if for instance, $y \in n^{-1}U$, then by the above definition there exists $x \in U$ such that $y \in n^{-1}x$. Hence, by Definition 7.1, it already follows that $ny = x \in U$.

By using Definition 7.1, we can also easily prove the following

Theorem 7.3. *If X is a groupoid, then for any $x \in X$ and $n \in \mathbb{N}$ we have*

(1) $n(n^{-1}x) \subset \{x\}$,

(2) $\{x\} \subset n^{-1}(nx)$.

Proof. Since $nx = nx$, it is clear that $x \in n^{-1}(nx)$. Therefore, (2) is true.

Moreover, if $z \in n(n^{-1}x)$ then there exists $y \in n^{-1}x$ such that $z = ny$. Hence, since $y \in n^{-1}x$ implies $ny = x$, we can infer that $z = x$. Therefore, (1) is also true. $\square$

Remark 7.4. Now, by using this theorem, for any $U \subset X$, we can also easily prove that $n(n^{-1}U) \subset U \subset n^{-1}(nU)$.

For instance, by using Theorem 7.3 and Remark 7.2, we can easily see that

$$U = \bigcup_{x \in U} \{x\} \subset \bigcup_{x \in U} n^{-1}(nx) = n^{-1}\Big(\bigcup_{x \in U} \{nx\}\Big) = n^{-1}(nU).$$

By using an obvious analogue of Definition 7.1, we can also easily prove the following

Theorem 7.5. *If X is a group, then for any $x \in X$ and $k \in \mathbb{Z}$ we have*

(1) $k\big(k^{-1}x\big) \subset \{x\}$,

(2) $\{x\} \subset k^{-1}\big(kx\big)$.

Remark 7.6. Now, by using this theorem, for any $U \subset X$, we can also easily prove that $k\big(k^{-1}U\big) \subset U \subset k^{-1}\big(kU\big)$.

However, it is now more important to note that, by using the corresponding definitions, we can also easily prove the following three theorems.

Theorem 7.7. *If X is a groupoid, then for any $U \subset X$ and $n \in \mathbb{N}$ the following assertions are equivalent:*

(1) *U is n-divisible,*

(2) $U \cap n^{-1}x \neq \emptyset$ *for all* $x \in U$.

Theorem 7.8. *If X is a groupoid, then for any $U \subset X$ and $n \in \mathbb{N}$ the following assertions are equivalent:*

(1) *U is uniquely n-divisible,*

(2) $\operatorname{card}\big(U \cap n^{-1}x\big) = 1$ *for all* $x \in U$.

Theorem 7.9. *If X is a groupoid, then for any $U \subset X$ and $n \in \mathbb{N}$ the following assertions are equivalent:*

(1) *U is n-cancellable,*

(2) $\operatorname{card}\big(U \cap n^{-1}(nx)\big) \leq 1$ *for all* $x \in U$.

Proof. If $x \in X$ and $y_1, y_2 \in U \cap n^{-1}(nx)$, then $y_1, y_2 \in U$ and $y_1, y_2 \in n^{-1}(nx)$, and thus $ny_1 = nx = ny_2$. Hence, if (1) holds, we can infer that $y_1 = y_2$, and thus (2) also holds.

Conversely, if $x, y \in U$ such that $nx = ny$, then by Definition 7.1 we have $y \in n^{-1}(nx)$. Moreover, by Theorem 7.3, we also have $x \in n^{-1}(nx)$. Therefore, $x, y \in U \cap n^{-1}(nx)$. Hence, if (2) holds, we can infer that $x = y$. Therefore, (1) also holds. $\square$

Analogously to the latter three theorems, we can also easily prove the following three theorems.

Theorem 7.10. *If X is a group, then for any $U \subset X$ and $k \in \mathbb{Z}$ the following assertions are equivalent:*

(1) *U is k-divisible,*

(2) $U \cap k^{-1}x \neq \emptyset$ *for all* $x \in U$.

Theorem 7.11. *If X is a group, then for any $U \subset X$ and $k \in \mathbb{Z}$ the following assertions are equivalent:*

(1) *U is uniquely k-divisible,*

(2) $\operatorname{card}\big(U \cap k^{-1}x\big) = 1$ *for all $x \in U$.*

Theorem 7.12. *If X is a group, then for any $U \subset X$ and $k \in \mathbb{Z}$ the following assertions are equivalent:*

(1) *U is k-cancellable,*

(2) $\operatorname{card}\big(U \cap k^{-1}(kx)\big) \leq 1$ *for all $x \in X$.*

Moreover, as a simple reformulation of Theorem 6.12, we can also state

Theorem 7.13. *A commutative group X, then for any $k \in \mathbb{Z}$ the following assertions are equivalent:*

(1) *X is k-cancellable,*

(2) $k^{-1}0 \subset \{0\}$,

(3) $k^{-1}0 = \{0\}$.

Remark 7.14. Quite similarly, by Remark 6.8, we can also state that if U is an n-cancellable subset of groupoid X with zero, for some $n \in \mathbb{N}$, such that $0 \in U$, then $U \cap n^{-1}0 = \{0\}$.

Remark 7.15. Moreover, by Remark 6.9, we can also state that if U is a k-cancellable subset of group X, for some $k \in \mathbb{Z}$, such that $0 \in U$, then $U \cap k^{-1}0 = \{0\}$.

In addition to Theorem 7.13 and Remarks 7.14 and 7.15, it is also worth proving

Theorem 7.16. *The following assertions hold:*

(1) *If X is a commutative group, then $k^{-1}0$ is a subgroup of X for all $k \in \mathbb{Z}$.*

(2) *If X is a commutative monoid, then $n^{-1}0$ is a submonoid of X for all $n \in \mathbb{N}_0$.*

However, it is now more important to note that in addition to Theorems 7.7, 7.10, 7.9 and 7.12, we can also easily prove the following four theorems.

Theorem 7.17. *If X is a groupoid, then for any $n \in \mathbb{N}$ the following assertions are equivalent:*

(1) *X is n-divisible,*

(2) $\{x\} \subset n\big(n^{-1}x\big)$ *for all $x \in X$,*

(3) $\{x\} = n\big(n^{-1}x\big)$ *for all $x \in X$.*

Proof. If (1) holds, then by Theorem 7.7, for every $x \in X$, we have $n^{-1}x \neq \emptyset$, and thus $n\big(n^{-1}x\big) \neq \emptyset$. Moreover, by Theorem 7.3, we also have $n\big(n^{-1}x\big) \subset \{x\}$. Therefore, (3) also holds. The implication (2) $\Longrightarrow$ (1) is even more obvious by Theorem 7.7. $\square$

Theorem 7.18. *If X is a group, then for any $k \in \mathbb{Z}$ the following assertions are equivalent:*

(1) *X is k-divisible,*

(2) *$\{x\} \subset k\big(k^{-1}x\big)$ for all $x \in X$,*

(3) *$\{x\} = k\big(k^{-1}x\big)$ for all $x \in X$.*

Theorem 7.19. *If X is a groupoid, then for any $n \in \mathbb{N}$ the following assertions are equivalent:*

(1) *X is n-cancellable,*

(2) *$n^{-1}\big(nx\big) \subset \{x\}$ for all $x \in X$,*

(3) *$n^{-1}\big(nx\big) = \{x\}$ for all $x \in X$.*

Proof. If (1) holds, then by Theorem 7.9, for every $x \in X$, we have

$$\operatorname{card}\big(n^{-1}(nx)\big) \leq 1.$$

Moreover, by Theorem 7.3, we also have $\{x\} \subset n^{-1}(nx)$. Therefore, (3) also holds. The implication (2) $\Longrightarrow$ (1) is even more obvious by Theorem 7.9. □

Theorem 7.20. *If X is a group, then for any $k \in \mathbb{Z}$ the following assertions are equivalent:*

(1) *X is k-cancellable,*

(2) *$k^{-1}\big(kx\big) \subset \{x\}$ for all $x \in X$,*

(3) *$k^{-1}\big(kx\big) = \{x\}$ for all $x \in X$.*

Now, as some immediate consequences of the latter four theorems, and Theorems 5.6 and 5.7, we can also state the following two theorems.

Theorem 7.21. *If X is a groupoid, then for any $n \in \mathbb{N}$ the following assertions are equivalent:*

(1) *X is uniquely n-divisible,*

(2) *$n^{-1}\big(nx\big) \subset \{x\} \subset n\big(n^{-1}x\big)$ for all $x \in X$,*

(3) *$n^{-1}\big(nx\big) = \{x\} = n\big(n^{-1}x\big)$ for all $x \in X$.*

Theorem 7.22. *If X is a group, then for any $k \in \mathbb{Z}$ the following assertions are equivalent:*

(1) *X is uniquely k-divisible,*

(2) *$k^{-1}\big(kx\big) \subset \{x\} \subset k\big(k^{-1}x\big)$ for all $x \in X$,*

(3) *$k^{-1}\big(kx\big) = \{x\} = k\big(k^{-1}x\big)$ for all $x \in X$.*

8. Some further results on the sets $n^{-1}x$ and $k^{-1}x$

In addition to Theorem 7.3, we can also prove the following

Theorem 8.1. *If X is a semigroup, then for any $x \in X$ and $m, n \in \mathbb{N}$ we have:*
(1) $m\big(n^{-1}x\big) \subset n^{-1}\big(mx\big)$,
(2) $m^{-1}\big(n^{-1}x\big) \subset (mn)^{-1}x$,
(3) $m\big((mn)^{-1}x\big) \subset n^{-1}x$,
(4) $n^{-1}x \subset (mn)^{-1}(mx)$.

Proof. If $y \in n^{-1}x$, then by Definition 7.1 we have $x = ny$. Hence, by using Theorem 1.4, we can infer that

$$mx = m(ny) = (mn)y = (nm)y = n(my).$$

Thus, by Definition 7.1, we also have

$$y \in (mn)^{-1}(mx) \qquad \text{and} \qquad my \in n^{-1}(mx).$$

Hence, we can already see that (4) and (1) are true.

On the other hand, if $y \in (mn)^{-1}x$, then by Definition 7.1 and Theorem 1.4 we have

$$x = (mn)y = (nm)y = n(my).$$

Thus, by Definition 7.1, we also have $my \in n^{-1}x$. Hence, we can already see that (3) is also true.

Finally, if $y \in m^{-1}\big(n^{-1}x\big)$, then by Remark 7.2, we have $my \in n^{-1}x$. Hence, by using Definition 7.1 and Theorem 1.4, we can infer that

$$x = n(my) = (nm)y = (mn)y.$$

Thus, by Definition 7.1, we also have $y = (mn)^{-1}x$. Hence, we can already see that (2) is also true. □

From this theorem, by Theorem 7.8, it is clear that in particular we also have

Corollary 8.2. *If X is a uniquely $\mathbb{N}$-divisible semigroup, then for any $x \in X$ and $m, n \in \mathbb{N}$ we have:*
(1) $m\big(n^{-1}x\big) = n^{-1}\big(mx\big)$,
(2) $m^{-1}\big(n^{-1}x\big) = (mn)^{-1}x$,
(3) $m\big((mn)^{-1}x\big) = n^{-1}x$,
(4) $n^{-1}x = (mn)^{-1}\big(mx\big)$.

Analogously to Theorem 8.1, we can also prove the following

Theorem 8.3. *If X is a group, then for any $x \in X$ and $k, l \in \mathbb{Z}$ we have:*
(1) $k\big(l^{-1}x\big) \subset l^{-1}\big(kx\big)$,
(2) $k^{-1}\big(l^{-1}x\big) \subset (kl)^{-1}x$,
(3) $k\big((kl)^{-1}x\big) \subset l^{-1}x$,
(4) $l^{-1}x \subset (kl)^{-1}\big(kx\big)$.

Hence, by Corollary 6.6 and Theorem 7.11, it is clear that in particular we have

Corollary 8.4. *If X is a uniquely $\mathbb{N}$-divisible group, then for any $x \in X$ and $k, l \in \mathbb{Z}_0$ we have:*
(1) $k\big(l^{-1}x\big) = l^{-1}\big(kx\big)$,
(2) $k^{-1}\big(l^{-1}x\big) = (kl)^{-1}x$,
(3) $k\big((kl)^{-1}x\big) = l^{-1}x$,
(4) $l^{-1}x = (kl)^{-1}\big(kx\big)$.

In addition to Theorem 8.1, we can also prove the following

Theorem 8.5. *If X is a commutative semigroup, then for any $x, y \in X$ and $n \in \mathbb{N}$ we have*
$$n^{-1}x + n^{-1}y \subset n^{-1}(x+y).$$

Proof. If $z \in n^{-1}x$ and $w \in n^{-1}y$, then by using Definition 7.1 and Theorem 1.5, we can see that
$$x + y = nz + nw = n(z+w).$$
Therefore, by Definition 7.1, we also have $z + w \in n^{-1}(x+y)$. Hence, we can already see that the required inclusion is also true. $\square$

From this theorem, by Theorem 7.8, it is clear that in particular we also have

Corollary 8.6. *If X is a uniquely $\mathbb{N}$-divisible commutative semigroup, then for any $x, y \in X$ and $n \in \mathbb{N}$ we have*
$$n^{-1}(x+y) = n^{-1}x + n^{-1}y.$$

Analogously to Theorem 8.5, we can also prove the following

Theorem 8.7. *If X is a commutative group, then for any $k \in \mathbb{Z}$ and $x, y \in X$ we have*
$$k^{-1}x + k^{-1}y \subset k^{-1}(x+y).$$

Hence, by Corollary 6.6 and Theorem 5.11, it is clear that in particular we also have

Corollary 8.8. *If X is a uniquely $\mathbb{N}$-divisible commutative semigroup, then for any $k \in \mathbb{Z}_0$ and $x, y \in X$ we have*
$$k^{-1}(x+y) = k^{-1}x + k^{-1}y.$$

Remark 8.9. In the latter two theorems and their corollaries, the commutativity assumptions on X can be weakened.

For instance, in Theorem 8.5 it would be enough to assume only that the sets $n^{-1}x$ and $n^{-1}y$ are elementwise commuting.

9. Uniquely $\mathbb{N}$-divisible semigroups

In addition to Corollary 8.2, we can also easily prove the following

Lemma 9.1. *If X is a uniquely $\mathbb{N}$-divisible semigroup and $m, n, p, q \in \mathbb{N}$ such that $m/n = p/q$, then for every $x \in X$ we have*

$$m\big(n^{-1}x\big) = p\big(q^{-1}x\big).$$

Proof. By Theorem 7.21, we have

$$n\big(n^{-1}x\big) = \{x\} = q\big(q^{-1}x\big).$$

Hence, by using that $mq = pn$, we can infer that

$$(mq)\,\big(n\big(n^{-1}x\big)\big) = (pn)\big(q\big(q^{-1}x\big)\big).$$

Now, by using Theorem 1.4, we can also see that

$$(nq)\big(m\big(n^{-1}x\big)\big) = (nq)\big(p\big(q^{-1}x\big)\big).$$

Hence, by using Theorem 5.6 and 5.11, we can see that the required equality is also true. □

Analogously to this lemma, we can also prove the following

Lemma 9.2. *If X is a uniquely $\mathbb{N}$-divisible group and $n, q \in \mathbb{N}$ and $m, p \in \mathbb{Z}$ such that $m/n = p/q$, then for every $x \in X$ we have*

$$m\big(n^{-1}x\big) = p\big(q^{-1}x\big).$$

Because of the above lemmas, we may naturally introduce the following two definitions.

Definition 9.3. If X is a uniquely $\mathbb{N}$-divisible semigroup, then for any $x \in X$ and $m, n \in \mathbb{N}$ we define

$$\big(m/n\big)x = m\big(n^{-1}x\big).$$

Definition 9.4. If X is a uniquely $\mathbb{N}$-divisible group, then for any $x \in X$, $n \in \mathbb{N}$ and $m \in \mathbb{Z}$ we define

$$\big(m/n\big)x = m\big(n^{-1}x\big).$$

By using Definition 9.3 and Corollary 8.2, we can easily prove the following

Theorem 9.5. *If X is a uniquely $\mathbb{N}$-divisible semigroup, then for any $x \in X$ and $r, s \in \mathbb{Q}$, with $r, s > 0$, we have*

(1) $(r+s)x = rx + sx$,

(2) $(rs)x = r(sx)$.

Proof. By the definition of $\mathbb{Q}$, there exists $m, n, p, q \in \mathbb{N}$ such that $r = m/n$ and $s = p/q$. Now, by using Theorems 7.8 and 1.4 and Corollary 8.2, we can see that

$$\begin{aligned}(r+s)x &= \big((m/n)+(p/q)\big)x = \big((mq+pn)/(nq)\big)x \\ &= (mq+pn)\big((nq)^{-1}x\big) = (mq)\big((nq)^{-1}x\big) + (pn)\big((nq)^{-1}x\big) \\ &= m\,\big(q\big((nq)^{-1}x\big)\big) + p\,\big(n\big((nq)^{-1}x\big)\big) = m\big(n^{-1}x\big) + p\big(q^{-1}x\big) \\ &= (m/n)x + (p/q)x = rx + sx\end{aligned}$$

and

$$\begin{aligned}(rs)x &= \big((m/n)(p/q)\big)x = \big((mp)/(nq)\big)x = (mp)\big((nq)^{-1}x\big) \\ &= m\,\big(p\big((nq)^{-1}x\big)\big) = m\,\big(p\big(n^{-1}\big(q^{-1}x\big)\big)\big) = m\,\big(n^{-1}\big(p\big(q^{-1}x\big)\big)\big) \\ &= m\,\big(n^{-1}\big((p/q)x\big)\big) = (m/n)\big((p/q)x\big) = r(sx).\end{aligned}$$ □

Analogously to this theorem, we can also prove the following

Theorem 9.6. *If X is a uniquely $\mathbb{N}$-divisible group, then for any $x \in X$ and $r, s \in \mathbb{Q}$ we have*

(1) $(r+s)x = rx + sx$,

(2) $(rs)x = r(sx)$.

By using Definition 9.3 and Corollary 8.6, we can also easily prove the following

Theorem 9.7. *If X is a uniquely $\mathbb{N}$-divisible commutative semigroup, then for any $x, y \in X$ and $r \in \mathbb{Q}$, with $r > 0$, we have*

$$r(x+y) = rx + ry.$$

Proof. By the definition of $\mathbb{Q}$, there exist $m, n \in \mathbb{N}$ such that $r = m/n$. Now, by using Corollary 8.6 and Theorem 1.5, we can see that

$$\begin{aligned}r(x+y) &= (m/n)(x+y) = m\big(n^{-1}(x+y)\big) = m\big(n^{-1}x + n^{-1}y)\big) \\ &= m\big(n^{-1}x\big) + m\big(n^{-1}y\big) = m\big(n^{-1}x\big) + m\big(n^{-1}y\big) \\ &= (m/n)x + (m/n)y = rx + ry.\end{aligned}$$ □

Analogously to this theorem, we can also prove the following

Theorem 9.8. *If X is a uniquely $\mathbb{N}$-divisible commutative group, then for any $x, y \in X$ and $r \in \mathbb{Q}$, we have*

$$r(x+y) = rx + ry.$$

Now, as an immediate consequence of Theorems 9.6 and 9.7, we can also state

Corollary 9.9. *If X is a uniquely $\mathbb{N}$-divisible commutative group, then X, with the multiplication given in Definition 9.4, is a vector space over $\mathbb{Q}$.*

Remark 9.10. Note that, by Remark 6.7, every vector space X over $\mathbb{Q}$ is uniquely $\mathbb{Q}_0$-divisible.

Now, by using Corollary 9.9, from the basic decomposition theorem of vector spaces, mentioned in Remark 3.10, we can immediately derive the following

Theorem 9.11. *If X is a uniquely $\mathbb{N}$-divisible commutative group, then for each $\mathbb{N}$-divisible subgroup U of X there exists an $\mathbb{N}$-divisible subgroup V of X such that $X = U \oplus V$.*

Remark 9.12. Note that now, by Theorem 5.6, X is $\mathbb{N}$-cancellable, and thus actually both U and V are also uniquely $\mathbb{N}$-divisible. Moreover, by Corollary 6.6, U, V and X are uniquely $\mathbb{Z}_0$-divisible.

Remark 9.13. To see that the $\mathbb{N}$-divisibility of U is an essential condition in the above theorem, we can note that $\mathbb{Z}$ is an additive subgroup of the field $\mathbb{Q}$ such that, for any $\mathbb{N}$-superhomogeneous subset V of $\mathbb{Q}$ with $\mathbb{Z} \cap V \subset \{0\}$, we have $V \subset \{0\}$, and thus $\mathbb{Z} + V \subset \mathbb{Z}$.

Namely, if $x \in V$, then since $V \subset \mathbb{Q}$ there exist $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ such that $x = m/n$. Moreover, since V is $\mathbb{N}$-superhomogeneous, we have

$$m = n(m/n) = nx \in V.$$

Hence, since $m \in \mathbb{Z}$ and $\mathbb{Z} \cap V \subset \{0\}$ also hold, we can infer that $m = 0$, and thus $x = 0$. Therefore, $V \subset \{0\}$, and thus $\mathbb{Z} + V \subset \mathbb{Z} + \{0\} = \mathbb{Z}$.

In addition to Remark 9.13, it is also worth proving the following

Theorem 9.14. *If X is an $\mathbb{N}$-cancellable group and $a \in X$, then $U = \mathbb{Z}a$ is a commutative subgroup of X such that, for every $\mathbb{N}$-divisible symmetric subset V of $X \setminus \{a\}$, we have $U \cap V \subset \{0\}$.*

Proof. By Theorems 1.8, 1.9 and 2.8, it is clear that U is a commutative subgroup of X. Therefore, we need only prove that $U \cap V \subset \{0\}$.

For this, assume on the contrary that there exists $x \in U \cap V$ such that $x \neq 0$. Then, by the definition of U, there exists $k \in Z$ such that $x = ka$. Hence, since $x \neq 0$, we can infer that $k \neq 0$. Therefore, by Corollary 6.2, there exists $v \in V$ such that $x = kv$. Thus, we have $ka = kv$. Hence, by using Corollary 6.4, we can infer that $a = v$, and thus $a \in V$. This contradiction proves the required inclusion. $\square$

From this theorem, by using Theorem 3.4, we can immediately derive

Corollary 9.15. *If X and U are as in Theorem 9.14, then for every $\mathbb{N}$-divisible subgroup V of X with $a \notin V$ and $X = U + V$ we have $X = U \oplus V$.*

Remark 9.16. Concerning Theorem 9.11, it is also worth mentioning that Baer [1] in 1936 already proved that if U is an $\mathbb{N}$-divisible subgroup of a commutative group X, then there exists a subgroup V of X such that $X = U \oplus V$.

Moreover, Kertész [11] in 1951 proved that if X is a commutative group such that the order of each element of X is a square-free number, then for every subgroup U of X there exists a subgroup V of X such that $X = U \oplus V$.

Surprisingly, the above two results were already considered to be well-known by Baer in [1, p.1] and [3, p. 504]. Moreover, it is also worth mentioning that Hall [9], analogously to Kertész [11], also proved an "if and only if result".

10. Operations with divisible and cancellable sets

Theorem 10.1. *If U is an n-divisible subset of a semigroup X, for some $n \in \mathbb{N}$, then for every $m \in \mathbb{N}$ the set mU is also n-divisible.*

Proof. If $x \in mU$, then by the definition of mU there exists $u \in U$ such that $x = mu$. Moreover, by the n-divisibility of U, there exists $v \in U$ such that $u = nv$. Hence, by using Theorem 1.4, we can see that $x = mu = m(nv) = n(mv)$. Thus, since $mv \in mU$, the required assertion is also true. □

Moreover, as a certain converse to this theorem, we can also prove

Theorem 10.2. *If U is an m-cancellable, n-superhomogeneous subset of a semigroup X, for some $m, n \in \mathbb{N}$, such that mU is n-divisible, then U is also n-divisible.*

Proof. If $x \in U$, then by the definition mU we also have $mx \in mU$. Therefore, by the n-divisibility of mU, there exists $v \in mU$ such that $mx = nv$. Moreover, by the definition of mU, there exists $y \in U$ such that $v = my$. Now, by using Theorem 1.4, we can see that $mx = nv = n(my) = m(ny)$. Hence, by using the m-cancellability of U and the fact that $ny \in U$, we can already infer that $x = ny$. Therefore, the required assertion is also true. □

Quite similarly to Theorems 10.1 and 10.2, we can also prove the following two theorems.

Theorem 10.3. *If U is a k-divisible subset of a group X, for some $k \in \mathbb{Z}$, then for every $l \in \mathbb{Z}$ the set lU is also k-divisible.*

Theorem 10.4. *If U is an l-cancellable, k-superhomogeneous subset of a group X, for some $l, k \in \mathbb{N}$, such that lU is k-divisible, then U is also k-divisible.*

In addition to Theorem 10.1, we can also easily prove the following

Theorem 10.5. *If U and V are elementwise commuting, n-divisible subsets of a semigroup X, for some $n \in \mathbb{N}$, then $U + V$ is also n-divisible.*

Proof. If $x \in U+V$, then by the definition of $U+V$ there exist $u \in U$ and $v \in V$ such that $x = u+v$. Moreover, since U and V are n-divisible, there exist $\omega \in U$ and $w \in V$ such that $u = n\omega$ and $v = nw$. Hence, by using Theorem 1.5, we can see that $x = u+v = n\omega + nw = n(\omega + w)$. Thus, since $\omega + w \in U+V$, the required assertion is also true. □

Moreover, as a certain converse to this theorem, we can also prove

Theorem 10.6. *If U and V are elementwise commuting, n-superhomogeneous subsets of a monoid X, for some $n \in \mathbb{N}$, such that $U+V$ is n-divisible, and $U+V = U \oplus V$ and $0 \in V$, then U is also n-divisible.*

Proof. If $x \in U$, then because of $0 \in V$ we also have $x \in U+V$. Thus, by the n-divisibility of $U+V$, there exists $y \in U+V$ such that $x = ny$. Moreover, by the definition of $U+V$, there exist $u \in U$ and $v \in V$ such that $y = u+v$. Now, by using Theorem 1.5, we can see that

$$x = ny = n(u+v) = nu + nv.$$

Moreover, we can also note that $x \in U+V$, $nu \in U$ and $nv \in V$. Hence, since $x = x+0$ also holds with $x \in U$ and $0 \in V$, by using the assumption $U+V = U \oplus V$, we can already infer that $x = nu$. Therefore, U is also n-divisible. □

Quite similarly to Theorems 10.5 and 10.6, we can also prove the following two theorems.

Theorem 10.7. *If U and V are elementwise commuting, k-divisible subsets of a semigroup X, for some $k \in \mathbb{Z}$, then $U+V$ is also k-divisible.*

Theorem 10.8. *If U and V are elementwise commuting, k-superhomogeneous subsets of a group X, for some $k \in \mathbb{Z}$, such that $U+V$ is k-divisible, and $U+V = U \oplus V$ and $0 \in V$, then U is also k-divisible.*

Hence, by Theorem 3.4, it is clear that in particular we also have

Corollary 10.9. *If U and V are elementwise commuting subgroups of a group X such that $U+V$ is k-divisible, for some $k \in \mathbb{Z}$ such that $U \cap V = \{0\}$, then U and V are also n-divisible.*

In addition to Theorem 10.5, we can also prove the following

Theorem 10.10. *If U and V are elementwise commuting, n-superhomogeneous subsets of a semigroup X, for some $n \in \mathbb{N}$ such that U and V are n-cancellable and $U+V = U \oplus V$, then $U+V$ is also n-cancellable.*

Proof. For this, assume that $x, y \in U+V$ such $nx = ny$. Then, by the definition of $U+V$, there exist $u, \omega \in U$ and $v, w \in V$ such that $x = u+v$ and $y = \omega + w$. Hence, by using Theorem 1.5, we can see that

$$nu + nv = n(u+v) = nx = ny = n(\omega + w) = n\omega + nw.$$

Moreover, we can also note that $nu, n\omega \in U$ and $nv, nw \in V$, and thus $nu + nv, n\omega + nw \in U + V$. Now, by using that $U + V = U \oplus V$, we can see that $nu = n\omega$ and $nv = nw$. Hence, by using the n-cancellability of U and V, we can already infer that $u = \omega$ and $v = w$. Therefore, $x = u + v = \omega + w = y$, and thus the required assertion is also true. $\square$

Remark 10.11. Now, as a trivial converse to this theorem, we can also state that if U and V subsets of a monoid X such that $U + V$ is n-cancellable, for some $n \in \mathbb{Z}$, and $0 \in U \cap V$, then U and V are also n-cancellable.

Quite similarly to Theorem 10.10, we can also prove the following

Theorem 10.12. *If U and V are elementwise commuting, k-superhomogeneous subsets of a group X, for some $k \in \mathbb{Z}$ such that U and V are k-cancellable and $U + V = U \oplus V$, then $U + V$ is also k-cancellable.*

Hence, by Theorem 3.4, it is clear that in particular we also have

Corollary 10.13. *If U and V are elementwise commuting subgroups of a group X such that U and V are k-cancellable for some $k \in \mathbb{Z}$, and $U \cap V = \{0\}$, then $U + V$ is also k-cancellable.*

Remark 10.14. In an immediate continuation of this paper, by using the notion of the order

$$n_a = \inf\{n \in \mathbb{N} : na = 0\}$$

of an element a of a monoid (resp. group) X, we shall investigate the divisibility and cancellability properties of the set $\mathbb{N}_0 a + V$ (resp. $\mathbb{Z}a + V$) for some substructures V of X.

Acknowledgements. The authors are indebted to the anonymous referee for pointing out several grammatical errors and misprints in the original manuscript.

References

[1] Baer, R., The subgroups of the elements of finite order of an Abelian group, *Ann. Math.* 37(1936), 766-781.

[2] Baer, R., Abelian groups that are direct summands of every containing Abalian group, *Bull. Amer. Math. Soc.* 46(1940), 800-806.

[3] Baer, R., Absolute retracts in group theory, *Bull. Amer. Math. Soc.* 52(1946), 501-506.

[4] Burton, D.M., Abstract Algebra, *WM.C. Brown Publishers*, Dubuque, Iowa, 1988.

[5] Cotlar, M., Cignoli, R., An Introduction to Functional Analysis, *North-Holland Publishing Company*, Amsterdam, 1974.

[6] Fuchs, L., Abelian Groups, *Publishing House of the Hungarian Academy of Sciences*, Budapest, 1958.

[7] Gacsályi, S., On pure subgroups and direct summands of abelian groups, *Publ. Math. Debrecen* 4(1955), 89–92.

[8] Glavosits, T., Száz, Á., Constrictions and extensions of free and controlled additive relations, *In: Th.M. Rassias (ed.), Handbook in Functional Equations: Functional Inequalities*, to appear.

[9] Hall, M., Complemented groups, *J. London Math. Soc.* 12(1937), 201–204.

[10] Hall, M., The Theory of Groups, *The Macmillan Company*, New York, 1959.

[11] Kertész, A., On groups every subgroup of which is a direct summand, *Publ. Math. Debrecen* 2(1951), 74–75.

[12] Kertész, A., Szele, T., On abelian groups every multiple of which is a direct summand, *Acta Sci. Math. Szeged* 14(1952), 157–166.

[13] Scott, W.R., Group Theory, *Prentice-Hall, INC.*, Englewood Cliffs, New Yersey, 1964.

[14] Taylor, A.E., Lay, D.C., Introduction to Functional Analysis, *Robert E. Krieger Publishing Company*, Malabar, Florida, 1986.

Annales Mathematicae et Informaticae
43 (2014) pp. 93–101
http://ami.ektf.hu

Commutator identities on group algebras*

Tibor Juhász

Institute of Mathematics and Informatics
Eszterházy Károly College
juhaszti@ektf.hu

Submitted October 30, 2014 — Accepted December 21, 2014

Abstract

Let K be a field of characteristic $p > 2$, and G a nilpotent group with commutator subgroup of order p^n. Denote by $(KG)_*$ the set of symmetric elements of the group algebra KG with respect to an oriented classical involution. Then KG satisfies all Lie commutator identities of degree $p^n + 1$ or more. We will show that $(KG)_*$ satisfies a Lie commutator identity of degree less than $p^n + 1$ if and only if G' is not cyclic. Consequently, if G' is cyclic, then the Lie nilpotency index and the Lie derived length of $(KG)_*$ are just the same as of KG, namely $p^n + 1$ and $\lceil \log_2(p^n + 1)\rceil$, respectively. The corresponding results on the set of symmetric units of KG are also obtained.

Keywords: Group ring, involution, polynomial identity, group identity, derived length, Lie nilpotency index, nilpotency class

MSC: 16W10, 16S34, 16U60, 16N40

1. Introduction

The Lie derived length and the Lie nilpotency index of group algebras and their certain subsets have been studied separately for many decades. Both of these properties can be characterized by specific polynomial identities, where the polynomials are multilinear Lie monomials. In this paper we investigate group algebras satisfying general multilinear Lie monomial (Lie commutator) identities, and from that draw conclusions about the above properties.

*This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP 4.2.4. A/2-11-1-2012-0001 'National Excellence Program'.

Let KG denote the group algebra of a group G over a field K. Then KG, with the Lie commutator $[x, y] = xy - yx$ serving as the Lie bracket, can be considered as a Lie algebra. Let S be a nonempty subset of KG. We will consider the elements of S as *Lie commutators of weight* 1 on S, and inductively, an element $[x, y]$ of KG, where x and y are Lie commutators of weight u and v on S with $u + v = r$, will be called a *Lie commutator of weight* r on S.

Denote by $K\langle x_1, \ldots, x_m\rangle$ the polynomial ring in the non-commuting indeterminates $x_1, \ldots, x_m$ over K. The set S is said to *satisfy a polynomial identity* if there exists a nonzero polynomial in $K\langle x_1, \ldots, x_m\rangle$ such that $f(s_1, \ldots, s_m) = 0$ for all $s_1, \ldots, s_m \in S$. Let now X be the set of the indeterminates in $K\langle x_1, \ldots, x_m\rangle$. A Lie commutator of weight r on X is called a *multilinear Lie monomial of degree* r, if it is linear in each of its indeterminates. We will say that the subset S of KG *satisfies a Lie commutator identity of degree* r, if there exists a nonzero multilinear Lie monomial f of degree r in $K\langle x_1, \ldots, x_m\rangle$ with $f(s_1, \ldots, s_m) = 0$ for all $s_1, \ldots, s_m \in S$. Then we also say: S satisfies the Lie commutator identity $f(x_1, \ldots, x_m) = 0$. We will denote by $f(S)$ the image of the set S under the polynomial function f.

For subsets $V, W \subseteq KG$, by the symbol $[V, W]$ we mean the subspace of KG generated by all Lie commutators $[v, w]$ with $v \in V, w \in W$. Set $\gamma_1(S) = \delta^{[0]}(S) = S$, and by induction, let $\gamma_{n+1}(S) = [\gamma_n(S), S]$ and $\delta^{[n+1]}(S) = [\delta^{[n]}(S), \delta^{[n]}(S)]$. S is said to be *Lie nilpotent*, if $\gamma_n(S) = 0$, and *Lie solvable*, if $\delta^{[n]}(S) = 0$ for some integer n. The first such n is called the *Lie nilpotency index* or the *Lie derived length* of S and denoted by $t_L(S)$ and $\mathrm{dl}_L(S)$, respectively. It is obvious that S is Lie nilpotent of index n, or Lie solvable of derived length n, if and only if it satisfies the polynomial identity

$$[x_1, \ldots, x_n] = 0, \tag{1.1}$$

or

$$[x_1, \ldots, x_{2^n}]^\circ = 0, \tag{1.2}$$

respectively, where the Lie commutators on the left-hand sides are defined inductively to be

$$[x_1, \ldots, x_n] = [[x_1, \ldots, x_{n-1}], x_n]$$

and

$$[x_1, \ldots, x_{2^n}]^\circ = [[x_1, \ldots, x_{2^{n-1}}]^\circ, [x_{2^{n-1}+1}, \ldots, x_{2^n}]^\circ]$$

with $[x_1, x_2]^\circ = [x_1, x_2]$, and n is the least such integer.
Besides Lie nilpotence and Lie solvability, many other properties can be originated from Lie commutator identities. For example, KG is said to be *Lie centre-by-metabelian* (or *Lie centrally metabelian*), if $\delta^{[2]}(KG)$ is central in KG, or, equivalently, KG satisfies the Lie commutator identity

$$[[[x_1, x_2], [x_3, x_4]], x_5] = 0 \tag{1.3}$$

of degree 5. However, as we will see, the identities (1.1) and (1.2) play special roles.

For a prime p we say that G is p-abelian, if its commutator subgroup G' is a finite p-group. By definition, the 0-abelian groups are the abelian groups. In what follows, p will always denote the characteristic of the field K. According to [7], KG satisfies a polynomial identity if and only if G has a p-abelian subgroup of finite index. Now, assume that the Lie ideal L of KG satisfies the Lie commutator identity $f(x_1,\dots,x_m)=0$. If f is of degree 1, then $f(x_1,\dots,x_m)=x_i$ for some $i\in\{1,\dots,m\}$, so $L=\delta^{[0]}(L)=f(L)$. Suppose that there exists k such that $\delta^{[k]}(L)\subseteq f(L)$ whenever f is of degree less than r. Let now f be of degree r. Then f can be expressed as the Lie commutator of the multilinear Lie monomials f_1 and f_2 of degrees less than r. By the inductive hypothesis, there exist k_1,k_2 such that $\delta^{[k_1]}(L)\subseteq f_1(L)$ and $\delta^{[k_2]}(L)\subseteq f_2(L)$. Assume that $k_1\le k_2$, and let $k=k_2+1$. Then

$$\begin{aligned}\delta^{[k]}(L)=[\delta^{[k_2]}(L),\delta^{[k_2]}(L)]&\subseteq[\delta^{[k_1]}(L),\delta^{[k_2]}(L)]\\ &\subseteq[f_1(L),f_2(L)]=f(L).\end{aligned}$$

We have just proved that if L satisfies a Lie commutator identity, then L is Lie solvable. The converse is trivial.

The Lie solvable group algebras are described in [6]: KG is Lie solvable if and only if one of the following conditions holds: (i) $p\neq 2$, and G is p-abelian; (ii) $p=2$, and G has a 2-abelian subgroup of index at most 2. Consequently, for $p=0$, KG satisfies a Lie commutator identity precisely if G is abelian, and then, of course, KG satisfies all Lie commutator identities of degree at least 2. Therefore, in the sequel we can restrict ourselves to the case only when $p>0$ and G is nonabelian. In [6], a necessary and sufficient condition can also be found for the Lie nilpotence of the group algebra KG: KG is Lie nilpotent if and only if G is nilpotent and p-abelian. It is easy to check that if $S\subseteq KG$ is Lie nilpotent of class n (in other words, S satisfies (1.1)), then S satisfies all Lie commutator identities of degree at least n.

Applying Theorems 3 and 6 of [5], it is not so hard to derive that on group algebras, all Lie commutator identities of degree r are equivalent while $r\le 4$. Nevertheless, according to [9], the group algebra $\mathbb{F}_3D_6$, where $\mathbb{F}_3$ denotes the field of three elements and D_6 the dihedral group of order 6, satisfies the identity (1.3), but, by [6], it does not satisfy (1.1) for $n=5$. It is worth mentioning here that the question of the equivalence of Lie commutator identities of the same degree is raised in the "Dniester Notebook: Unsolved Problems in the Theory of Rings and Modules" (see Problem 2.6 in [8, p. 482]).

Let now $*$ be an involution of the group algebra KG, and let $(KG)_*=\{x\in KG : x^*=x\}$ the set of symmetric elements with respect to $*$. Evidently, $(KG)_*$ is a subspace of KG, but not always closed under Lie commutator. Although the classification of all involutions of group algebras is still open, the exploration of the algebraic properties of symmetric elements is an extensively studied area of group algebras. Most of the results are known with respect to the so-called *classical involution*, which sends every element of G into its inverse. By $*$ we will understand a more general involution introduced by S. P. Novikov. Let $\sigma\colon G\to$

$\{\pm 1\}$ a homomorphism and let $*\colon KG \to KG$ be given by

$$\left(\sum_{g\in G}\alpha_g g\right)^* = \sum_{g\in G}\alpha_g\sigma(g)g^{-1}.$$

This involution is called *oriented classical involution* of KG. According to [3], it can happen that $(KG)_*$ satisfies a Lie commutator identity, but the whole KG does not satisfy the same identity.

Now, we will assign group commutators to Lie monomials. Let τ be the mapping from the set of all Lie commutators on the subset $X = \{x_1, \dots, x_m\}$ of $K\langle x_1, \dots, x_m\rangle$ into the free group F with generators $u_1, \dots, u_n$, given by $\tau(x_i) = u_i$, and for the Lie commutator $[x, y]$ of weight $r > 1$ on X, let $\tau([x, y])$ be the group commutator of $\tau(x)$ and $\tau(y)$. The word w in F will be called a *multilinear group commutator of degree* r, if it is the image of a multilinear Lie monomial of degree r under τ. Denote by $U(S)$ the set of units of the set $S \subseteq KG$. We will say that $U(S) \neq \emptyset$ satisfies a *group commutator identity of degree* r, if there exists a nontrivial multilinear group commutator $w(u_1, \dots, u_n)$ of degree r in the free group with generators $v_1, \dots, v_n$ such that $w(h_1, \dots, h_n) = 1$ for all $h_1, \dots, h_n \in U(S)$.

We will say that $U(S)$ is *nilpotent of class* $n-1$, or *solvable of length* n, if $U(S)$ satisfies the group commutator identity $(v_1, \dots, v_n) = 1$, or $(v_1, \dots, v_{2^n})^\circ = 1$, respectively, where the group commutators $(v_1, \dots, v_n)$ and $(v_1, \dots, v_{2^n})^\circ$ are defined by induction, analogously to (1.1) and (1.2), and n is the first such integer. The nilpotency class and the derived length of $U(S)$ will be denoted by $\mathrm{cl}(U(S))$ and $\mathrm{dl}(U(S))$, respectively.

Our main theorem is the following.

Theorem 1.1. *Let K be a field of characteristic $p > 2$, and let G be a nilpotent p-abelian group with cyclic commutator subgroup. Then:*

(i) $(KG)_*$ satisfies no Lie commutator identity of degree less than $|G'| + 1$;

(ii) provided that G is torsion, $U_*(KG)$ satisfies no group commutator identity of degree less than $|G'| + 1$.

By Theorem 1 of [2], if G' is not cyclic, then $t_L(KG) \leq |G'|$, or in other words, KG satisfies all Lie commutator identities of degree at least $|G'|$. Combining this result with Theorem 1.1, we can state the next corollary.

Corollary 1.2. *Let K be a field of characteristic $p > 2$, and let G be a nilpotent p-abelian group. Then the group algebra KG satisfies all Lie commutator identities of degree $|G'| + 1$ or more, and $U(KG)$ satisfies all group commutator identities of degree $|G'|+1$ or more. Furthermore, $(KG)_*$ satisfies a Lie commutator identity of degree less than $|G'| + 1$ if and only if G' is not cyclic. Provided that G is torsion, $U_*(KG)$ satisfies a group commutator identity of degree less than $|G'| + 1$ if and only if G' is not cyclic.*

Finally, we draw conclusions about the Lie nilpotency index and the Lie derived length of $(KG)_*$, such as the nilpotency class and derived length of $U_*(KG)$.

Corollary 1.3. *Let K be a field of characteristic $p > 2$, and let G be a nilpotent p-abelian group. Then $t_L((KG)_*) \leq |G'| + 1$, with equality if and only if G' is cyclic. Provided that G is torsion, $\mathrm{cl}(U_*(KG)) \leq |G'|$, with equality if and only if G' is cyclic.*

Corollary 1.4. *Let K be a field of characteristic $p > 2$, and let G be a nilpotent p-abelian group with cyclic commutator subgroup. Then*

$$\mathrm{dl}_L((KG)_*) = \mathrm{dl}_L(KG) = \lceil \log_2(|G'| + 1) \rceil.$$

In addition, if G is torsion, then $\mathrm{dl}(U_*(KG)) = \mathrm{dl}_L(KG)$.

2. Proof of Theorem 1.1

Let G be a finite p-group with cyclic commutator subgroup of order p^n, where p is an odd prime, and let K be a field of characteristic p. We will denote by $\omega(KG)$ and $\omega(KG')$ the augmentation ideals of KG and KG', respectively. The assumption guarantees that they are nilpotent ideals, and by Lemma 3 of [1], the relations

$$\begin{aligned}
&[\omega(KG')^m, \omega(KG)^l] \subseteq \omega(KG)^{l-1}\omega(KG')^{m+1};\\
&[\omega(KG)^k, \omega(KG)^l] \subseteq \omega(KG)^{k+l-2}\omega(KG');\\
&[\omega(KG)^k\omega(KG')^m, \omega(KG)^l\omega(KG')^n] \subseteq \omega(KG)^{k+l-2}\omega(KG')^{n+m+1}
\end{aligned} \tag{2.1}$$

hold for all $k, l, m, n \geq 1$. By definition, $\omega(KG)^0 = KG$.

We will also use the following well-known identity: for any $g \in G$ and integer k

$$g^k - 1 \equiv k(g-1) \pmod{\omega(KG)^2}. \tag{2.2}$$

Let I_r denote the ideal $\omega(KG)^3\omega(KG')^{r-1} + KG\omega(KG')^r$ of KG, and let S be the subspace of KG spanned by the elements

$$(a-1)(a^{-1}-1),\ (b-1)(b^{-1}-1),\ (ab-1)((ab)^{-1}-1),$$

with $a, b \in G$ such that the commutator $x = (a, b)$ is of order p^n. For the multilinear Lie monomial f we will denote by w_f the multilinear group commutator $\tau(f)$.

Lemma 2.1. *S satisfies no Lie commutator identity of degree less than $p^n + 1$, and $1 + S$ satisfies no group commutator identity of degree less than $p^n + 1$.*

Proof. We show that for arbitrary multilinear Lie commutator $f(x_1, \dots, x_m)$ of degree r, and for any element v of the set $V = \{(a-1)^2, (b-1)^2, (a-1)(b-1)\}$ there exist $s_1, \dots, s_m \in S$ such that

$$f(s_1, \dots, s_m) \equiv w_f(1 + s_1, \dots, 1 + s_m) - 1 \equiv v(x-1)^{r-1} \pmod{I_r}.$$

This goes by induction on r. If $r=1$, then $f(S)=S$, and using (2.2) we have

$$-(a-1)(a^{-1}-1)\equiv(a-1)^2 \pmod{\omega(KG)^3},$$
$$-(b-1)(b^{-1}-1)\equiv(b-1)^2 \pmod{\omega(KG)^3}$$

and

$$\begin{aligned}-(ab-1)((ab)^{-1}-1)&\equiv(ab-1)^2=((a-1)(b-1)+(a-1)+(b-1))^2\\ &\equiv(a-1)^2+(b-1)^2+2(a-1)(b-1) \pmod{\omega(KG)^3}.\end{aligned}$$

Hence,

$$\begin{aligned}2^{-1}((a-1)(a^{-1}-1)+(b-1)(b^{-1}-1)-(ab-1)((ab)^{-1}-1))\\ \equiv(a-1)(b-1) \pmod{\omega(KG)^3}.\end{aligned}$$

As $\omega(KG)^3\subseteq I_1$, the claim is true for $r=1$. Assume the claim for all Lie commutator identity of degree less than r, and let f be a multilinear Lie commutator of degree r. Then f can be expressed as a Lie commutator of the multilinear Lie commutators f_1 and f_2 of degree d and $r-d$, respectively. By the inductive hypothesis, for all $v_1,v_2\in V$ there exist $s_1,\dots,s_m\in S$ such that

$$f_1(s_1,\dots,s_m)\equiv w_{f_1}(1+s_1,\dots,1+s_m)-1\equiv v_1(x-1)^{d-1} \pmod{I_d},$$
$$f_2(s_1,\dots,s_m)\equiv w_{f_2}(1+s_1,\dots,1+s_m)-1\equiv v_2(x-1)^{r-d-1} \pmod{I_{r-d}}.$$

Now we can apply (2.1) and the equality

$$KG\omega(KG')^k=\omega(KG')^k+\omega(KG)\omega(KG')^k$$

which holds for any $k\geq 1$ to get that both $[I_s,I_t]$ and $[\omega^2(KG)\omega(KG')^{s-1},I_t]$ are subsets of I_{s+t} for any $s,t\geq 1$. Then

$$f_(s_1,\dots,s_m)\equiv[v_1(x-1)^{d-1},v_2(x-1)^{r-d-1}] \pmod{I_r},$$

furthermore,

$$\begin{aligned}&[v_1(x-1)^{d-1},v_2(x-1)^{r-d-1}]\\ &\quad=v_1[(x-1)^{d-1},v_2(x-1)^{r-d-1}]+[v_1,v_2(x-1)^{r-d-1}](x-1)^{d-1}\\ &\quad=v_1[(x-1)^{d-1},v_2](x-1)^{r-d-1}+[v_1,v_2](x-1)^{r-2},\end{aligned}$$

and by using the first relation of (2.1) we have

$$f(s_1,\dots,s_m)\equiv[v_1,v_2](x-1)^{r-2} \pmod{I_r}. \tag{2.3}$$

It remains to compute the Lie commutators $[v_1,v_2]$ for all possible v_1 and v_2. According to [1] (see p. 4911),

$$\begin{aligned}[(a-1)^2,(b-1)^2]&\equiv 4(a-1)(b-1)(x-1) \pmod{I_2},\\ [(a-1)^2,(a-1)(b-1)]&\equiv 2(a-1)^2(x-1) \pmod{I_2},\\ [(b-1)^2,(a-1)(b-1)]&\equiv 2(b-1)^2(x-1) \pmod{I_2}.\end{aligned} \tag{2.4}$$

For the sake of completeness, we confirm here the first congruence, the other two can be obtained similarly. Clearly,

$$\begin{aligned}[(a-1)^2,(b-1)^2] &= (a-1)[a,(b-1)^2]+[a,(b-1)^2](a-1)\\ &= (a-1)(b-1)[a,b]+(a-1)[a,b](b-1)\\ &\quad + (b-1)[a,b](a-1)+[a,b](b-1)(a-1).\end{aligned}$$

Furthermore, $[a,b]=ba(x-1)=(ba-1)(x-1)+(x-1)$ and $(g-1)(h-1)=(h-1)(g-1)+hg((g,h)-1)$ for any $g,h\in G$, so every summand on the right hand side is congruent to $(a-1)(b-1)(x-1)$ modulo I_2. This implies the required congruence.

So, by (2.4), for any $v\in V$ we can choose v_1 and v_2 such that

$$f(s_1,\dots,s_m)\equiv \alpha v(x-1)^{r-1} \pmod{I_r},$$

for some $\alpha\in K\setminus\{0\}$.

For the sake of brevity, we write $1+\underline{s}$ instead of $(1+s_1,\dots,1+s_m)$. Then

$$\begin{aligned} w_f(1+\underline{s}) &= (w_{f_1}(1+\underline{s}),w_{f_2}(1+\underline{s}))\\ &= 1+w_{f_1}(1+\underline{s})^{-1}w_{f_2}(1+\underline{s})^{-1}[w_{f_1}(1+\underline{s}),w_{f_2}(1+\underline{s})]\\ &\equiv 1+w_{f_1}(1+\underline{s})^{-1}w_{f_2}(1+\underline{s})^{-1}f(s_1,\dots,s_m)\\ &\equiv 1+\alpha v(x-1)^{r-1} \pmod{I_r}.\end{aligned}$$

Let k be an integer for which x_k divides the polynomial $f(x_1,\dots,x_m)$; let $s'_k=\alpha^{-1}s_k$, and $s'_i=s_i$ for all $i\neq k$. Then

$$f(s'_1,\dots,s'_m)\equiv w_f(1+s'_1,\dots,1+s'_m)-1\equiv v(x-1)^{r-1} \pmod{I_r},$$

and the induction is done.

Now, applying the results of [4] we show that $w=v(x-1)^{r-1}\notin I_r$ for $r=p^n$. Denote by t the weight of the element $x-1$. Then $t\geq 2$, and $w\in\omega(KG)^{2+t(r-1)}\setminus\omega(KG)^{3+t(r-1)}$. Since $\omega(KG)^i$ has a basis over K consisting of regular elements of weight not less than i, we have that $I_r=\omega(KG)^3\omega(KG')^{r-1}\subseteq\omega(KG)^{3+t(r-1)}$. Consequently, $w\notin I_r$. This means that $f(S)$ contains a nonzero element for any Lie commutator identity f of degree p^n or less. □

As every element of G has odd order, the orientation σ has to be trivial, so all elements of S belong to $(KG)_*$, further $1+S\subseteq U_*(KG)$. This implies the following statement.

Lemma 2.2. *Let K be a field of characteristic $p>2$, and let G be a finite p-group with cyclic commutator subgroup. Then*

(i) $(KG)_*$ satisfies no Lie commutator identity of degree less than $|G'|+1$;

(ii) $U_*(KG)$ satisfies no group commutator identity of degree less than $|G'|+1$.

Now, we are ready to prove our main theorem. We will use that the subspace $(KG)_*$ of KG is spanned by the set $\{g + \sigma(g)g^{-1} : g \in G\}$.

Proof of Theorem 1.1. Let $f(x_1, \dots, x_m)$ be a multilinear Lie commutator of degree less than $|G'| + 1$.

According to Theorem 1.7 of [10], the FC-group G is isomorphic to a subgroup of the direct product of the torsion FC-group G/A and the torsion-free abelian group G/T, where A is a maximal torsion free central subgroup, and T is the torsion part of G. Hence, $G' \cong (G/A)'$. Assume that $A \subseteq \ker\sigma$. Then the involution $*$ induces the involution

$$\left(\sum_{\overline{g}\in G/A} \alpha_g \overline{g}\right)^{\star} = \sum_{\overline{g}\in G/A} \alpha_g \sigma(g)\overline{g}^{-1},$$

on $K[G/A]$, which is also an oriented classical involution, and the elements of $(K[G/A])_\star$ are exactly the homomorphic images of the elements of $(KG)_*$ under the natural homomorphism $\varphi\colon KG \to K[G/A]$. Choose the elements $\overline{g}, \overline{h} \in G/A$ such that $(G/A)' = \langle(\overline{g}, \overline{h})\rangle$. As a finitely generated torsion nilpotent group, $H = \langle \overline{g}, \overline{h}\rangle$ is finite, and it is the direct product of its Sylow subgroups. Denote by P the Sylow p-subgroup of H. Since G' is a p-group, we have that $P' = H' \cong G'$. By applying (i) of Lemma 2.2 for the finite p-group P, we obtain that there exist elements $s_1, \dots, s_m \in (KG)_*$ such that $\varphi(s_1), \dots, \varphi(s_m) \in (KP)_\star$ and

$$f(\varphi(s_1), \dots, \varphi(s_m)) \neq 0.$$

Then $\varphi(f(s_1, \dots, s_m)) \neq 0$, and $f(s_1, \dots, s_m) \neq 0$, as desired.

In the remaining case when $A \not\subseteq \ker\sigma$, let us take an element a from $A \setminus \ker\sigma$. Then $G = \ker\sigma \cup a\ker\sigma$, and as a is central in G, it follows that $(\ker\sigma)' = G'$. Now we may repeat the proof to have that $(K\ker\sigma)_*$ does not satisfy f. Since $(K\ker\sigma)_* \subseteq (KG)_*$, the first part of the theorem is proved.

Assume that G is torsion, and denote by P the Sylow p-subgroup of the finite nilpotent group $H = \langle g, h\rangle$, where $g, h \in G$ such that $\langle(g, h)\rangle = G'$. Then $P' = G'$, and by (ii) of Lemma 2.2, $U_*(KP)$ satisfies no Lie commutator identity of degree less than $|G'| + 1$, so is $U_*(KG)$. □

References

[1] Bagiński, C., A note on the derived length of the unit group of a modular group algebra, *Comm. Algebra*, Vol. 30 (2002), no. 10, 4905–4913.

[2] Bovdi, V., Spinelli, E., Modular group algebras with maximal Lie nilpotency indices, *Publ. Math. (Debrecen)*, Vol. 65 (2004), no. 1-2, 243–252.

[3] Castillo, J.H., Polcino, C.M., Lie properties of symmetric elements under oriented involutions, *Comm. Algebra*, Vol. 40 (2012), no. 12, 4404–4419.

[4] Jennings, S.A., The structure of the group ring of a p-group over a modular field, *Trans. Amer. Math. Soc.*, Vol. 50 (1941), 175–185.

[5] Levin, F., Rosenberger, G., Lie metabelian group rings, *North-Holland Math. Stud.*, Vol. 126 (1986), 153–161 (to appear in Group and semigroup rings (Johannesburg, 1985)).

[6] Passi, I.B.S., Passman, D.S., Sehgal, S.K., Lie solvable group rings, *Canad. J. Math.*, Vol. 25 (1973), 748–757.

[7] Passman, D.S. Group rings satisfying a polynomial identity, *J. Algebra*, Vol. 20 (1972), 221–225.

[8] Sabinin, L., Sbitneva, L., Shestakov, I., Non-associative algebras and its applications, CRC Press, (2006)

[9] Sharma, R.K., Srivastava, J.B., Lie centrally metabelian group rings, *J. Algebra*, Vol. 151 (1992), 476–486.

[10] Tomkinson, M.J., FC-groups, Research Notes in Mathematics, Vol. 96, Pitman (Advanced Publishing Program), Boston, (1984).

Annales Mathematicae et Informaticae
43 *(2014) pp. 103–112*
http://ami.ektf.hu

Generalized binary recurrent quasi-cyclic matrices

E. Kılıç[a], Y. T. Ulutaş[b], I. Akkus[c*], N. Ömür[b]

[a]TOBB University of Economics and Technology Mathematics Department, Ankara, Turkey
ekilic@etu.edu.tr

[b]Kocaeli University Mathematics Department, Izmit Kocaeli, Turkey
turkery@kocaeli.edu.tr,neseomur@kocaeli.edu.tr

[c]Kırıkkale University, Faculty of Arts and Sciences, Department of Mathematics, Yahsihan, Kırıkkale, Turkey
iakkus.tr@gmail.com

Submitted February 1, 2014 — Accepted August 15, 2014

Abstract

In this paper, we obtain solutions to infinite family of Pell equations of higher degree based on the more generalized Fibonacci and Lucas sequences as well as their all subsequences of the form $\{u_{kn}\}$ and $\{v_{kn}\}$ for odd $k > 0$.

Keywords: Quasi-cyclic matrices, binary linear recurrences, Pell equation.

MSC: 11B37, 15A15.

1. Introduction

The generalized Fibonacci and Lucas sequences are defined by

$$u_{n+1} = Au_n + Bu_{n-1} \tag{1.1}$$

and

$$v_{n+1} = Av_n + Bv_{n-1}, \tag{1.2}$$

*Corresponding author

where $u_0 = 0$, $u_1 = 1$ and $v_0 = 2$, $v_1 = A$, respectively.

For $k \geq 0$ and $n > 1$, the sequences $\{u_{kn}\}$ and $\{v_{kn}\}$ satisfy the recursions (see [1]):

$$u_{kn} = v_k u_{k(n-1)} - (-B)^k u_{k(n-2)} \text{ and } v_{kn} = v_k v_{k(n-1)} - (-B)^k v_{k(n-2)}. \quad (1.3)$$

The Binet formulae are

$$u_n = \frac{\alpha^n - \beta^n}{\alpha - \beta} \text{ and } v_n = \alpha^n + \beta^n,$$

where $\alpha, \beta = A \pm \sqrt{A^2 + 4B}$.

By the Binet formulae note that for a fixed $k > 0$,

$$u_{-kn} = (-1)^{kn+1} u_{kn} \text{ and } u_{2kn} = v_{kn} u_{kn}. \quad (1.4)$$

A $n \times n$ quasi-cyclic matrix $R(D; x_1, x_2, ... x_n)$ (or shortly R) has the form (see [2, 4, 5]):

$$R = \begin{pmatrix} x_1 & Dx_n & Dx_{n-1} & \cdots & Dx_3 & Dx_2 \\ x_2 & x_1 & Dx_n & \cdots & Dx_4 & Dx_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n-1} & x_{n-2} & x_{n-3} & \cdots & x_1 & Dx_n \\ x_n & x_{n-1} & x_{n-2} & \cdots & x_2 & x_1 \end{pmatrix}.$$

The classical Pell equation $x^2 - dy^2 = \pm 1$ $(d \in Z)$ can be rewritten as

$$\det \begin{pmatrix} x & dy \\ y & x \end{pmatrix} = \pm 1.$$

By means of quasi-cyclic determinants, the equation

$$\det \begin{pmatrix} x_1 & Dx_n & Dx_{n-1} & \cdots & Dx_3 & Dx_2 \\ x_2 & x_1 & Dx_n & \cdots & Dx_4 & Dx_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n-1} & x_{n-2} & x_{n-3} & \cdots & x_1 & Dx_n \\ x_n & x_{n-1} & x_{n-2} & \cdots & x_2 & x_1 \end{pmatrix} = \pm 1$$

is called Pell's equation of degree n.

In [2], the author gave a method to generalize the classical Pell equation whose degree is $n = 2$ to a Pell equation of degree $n \geq 2$ by some $n \times n$ quasi-cyclic determinants. In particular, the author proved that for $n \geq 2$,

$$\det (R(L_n; F_{2n-1}, F_{2n-2}, ..., F_n)) = 1, \quad (1.5)$$

where L_n and F_n denote the nth Lucas and Fibonacci number, respectively. Further it was showed that

$$\det (R(L_n; F_{2n-1+k}, F_{2n-2+k}, ..., F_{n+k})) = (-1)^{n-1} L_n F_k^n + F_{k-1}^n,$$

where k is an integer.

In [3], the author generalized the results given in [2] by giving a relationship between certain Pell equations of degree n and general Fibonacci and Lucas sequences. For example, for $k = 1$ in (1.3) and (1.4) and $n > 1$, we have

$$\det\left(R\left(v_n; u_{2n-1}, u_{2n-2}, ..., u_n\right)\right) = B^{n(n-1)}, \tag{1.6}$$

where B is defined as before.

From [4, 5], the following two propositions are known:

Proposition 1. *For $n > 0$,*

$$\det\left(R\right) = \prod_{k=0}^{n-1}\left(\sum_{i=1}^{n} x_i d^{i-1}\varepsilon^{k(i-1)}\right), \tag{1.7}$$

where $d = \sqrt[n]{D}$, $\varepsilon = e^{2\pi i/n}$ and each factor $\sum_{i=1}^{n} x_i d^{i-1}\varepsilon^{k(i-1)}$ of the RHS of (1.7) is an eigenvalue of the matrix R.

Proposition 2. *Let n and D be fixed. Then the sum, differences, and product of two quasi-cyclic matrices is also quasi-cyclic. The inverse of a quasi-cyclic matrix is quasi-cyclic.*

In this paper, we generalize the results of [2, 3] and so obtain solutions to infinite family of Pell equations of higher degree based on more generalized Fibonacci and Lucas sequences as well as their all subsequences of the form $\{u_{kn}\}$ and $\{v_{kn}\}$, for odd $k > 0$.

2. Quasi-cyclic matrices via the generalized Fibonacci and Lucas numbers

We obtain some results about infinite family of Pell equations of higher degree by using certain quasi-cyclic determinants with the generalized Fibonacci and Lucas numbers. We give some auxiliary results for further use and denote $(-B)^k$ by b for easy writing.

Lemma 2.1. *For positive integers k and n,*

$$\begin{aligned} v_k u_{k(2n-1)} - v_{kn} u_{kn} &= b u_{k(2n-2)}, \\ b\left(u_{k(2n-1)} - v_{kn} u_{k(n-1)}\right) &= b^n u_k, \\ u_{kn}^2 - u_{k(n+1)} u_{k(n-1)} &= b^{(n-1)} u_k^2. \end{aligned}$$

Proof. The claimed identities follows from the Binet formulae. □

Theorem 2.2. *For $n \geq 2$,*

$$\det\left(R\left(v_{kn}; u_{k(2n-1)}, u_{k(2n-2)}, ..., u_{kn}\right)\right) = b^{n(n-1)} u_k^n. \tag{2.1}$$

Proof. For $n=2$,

$$\det\left(R\left(v_{2k};u_{3k},u_{2k}\right)\right)=\begin{vmatrix} u_{3k} & v_{2k}u_{2k} \\ u_{2k} & u_{3k} \end{vmatrix}=u_{3k}^2-v_{2k}u_{2k}^2=b^2u_k^2.$$

For $n>2$, consider the upper triangular matrix

$$T=\begin{pmatrix} 1 & -v_k & b & & 0 \\ & 1 & -v_k & \ddots & \\ & & \ddots & \ddots & b \\ & & & 1 & -v_k \\ & & & & 1 \end{pmatrix}. \tag{2.2}$$

From a matrix multiplication and by Lemma 2.1, we get

$$RT=\begin{pmatrix} u_{k(2n-1)} & -bu_{k(2n-2)} & b^nu_k & 0 & \dots & 0 \\ u_{k(2n-2)} & -bu_{k(2n-3)} & 0 & b^nu_k & \ddots & \vdots \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & b^nu_k \\ u_{k(n+1)} & -bu_{kn} & 0 & 0 & \dots & 0 \\ u_{kn} & -bu_{k(n-1)} & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{2.3}$$

Then we write

$$\begin{aligned} \det R &= (\det R)(\det T)=\det(RT) \\ &= \left(bu_{kn}^2-bu_{k(n+1)}u_{k(n-1)}\right)\det\begin{pmatrix} b^nu_k & 0 & \cdots & 0 \\ 0 & b^nu_k & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b^nu_k \end{pmatrix} \\ &= \left(bu_{kn}^2-bu_{k(n+1)}u_{k(n-1)}\right)\left(b^nu_k\right)^{n-2} \\ &= b^{n(n-1)}u_k^n, \end{aligned}$$

as claimed. $\square$

Corollary 2.3. *For $n\geq 2$,*

$$\prod_{k=0}^{n-1}\left(\sum_{j=1}^{n}u_{k(2n-j)}\left(\sqrt[n]{v_{kn}}\right)^{j-1}\varepsilon^{k(j-1)}\right)=b^{n(n-1)}u_k^n,$$

where $\sqrt[n]{v_{kn}}$ is the nth complex root of v_{kn} and $\varepsilon=e^{2\pi i/n}$.

We shall need the following identities:

1. $-bu_{k(2n-3)} + v_k u_{k(2n-2)} - u_{k(2n-1)} = 0, ..., -bu_{kn} + v_k u_{k(n+1)} - u_{k(n+2)} = 0,$
2. $u_{k(2n-1)} - v_{kn} u_{k(n-1)} = b^{n-1} u_k,$
3. $E_n^{n+1} = v_{kn} E_n$ and $E_n^n = v_{kn} I_n$, where

$$E_n = \begin{pmatrix} 0 & 0 & \cdots & 0 & v_{kn} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Theorem 2.4. *For $n \geq 3$, the matrix $R\left(v_{kn}; u_{k(2n-1)}, u_{k(2n-2)}, ..., u_{kn}\right)$ is invertible and its inverse matrix R^{-1} is given by*

$$R^{-1}\left(v_{kn}; u_{k(2n-1)}, u_{k(2n-2)}, ..., u_{kn}\right) = -\frac{1}{u_k b^n}\left(-bI_n + v_k E_n - E_n^2\right), \tag{2.4}$$

where I_n is the $n \times n$ identity matrix and the matrix E_n is defined as before.

Proof. Since $\det\left(R\left(v_{kn}; u_{k(2n-1)}, u_{k(2n-2)}, ..., u_{kn}\right)\right) \neq 0$ by Theorem 2.2, its inverse exists. It is easy to see that

$$R\left(v_{kn}; u_{k(2n-1)}, u_{k(2n-2)}, ..., u_{kn}\right) = \left(u_{k(2n-1)} I_n + u_{k(2n-2)} E_n + ... + u_{kn} E_n^{n-1}\right).$$

Hence,

$$\begin{aligned} & R\left(v_{kn}; u_{k(2n-1)}, u_{k(2n-2)}, ..., u_{kn}\right) R^{-1}\left(v_{kn}; u_{k(2n-1)}, u_{k(2n-2)}, ..., u_{kn}\right) \\ &= \left(u_{k(2n-1)} I_n + u_{k(2n-2)} E_n + ... + u_{kn} E_n^{n-1}\right)\left(\frac{-1}{u_k b^n}\right)\left(-\left(-B\right)^k I_n + v_k E_n - E_n^2\right) \\ &= \left(-bu_{k(2n-1)} I_n + \left(u_{2kn} - u_{kn} v_{kn}\right) E_n + \left(v_k u_{kn} - u_{k(n+1)}\right) v_{kn} I_n\right)\left(\frac{-1}{u_k b^n}\right) \\ &= -b\left(u_{k(2n-1)} - v_{kn} u_{k(n-1)}\right) I_n \left(\frac{-1}{u_k b^n}\right) \\ &= -b\left(b^{(n-1)} u_k\right) I_n \left(\frac{-1}{u_k b^n}\right) = I_n, \end{aligned}$$

as claimed. ☐

3. The determinants of quasi-cyclic matrices

For all integer t, define the $n \times n$ quasi-cyclic matrix $R_{k,n,t}$ as

$$R_{k,n,t} = R\left(v_{kn}; u_{k(2n-1+t)}, u_{k(2n-2+t)}, ..., u_{k(n+t)}\right).$$

By Theorem 2.2, we have

$$\det\left(R_{k,n,0}\right) = b^{n(n-1)} u_k^n.$$

For $\det R_{k,n,1}$, $\det R_{k,n,2}$,..., $\det R_{k,n,-1}$, $\det R_{k,n,-2}$,..., we can obtain corresponding results.

Define the $n \times n$ matrices $g_{k,n,t}$ and $h_{k,n,t}$ as shown:

$$g_{k,n,t} = \begin{pmatrix} u_{k(2n+t-1)} & -bu_{k(2n+t-2)} & -b^{n+1}u_{k(t-1)} & & 0 \\ u_{k(2n+t-2)} & -bu_{k(2n+t-3)} & b^n u_{kt} & \ddots & \\ \vdots & \vdots & 0 & \ddots & -b^{n+1}u_{k(t-1)} \\ u_{k(n+t+1)} & -bu_{k(n+t)} & \vdots & \ddots & b^n u_{kt} \\ u_{k(n+t)} & -bu_{k(n+t-1)} & 0 & \dots & 0 \end{pmatrix}$$

and

$$h_{k,n,t} = \begin{pmatrix} u_{k(2n+t-1)} & b^n u_{kt} & -b^{n+1}u_{k(t-1)} & & & 0 \\ u_{k(2n+t-2)} & 0 & b^n u_{kt} & -b^{n+1}u_{k(t-1)} & & \\ \vdots & \vdots & 0 & b^n u_{kt} & \ddots & \\ \vdots & \vdots & \vdots & 0 & \ddots & -b^{n+1}u_{k(t-1)} \\ u_{k(n+t+1)} & 0 & 0 & \dots & \ddots & b^n u_{kt} \\ u_{k(n+t)} & 0 & 0 & \dots & \dots & 0 \end{pmatrix}.$$

We give some auxiliary Lemmas before the proof of main Theorem.

Lemma 3.1. *(The recurrence of* $\det g_{k,n,t}$*)*

$$\det g_{k,n,t} = (-1)^n \, b^{\left(n^2-n+t\right)} u_k u_{k(n-1)} u_{kt}^{n-2} - b^{(2n-1)} u_{k(t-1)} \det g_{k,n-1,t}. \qquad (3.1)$$

Proof. Clearly

$$\det g_{k,n,t}$$

$$= -b^{n(n-2)+1} \begin{vmatrix} u_{k(2n+t-1)} & u_{k(2n+t-2)} & -bu_{k(t-1)} & 0 & \dots & 0 \\ u_{k(2n+t-2)} & u_{k(2n+t-3)} & u_{kt} & -bu_{k(t-1)} & \ddots & \vdots \\ \vdots & \vdots & 0 & u_{kt} & \ddots & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & -bu_{k(t-1)} \\ u_{k(n+t+1)} & u_{k(n+t)} & \vdots & \vdots & \ddots & u_{kt} \\ u_{k(n+t)} & u_{k(n+t-1)} & 0 & \dots & \dots & 0 \end{vmatrix}.$$

By subtracting the second column of $g_{k,n,t}$ from the first column by multiplying

v_k gives us

$$\det g_{k,n,t} = -b^{n(n-2)+1} \begin{vmatrix} bu_{k(2n+t-3)} & u_{k(2n+t-2)} & -bu_{k(t-1)} & 0 & \cdots & 0 \\ bu_{k(2n+t-4)} & u_{k(2n+t-3)} & u_{kt} & -bu_{k(t-1)} & \ddots & \vdots \\ \vdots & \vdots & 0 & u_{kt} & \ddots & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & -bu_{k(t-1)} \\ u_{k(n+t-1)} & u_{k(n+t)} & \vdots & \vdots & \ddots & u_{kt} \\ bu_{k(n+t-2)} & u_{k(n+t-1)} & 0 & \cdots & \cdots & 0 \end{vmatrix}.$$

So on after $n+t-1$ subtractions between the two columns, we get finally

$$\det g_{k,n,t} = -b^{n(n-2)+n+t} \begin{vmatrix} u_{kn} & u_{k(n-1)} & -bu_{k(t-1)} & 0 & \cdots & 0 \\ u_{k(n-1)} & u_{k(n-2)} & u_{kt} & -bu_{k(t-1)} & \ddots & \vdots \\ \vdots & \vdots & 0 & u_{kt} & \ddots & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & -bu_{k(t-1)} \\ u_{2k} & u_1 & \vdots & \vdots & \ddots & u_{kt} \\ u_k & u_0 & 0 & \cdots & \cdots & 0 \end{vmatrix}.$$

Expanding the determinant above with respect to the first row and by $u_0 = 0$, we get

$$\det g_{k,n,t} = b^{(n^2-n+t)} u_{k(n-1)} \begin{vmatrix} u_{k(n-1)} & u_{kt} & \cdots & \vdots \\ \vdots & 0 & \cdots & -bu_{k(t-1)} \\ \vdots & \vdots & \cdots & u_{kt} \\ u_k & 0 & \cdots & 0 \end{vmatrix}$$

$$+ b^{n^2-n+t+1} u_{k(t-1)} \begin{vmatrix} u_{k(n-1)} & u_{k(n-2)} & -bu_{k(t-1)} & 0 & 0 \\ u_{k(n-2)} & u_{k(n-3)} & u_{kt} & \cdots & \vdots \\ \vdots & \vdots & 0 & \cdots & -bu_{k(t-1)} \\ \vdots & \vdots & \vdots & \cdots & u_{kt} \\ u_k & u_0 & 0 & \cdots & 0 \end{vmatrix}$$

$$= (-1)^n b^{n^2-n+t} u_{k(n-1)} u_k u_{kt}^{n-2} + b^{n^2-n+t+1} u_{k(t-1)} \left(\frac{-1}{b^{n^2-3n+t+2}} \right) \det g_{k,n-1,t}$$

$$= (-1)^n b^{n^2-n+t} u_{k(n-1)} u_k u_{kt}^{n-2} - b^{2n-1} u_{k(t-1)} \det g_{k,n-1,t}.$$

Thus we have the conclusion. □

Lemma 3.2. *For odd $k > 0$,*

$$\det g_{k,n,t} = \frac{(-1)^{kn}}{u_k}\left[b^{n^2-n+1}u_{k(n-1)}u_{kt}^n + b^{n^2}u_{k(t-1)}^n u_k - b^{n^2-n+1}u_{k(t-1)}u_{kn}u_{kt}^{n-1}\right] \tag{3.2}$$

Proof. (Induction on n) When $n = 2$, we have

$$\det g_{k,2,t} = \begin{vmatrix} u_{(3+t)} & -bu_{(2+t)} \\ u_{(2+t)} & -bu_{(1+t)} \end{vmatrix} = -b\left(u_{(3+t)}u_{(1+t)} - u_{(2+t)}^2\right) = b^{t+2}u_k^2.$$

Substituting $n = 2$ in the RHS of (3.2), we get

$$\begin{aligned} &\frac{(-1)^{2k}}{u_k}\left[b^3 u_k u_{kt}^2 + b^4 u_{k(t-1)}^2 u_k - b^3 u_{k(t-1)}u_{2k}u_{kt}\right] \\ &= b^3\left(u_{kt}^2 + bu_{k(t-1)}^2 - u_{k(t-1)}v_k u_{kt}\right) \\ &= b^3\left(u_{kt}^2 - u_{k(t+1)}u_{k(t-1)}\right) = b^{t+2}u_k^2, \end{aligned}$$

as claimed. We assume that the claim is true for $n-1$. Now we prove that the claim is true for n. By the induction hypothesis and (3.1), we write for odd integer k,

$$\begin{aligned} &\det g_{k,n,t} \\ &= (-1)^n b^{n^2-n+t}u_{k(n-1)}u_k u_{kt}^{n-2} - b^{2n-1}u_{k(t-1)}\frac{(-1)^{k(n-1)}}{u_k} \\ &\times\left[b^{n^2-3n+3}u_{k(n-2)}u_{kt}^{n-1} + b^{(n-1)^2}u_{k(t-1)}^{n-1}u_k - b^{n^2-3n+3}u_{k(t-1)}u_{k(n-1)}u_{kt}^{n-2}\right] \\ &= (-1)^{k(n-1)+1}b^{n^2}u_{k(t-1)}^n + (-1)^{k(n-1)}b^{n^2-n+1}\frac{u_{k(t-1)}u_{kt}^{n-1}u_{kn}}{u_k} + \\ &+ u_{kt}^{n-2}u_{k(n-1)}\left[(-1)^{kn}b^{n^2-n+t}u_k - (-1)^{k(n-1)}b^{n^2-n+1}\frac{u_{k(t+1)}u_{k(t-1)}}{u_k}\right] \\ &= (-1)^{k(n-1)+1}b^{n^2}u_{k(t-1)}^n + (-1)^{k(n-1)}b^{n^2-n+1}\frac{u_{k(t-1)}u_{kt}^{n-1}u_{kn}}{u_k} + \\ &+ (-1)^{kn}b^{n^2-n+1}\frac{u_{kt}^{n-2}u_{k(n-1)}}{u_k}\left[b^{t-1}u_k^2 + u_{k(t+1)}u_{k(t-1)}\right] \\ &= \frac{(-1)^{kn}}{u_k}\left[b^{n^2-n+1}u_{k(n-1)}u_{kt}^n + b^{n^2}u_{k(t-1)}^n u_k - b^{n^2-n+1}u_{k(t-1)}u_{kn}u_{kt}^{n-1}\right]. \end{aligned}$$

Thus the proof is complete. □

Lemma 3.3. *For $n > 1$,*

$$\det h_{k,n,t} = (-1)^{n+1}b^{n(n-1)}u_{k(n+t)}u_{kt}^{n-1}.$$

Proof. Expanding $\det h_{k,n,t}$ with respect to the last row gives us

$$
\begin{aligned}
&\det h_{k,n,t}\\
&= \begin{vmatrix}
u_{k(2n+t-1)} & b^n u_{kt} & -b^{n+1}u_{k(t-1)} & 0 & \cdots & 0 \\
u_{k(2n+t-2)} & 0 & b^n u_{kt} & -b^{n+1}u_{k(t-1)} & \ddots & \vdots \\
\vdots & \vdots & 0 & b^n u_{kt} & \ddots & 0 \\
\vdots & \vdots & \vdots & 0 & \ddots & -b^{n+1}u_{k(t-1)} \\
u_{k(n+t+1)} & 0 & \vdots & \vdots & \ddots & b^n u_{kt} \\
u_{k(n+t)} & 0 & 0 & \cdots & \cdots & 0
\end{vmatrix}\\
&= u_{k(n+t)} \left(-1\right)^{n+1} \left(b^n u_{kt}\right)^{n-1}\\
&= \left(-1\right)^{n+1} b^{n(n-1)} u_{k(n+t)} u_{kt}^{n-1},
\end{aligned}
$$

as claimed. □

Lemma 3.4. *For $n > 1$ and $k, t > 0$,*

$$
\begin{aligned}
v_{kn} &= \left(v_k u_{kn} - 2b u_{k(n-1)}\right)/u_k,\\
u_{k(n+t)} &= \left(u_{k(n+1)} u_{kt} - b u_{kn} u_{k(t-1)}\right)/u_k.
\end{aligned}
$$

Proof. The claims are obtained from the Binet formulae of $\{u_n\}$ and $\{v_n\}$. □

Theorem 3.5. *For $n \geq 2$ and all integer t,*

$$
\det R_{k,n,t} = b^{n(n-1)} \left(\left(-1\right)^{kn-1} v_{kn} u_{kt}^n + \left(-1\right)^{kn} b^n u_{k(t-1)}^n \right), \tag{3.3}
$$

where k is an odd integer.

Proof. From the definitions of $g_{k,n,t}$ and $h_{k,n,t}$, we see that

$$
\det R_{k,n,t} = \det g_{k,n,t} + \det h_{k,n,t}.
$$

So the proof follows from Lemmas 3.2, 3.3 and 3.4. □

When $t = n$ in (3.2) and (3.3), we have the following result.

Corollary 3.6. *For $n > 1$,*

$$
\begin{aligned}
\det g_{k,n,n} &= \left(-1\right)^{kn} b^{n^2} u_{k(n-1)}^n,\\
\det R_{k,n,n} &= \left(-1\right)^{kn} b^{n(n-1)} \left(-v_{kn} u_{kn}^n + b^n u_{k(n-1)}^n\right).
\end{aligned}
$$

References

[1] Kilic, E., Stanica, P., Factorizations and representations of second linear recurrences with indices in arithmetic progressions, *Bul. Mex. Math. Soc.*, Vol. 15(1) (2009), 23–36.

[2] Dazheng, L., Fibonacci-Lucas quasi-cyclic matrices, *The Fibonacci Quarterly*, Vol. 40(2) (2002), 280–286.

[3] Matyas, F., Second order linear recurrences and Pell's equations of higher degree, *Acta Acad. Paed. Agriensis, Sectio Mathematicae*, Vol. 29 (2002), 47–53.

[4] Guangxing, S., On eigenvalues of some cyclic matrices, *Math. Application*, Vol. 4(3) (1991), 76–82.

[5] Taiming, T., Dazheng, L., Diophantine approximation circulant matrix and Pell equation, *J. Shaansi Normal Univ.*, Vol. 28(14) (2000), 6–11.

Annales Mathematicae et Informaticae
43 (2014) pp. 113–121
http://ami.ektf.hu

Coincidences in numbers of graph vertices corresponding to regular planar hyperbolic mosaics

László Németh, László Szalay

Institute of Mathematics, University of West Hungary
nemeth.laszlo@emk.nyme.hu
szalay.laszlo@emk.nyme.hu

Submitted September 02, 2014 — Accepted October 20, 2014

Abstract

The aim of this paper is to determine the elements which are in two pairs of sequences linked to the regular mosaics $\{4,5\}$ and $\{p,q\}$ on the hyperbolic plane. The problem leads to the solution of diophantine equations of certain types.

Keywords: regular planar hyperbolic mosaics, linear recurrences, diophantine equations.

MSC: 11B37, 51M10.

1. Introduction

Consider a regular mosaic on the hyperbolic plane. Such a mosaic is characterized by the Schläfli's symbol $\{p,q\}$. It is known that we can define belts of cells around a given vertex of the mosaic (see [4]). Let's say that belt $\mathcal{B}_0$ is the aforesaid fixed vertex itself denoted by B_0. The first belt $\mathcal{B}_1$ consists of the cells which connect to B_0. Assume now that the belts $\mathcal{B}_{i-1}$ and $\mathcal{B}_i$ are known ($i \geq 1$). Let belt $\mathcal{B}_{i+1}$ be created by the cells that have common point (not necessarily common vertex) with $\mathcal{B}_i$, but not with $\mathcal{B}_{i-1}$. Figure 1 shows the first three belts in the mosaic corresponding to $\{4,5\}$. One important question is to study the phenomenon of the growing of belts ([1], [2], [3]), even in higher dimensions, too.

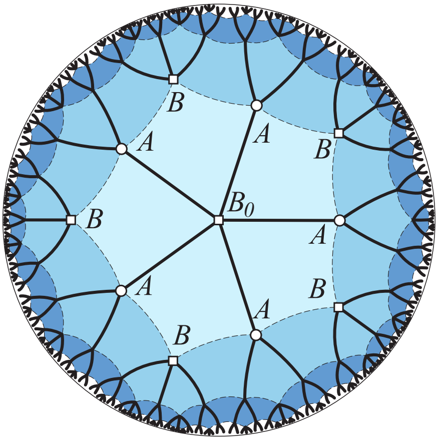


Figure 1: *Trees of the mosaic* $\{p, q\} = \{4, 5\}$

Take vertex B_0 as a main root of a will-be-graph (this is level 0). In general, let the outer boundary of belt $\mathcal{B}_i$ be called level i. Connect the vertices of level 1 to B_0 along the edges between the two levels of the lattice. By this way we have started to build trees. Then use always the maximum number of edges between level $(i-1)$ and level i. All vertices on level i are connected to only one vertex of the previous level, such that no unconnected leaves on level $(i-1)$ are remained. We never connect edges on the same level. The rest vertices on layer i will be roots of new trees. In this way, we obtain infinitely many trees, each of them contains infinitely many vertices. Let $\bar{A}$ denote the set of roots and $\bar{B}$ the set of other vertices. In Figure 1 and 2 the thick edges show the trees from level 0 to level 4. (We remark, that the dual problem is when we establish trees by connecting the centres of the cells of the mosaic.)

The case $q = 3$ provides no any tree since only one edge is not enough to connect the consecutive levels. If $p = 3$ the algorithm, apart from B_0, does not give roots. Therefore we may assume $p \geq 4$, $q \geq 4$, and since $(p-2)(q-2) = 4$ is the Euclidean lattice we also suppose $(p-2)(q-2) > 4$.

Let a_i and b_i denote the number of the vertices of $\bar{A}$ and $\bar{B}$ on level i, respectively. In this paper, we compare the terms a_i (and later b_i) of sequences belonging to different Schläfli's symbols $\{p, q\}$.

In the following, we recall some properties of the sequences a_i and b_i corresponding to hyperbolic planar lattice $\{p, q\}$ (see [4]). Simple geometric consideration shows $a_1 = q$, $b_1 = (p-3)q$, further the recursive system

$$a_n = (q-3)a_{n-1} + (q-2)b_{n-1}, \tag{1.1}$$

$$b_n = \big((q-3)(p-3)-1\big)a_{n-1} + \big((q-2)(p-3)-1\big)b_{n-1} \tag{1.2}$$

holds $(n \geq 2, p \geq 4, q \geq 4)$.

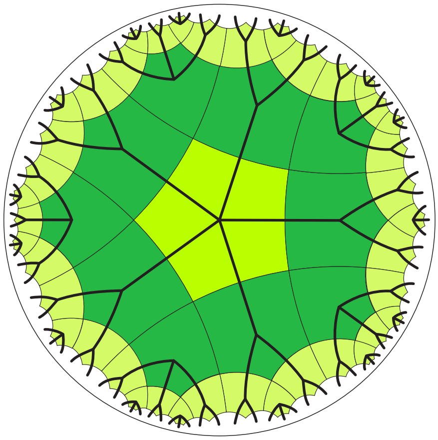


Figure 2: *Trees of the mosaic* $\{5,4\}$*, dual of mosaic* $\{4,5\}$

It is easy to separate the sequences $\{a_n\}$ and $\{b_n\}$, and it turns out that

$$a_n = \kappa a_{n-1} - a_{n-2} \qquad \text{and} \qquad b_n = \kappa b_{n-1} - b_{n-2}, \tag{1.3}$$

where $\kappa = (p-2)(q-2)-2$ $(\kappa \geq 4)$. Thus both sequences satisfy the same recurrence relation of order two, and they differ in their initials values. Indeed, to use (1.3) we need also the terms a_2 and b_2. Obviously, by (1.1) and (1.2), $a_2 = (\kappa+1)q$, $b_2 = (\kappa(p-3)-1)q$, and $(a_1, a_2) \neq (b_1, b_2)$. Later we also use the term $a_3 = (\kappa^2+\kappa-1)q$. Although a_0 and b_0 have no geometrical meaning, (1.3) provides the values $a_0 = -q$, $b_0 = q$, and this sometimes makes the calculations easier.

To achieve the investigations, we introduce the sufficient notations and recall some facts from the theory of linear recurrences. In general, let r and s denote arbitrary complex numbers. The sequence $\{G\}_{n=0}^{\infty}$ given by the initial values $G_0 \in \mathbb{C}$ and $G_1 \in \mathbb{C}$, and by the recursive relation

$$G_n = rG_{n-1} + sG_{n-2} \qquad (n \geq 2), \tag{1.4}$$

is called binary recurrence. For brevity, we often write $G(r, s, G_0, G_1)$ to indicate the parameters of the sequence $\{G\}$.

For any binary recurrence $G(r, s, G_0, G_1)$, the associate sequence of $\{G\}$ is the sequence $H(r, s, H_0, H_1)$ with

$$H_0 = 2G_1 - rG_0 \qquad \text{and} \qquad H_1 = rG_1 + 2sG_0. \tag{1.5}$$

Put $C_G = G_1^2 - rG_0G_1 - sG_0^2$. It is known that the terms of a binary recurrence $\{G\}$ and its associate sequence $\{H\}$ satisfy the equality

$$H_n^2 - DG_n^2 = 4C_G(-s)^n, \tag{1.6}$$

where $D = r^2 + 4s$.

2. Preparation and results

By (1.3) it follows that the coefficients of the investigated linear recurrences are $r = \kappa$ and $s = -1$. Thus $D = \kappa^2 - 4$, moreover

$$C_a = a_1^2 - ra_0a_1 - sa_0^2 = (\kappa + 2)q^2$$

and

$$C_b = b_1^2 - rb_0b_1 - sb_0^2 = ((p-3)^2 - \kappa(p-3) + 1)q^2.$$

Now we fix a mosaic given by $\{\tilde{p}, \tilde{q}\} = \{4, 5\}$. Then $\tilde{\kappa} = 4$, $\tilde{a}_n = 4\tilde{a}_{n-1} - \tilde{a}_{n-2}$, $\tilde{a}_1 = 5$, $\tilde{a}_2 = 25$, and $\tilde{b}_n = 4\tilde{b}_{n-1} - \tilde{b}_{n-2}$, $\tilde{b}_1 = 5$, $\tilde{b}_2 = 15$, moreover $\tilde{D} = 12$. The first ten terms of the sequences are given by the following table.

i	1	2	3	4	5	6	7	8	9	10
$\tilde{a}_i$	5	25	95	355	1325	4945	18455	68875	257045	959305
$\tilde{b}_i$	5	15	55	205	765	2855	10655	39765	148405	553855

Table 1: *Numbers of leaves and roots on level i connected with mosaic* $\{4, 5\}$

The associate sequences of $\{\tilde{a}_n\}$ and $\{\tilde{b}_n\}$ satisfy

$$\tilde{A}_n = 4\tilde{A}_{n-1} - \tilde{A}_{n-2} \quad \text{with} \quad \tilde{A}_1 = 30,\ \tilde{A}_2 = 90, \tag{2.1}$$
$$\tilde{B}_n = 4\tilde{B}_{n-1} - \tilde{B}_{n-2} \quad \text{with} \quad \tilde{B}_1 = 10,\ \tilde{B}_2 = 50, \tag{2.2}$$

respectively. Since $C_{\tilde{a}} = 150$, $C_{\tilde{b}} = -50$, by (1.6) we obtain the identities

$$\tilde{A}_n^2 - 12\tilde{a}_n^2 = 600 \qquad \text{and} \qquad \tilde{B}_n^2 - 12\tilde{b}_n^2 = -200. \tag{2.3}$$

In this paper, we target to solve

I. the diophantine equation $a_k = \tilde{a}_\ell$ in k and ℓ for certain mosaics $\{p, q\}$ (Section 3); further

II. the equations $a_\varepsilon = \tilde{a}_\ell$ in ℓ if $\varepsilon \in \{1, 2, 3\}$ and one of p and q is fixed (Section 4 and 5).

For the sequence $\{b_n\}$ analogous problems are examined.

The first question leads to simultaneous Pellian equations. The second problem requires different approaches depending on ε and the sequence $\{a_n\}$ (or $\{b_n\}$).

The observations are contained in the following theorems and Result 2.2. We always assume that

$$\{p, q\} \neq \{4, 4\}, \{4, 5\}.$$

Theorem 2.1. (1) *Let* $4 \le p \le 25$ *and* $4 \le q \le 18$. *Then the equation* $a_k = \tilde{a}_\ell$ *has only the trivial solution* $a_1 = \tilde{a}_1 = 5$ *for* $q = 5$ *and any* p.

(2) *If* $4 \le p, q \le 10$, *or* $11 \le p \le 25$ *and* $4 \le q \le 8$, *then the equation* $b_k = \tilde{b}_\ell$ *possesses only the solutions*

- $\{p, q\} = \{6, 5\}$, $b_1 = \tilde{b}_2 = 15$,
- $\{p, q\} = \{10, 5\}$, $b_2 = \tilde{b}_5 = 765$,
- $\{p, q\} = \{14, 5\}$, $b_1 = \tilde{b}_3 = 55$.

Result 2.2. (1) *If* $4 \le p \le 1\,600$, *then* $a_2 = \tilde{a}_\ell$ *is satisfied by*

- $\{p, q\} = \{26, 5\}$, $a_2 = \tilde{a}_4 = 335$,
- $\{p, q\} = \{90, 29\}$, $a_2 = \tilde{a}_8 = 68\,875$,
- $\{p, q\} = \{332, 5\}$, $a_2 = \tilde{a}_6 = 4\,945$,

(2) *In case of* $4 \le q \le 10\,000$, $a_3 = \tilde{a}_\ell$ *has no non-trivial small solution (i.e.* $p \le 10\,000$*).*

(3) *Assume* $4 \le p \le 10\,000$ *or* $4 \le q \le 2\,800$. *Then* $\{p, q\} = \{10, 5\}$, $b_2 = \tilde{b}_5 = 765$ *satisfy the equation* $b_2 = \tilde{b}_\ell$.

Theorem 2.3. (1) *All the solutions to* $a_2 = \tilde{a}_\ell$, *with* $5 \le q \le 25$ *are given by*

- $q = 5$, $\ell = 2 + 2t$ $(t \in \mathbb{N}^+)$,
- $q = 19$, $\ell = 58 + 90t$ *and* $\ell = 78 + 90t$ $(t \in \mathbb{N})$,
- $q = 23$, $\ell = 28 + 88t$ $(t \in \mathbb{N})$,
- $q = 25$, $\ell = 32 + 33t$ $(t \in \mathbb{N})$.

(2) *All the solutions to* $b_1 = \tilde{b}_\ell$, *with* $5 \le q \le 25$ *are given by*

- $q = 9$, $\ell = 5 + 18t$ *and* $\ell = 14 + 18t$ $(t \in \mathbb{N})$,
- $q = 11$, $\ell = 3 + 10t$ *and* $\ell = 8 + 10t$ $(t \in \mathbb{N})$,
- $q = 15$, $\ell = 2 + 6t$ *and* $\ell = 5 + 6t$ $(t \in \mathbb{N})$,
- $q = 17$, $\ell = 5 + 18t$ *and* $\ell = 14 + 18t$ $(t \in \mathbb{N})$.

3. Type I: $a_k = \tilde{a}_\ell$ and $b_k = \tilde{b}_\ell$ with certain p and q (Proof of Theorem 2.1)

It is known that the binary recurrence sequences are periodic modulo any positive integer. A simple consideration shows that the terms $\tilde{a}_n$ are never divisible by 2, 3, 7, 11, 13, 17 (primes up to 25), while $\tilde{b}_n$ are never a multiple of 2, 7, 13, 19, 23 (primes also up to 25). On the other hand, $q \mid a_n$ and $q \mid b_n$ hold for any n. Consequently, there is no solution to the equation $a_k = \tilde{a}_\ell$ unless $q = 5, 19, 23, 25$. Indeed, by $q \mid a_n$, one needs only to check one period of $\{\tilde{a}_n\}$ modulo q. Similarly, $b_k = \tilde{b}_\ell$ may possess solution only when $q = 5, 9, 11, 15, 17, 25$. Unfortunately, we could achive the computations only for $q = 5$ regarded to $a_k = \tilde{a}_\ell$, and for $q = 5$ and $q = 9$ regarded to $b_k = \tilde{b}_\ell$ since the time demand of evaluation of the algorithm decribed below seemed to be too much for larger q values.

Suppose that p and q are given, and consider $a_k = \tilde{a}_\ell$. Assume that $x = a_k$ satisfies this equation. Then, by (1.6)

$$y^2 - (\kappa^2 - 4)x^2 = 4(\kappa + 2)q^2 \tag{3.1}$$

holds for some positive integer y. On the other hand, in the virtue of (2.3) (the source of (2.3) is (1.6)), $x = \tilde{a}_\ell$ is also a zero of the equation

$$z^2 - 12x^2 = 600 \tag{3.2}$$

for some positive suitable integer z. Clearly, (3.1) and (3.2) form a system of simultaneous Pellian equations. The `PellianSystem()` procedure, developed in [6] and implemented in MAGMA [5] is able to solve such a system if the coefficients are not too large.

If we take $b_k = \tilde{b}_\ell$, then (3.1) and (3.2) must be replaced by

$$y^2 - (\kappa^2 - 4)x^2 = 4((p-3)^2 - \kappa(p-3) + 1)q^2 \tag{3.3}$$

and

$$z^2 - 12x^2 = -200, \tag{3.4}$$

respectively.

We have checked the solutions of the appropriate system of Pellian equations by MAGMA, and the result of the computations is reported in Theorem 2.1.

To illustrate the time demand of the computations, we note that the MAGMA server needed approximately 21 days to show that $b_k = \tilde{b}_\ell$ has no solution in the case $\{p, q\} = \{8, 9\}$ (this was the worst case we considered).

4. Type II: $a_\varepsilon = \tilde{a}_\ell$, $b_\varepsilon = \tilde{b}_\ell$, part 1. (Background behind Result 2.2)

This section is devoted to deal with the equations above in the specific cases

1. $a_2 = \tilde{a}_\ell$, when parameter p of $\{a_n\}$ is fixed in the range $4 \le p \le 1\,600$,
2. $a_3 = \tilde{a}_\ell$, when parameter q of $\{a_n\}$ satisfies $4 \le q \le 10\,000$,
3. $b_2 = \tilde{b}_\ell$, when $p \in [4; 10\,000]$,
4. $b_2 = \tilde{b}_\ell$, when $q \in [4; 2\,800]$.

The common background behind the four problems is that all of them are linked to hyperelliptic diophantine equations of degree four. Observe, that a_2 and b_2 is a quadratic polynomial in q, similarly a_3 and b_2 has degree two in p.

Consider first

$$a_2 = \tilde{a}_\ell$$

with fixed p. Then, by the first identity of (2.3), a_2 satisfies

$$y^2 - 12a_2^2 = 600,$$

where $a_2 = f(q) = (\kappa + 1)q$ is a quadratic polynomial of q. Consequently we need to solve the quartic hyperelliptic equation

$$y^2 = 12f^2(q) + 600. \tag{4.1}$$

We use the `IntegralQuarticPoints()` procedure of MAGMA package to handle (4.1). Note that if the constant term of the polynomial on the right hand side of (4.1) is not a full square, then the procedure requires a solution (as input) to the equation to determine all solutions. In this case we scanned the interval $J = [-10\,000; 10\,000]$ for q to find a solution. It might occur that there is a solution outside J and not inside J, but we found no example to this.

If once we have determined a q, then we search back the corresponding subscript ℓ.

The analogy to the other 3 cases of this section is obvious: in the right hand side of (4.1) the polyomial f is being replaced by $f(p) = (\kappa^2 + \kappa - 1)q$, $f(q) = (\kappa(p-3) - 1)q$ and $f(p) = (\kappa(p-3) - 1)q$, respectively.

Solutions we found are listed in Result 2.2 (the list might be not full in accordance with the basic interval J which was used for finding a solution).

5. Type III: $a_\varepsilon = \tilde{a}_\ell$, $b_\varepsilon = \tilde{b}_\ell$, part 2. (Proof of Theorem 2.3)

Here we study the title equation in a few cases with small ε, which differ from the previous section. Recall that both of the sequences $\{\tilde{a}_n\}$ and $\{\tilde{b}_n\}$ are purely periodic for any positive integer modulus.

Since $a_1 = q$ the equation $a_1 = \tilde{a}_\ell$ has, trivially, infinitely many solutions.

The next problem is $a_2 = \tilde{a}_\ell$ with fixed q. (The case with fixed p has already been studied in Section 4.) Recall that $a_2 = (\kappa + 1)q$, more precisely

$$a_2 = q(q-2)(p-2) - q$$

is linear in p. Therefore we need to determine the common terms of an arithmetic progression and the sequence $\{\tilde{a}_n\}$. The situation does not change if we consider $b_1 = \tilde{b}_\ell$ with either fixed p or fixed q. Indeed, $b_1 = (p-3)q$ is linear both in p and q.

Obviously, $a_2 \equiv -q \pmod{q(q-2)}$. Consequently, the equation $a_2 = \tilde{a}_\ell$ is soluble if and only if we find at least one element in the sequence $\{\tilde{a}_n\}$, which is congruent $-q$ modulo $q(q-2)$. Because of the periodicity, one must check only one period of $\{\tilde{a}_n\}$ modulo $q(q-2)$.

Assume first that $q = 5$. Then for the modulus $q(q-2) = 15$ we have $\tilde{a}_{2+2t} \equiv -5$ (the cycle's length is 2, and $t \in \mathbb{N}$). Hence $a_2 = \tilde{a}_{2+2t}$, further

$$p = \frac{\tilde{a}_{2+2t} + q}{q(q-2)} + 2.$$

t	0	1	2	3	4	5
$a_2 = \tilde{a}_{2+2t}$	25	355	4945	68875	959305	13361395
p	4	26	332	4594	63956	890762

Table 2: *First few solutions to $a_2 = \tilde{a}_\ell$ when $q = 5$*

The first six t values yield the following solutions. (If $t = 0$ then the two sequences $\{a_n\}$ and $\{\tilde{a}_n\}$ coincide.)

If $q > 5$ the first non-trivial solution is occurred when $q = 19$. Here the length of the cycle is 90, and $q(q-2) \mid \tilde{a}_{58} + 19$, $q(q-2) \mid \tilde{a}_{78} + 19$. That is $a_2 = \tilde{a}_{58+90t}$ and $a_2 = \tilde{a}_{78+90t}$ $(t \in \mathbb{N})$ provide all solutions for suitable values p. For instance, $t = 0$ gives

$$p = 8\,437\,940\,669\,128\,098\,583\,408\,551\,589\,590$$

and

$$p = 2\,318\,394\,927\,973\,629\,460\,854\,893\,981\,169\,574\,319\,067\,870,$$

respectively.

The treatment is similar for $b_1 = \tilde{b}_\ell$. If $q = 5$, then solution always exists since $b_1 = (p-3)q$, $5 \mid \tilde{b}_\ell$, therefore $p = \tilde{b}_\ell/5 + 3$. ($\tilde{b}_2$ and $\tilde{b}_3$ give back solutions have already been appeared in Theorem 2.1.) Now $b_1 \equiv 0 \pmod{q}$, and fixing $q \geq 6$ the first solution appears for $q = 9$, when the cycle length is 18 (modulo q), and we have $b_1 = \tilde{b}_{5+18t}$ and $b_1 = \tilde{b}_{14+18t}$ $(t \in \mathbb{N})$. These results can be directly converted the results corresponding to p, therefore we omit the appropriate analysis.

The results we obtained are summarized in Theorem 2.3.

Finally, we examine the equation $a_3 = \tilde{a}_\ell$ with fixed q, further $b_3 = \tilde{b}_\ell$ when exactly one of p and q is given. In each case we have a polynomial of degree three, let say $\phi(x)$, and we look for the common values of the polynomial and a binary recurrence. By (1.6), the problem leads to the hyperelliptic equation

$$y^2 = 12\phi^2(x) + c$$

of degree 6, where the constant c is either 600 or -200. Since the leading coefficient on the right hand side is not a square, there is no genearal algorithm to solve. For example, $p = 5$ provides now

$$y^2 = 12q^2(9q^2 - 45q + 55)^2 + 600.$$

After dividing by 4, we have

$$y_1^2 = 243q^6 - 2430q^5 + 9045q^4 - 14850q^3 + 9075q^2 + 150,$$

and the techique of the solution is not known.

Acknowledgements. The authors thank P. Olajos for his valuable help in using MAGMA package.

References

[1] Horváth, J., Über die regulären Mosaiken der hyperbolishen Ebene, *Annales Univ. Sci.*, Sectio Math. 7 (1964), 49–53.

[2] Németh, L., Combinatorial examination of mosaics with asymptotic pyramids and their reciprocals in 3-dimensional hyperbolic space, *Studia Sci. Math.*, 43 (2) (2006), 247–265.

[3] Németh, L., On the 4–dimensional hyperbolic hypercube mosaic, *Publ. Math. Debrecen*, 70/3–4 (2007), 291-305.

[4] Németh, L., Trees on the hyperbolic honeycombs, accepted in *Miskolc Math. Notes.*

[5] MAGMA Handbook, `http://magma.maths.usyd.edu.au/magma/handbook/`

[6] Szalay, L., On the resolution of simultaneous Pell equations, *Ann. Math. Inform.*, 34 (2007), 77–87.

Annales Mathematicae et Informaticae
43 (2014) pp. 123–144
http://ami.ektf.hu

Probabilistic model checking on HPC systems for the performance analysis of mobile networks

Wolfgang Schreiner[a], **Tamás Bérczes**[b], **János Sztrik**[b]

[a]Research Institute for Symbolic Computation (RISC), Johannes Kepler University Linz
Wolfgang.Schreiner@risc.jku.at

[b]Faculty of Informatics, University of Debrecen
berczes.tamas@inf.unideb.hu, sztrik.janos@inf.unideb.hu

Submitted March 28, 2014 — Accepted May 27, 2014

Abstract

We report on the use of HPC resources for the performance analysis of the mobile cellular network model described in "A New Finite-Source Queueing Model for Mobile Cellular Networks Applying Spectrum Renting" by Tien Van Do *et al.* That paper proposed a new finite-source retrial queueing model with spectrum renting that was analyzed with the MOSEL-2 tool. Our results show how this model can be also appropriately described and analyzed with the probabilistic model checker PRISM, although at some cost considering the formulation of the model; in particular, we are able to accurately reproduce most of the analytical results presented in that paper and thus increase the confidence in the previously presented results. However, we also outline some discrepancies which may hint to deficiencies of the original analysis. Moreover, by applying a parallel computing framework developed for this purpose, we are able to considerably speed up studies performed with the PRISM tool. The investigations are illustrated by figures and conclusions are drawn.

1. Introduction

We report in this paper on the application of high performance computing (HPC) resources for the performance analysis of mobile networks. We use the mobile cellular network system that was introduced in [4] where a number of sources (cell phone subscribers) compete for access to a number of servers (channels). Sources produce requests at rate λ; a free server processes these requests at rate μ. However, the number of available channels varies: if it gets to small, the cell phone operator may rent additional frequency blocks from another operator, partition these blocks into channels, and add these new channels to its own pool. If sufficiently many channels have become free again, the rented blocks may be released. This model is an extension of the mobile cellular network model presented in [2] which for the first time considered the renting of frequencies that are organized in blocks but did not yet consider the retrial phenomenon; in [3], the phenomenon of impatient customers waiting in the orbit was further investigated.

In [4], this model was originally analyzed with the help of the performance modeling tool MOSEL-2 [1] which is however not supported any more. Our own results are derived with the help of the probabilistic model checker PRISM [6, 7] which is actively developed and has been used for numerous purposes, among them the performance analysis of computing systems. In [9], we have developed an initial version of the model in PRISM which was subsequently refined and corrected in [8]. Furthermore, we have in [8] described a parallel computing framework that we applied to analyze the PRISM model with the use of HPC resources, i.e., we have sped up the analysis of our model by running experiments on a massively parallel non-uniform memory architecture (NUMA).

However, in [9, 8] only a small number of experiments were performed, some of which derived different results than were originally reported in [4]. This paper complements our work by presenting all experiments that were also described in [4] and illustrating for the whole set of experiments the speedup that can be achieved by their execution in our parallel computing framework. Tool supported analysis of cellular networks with finite-source retrial queuing system was treated in [5], too.

The remainder of this paper is structured as follows: to make this paper self-reliant, we summarize in Section 2 the previously introduced model and the parallel execution framework. In Section 3, we present our new results and contrast them to those reported in [4]. Section 4 presents our conclusions and open issues for further work.

2. The model

Appendix B presents the PRISM model that was introduced in [9]: it applies the concept of spectrum renting for mobile cellular networks introduced in [4]. However, that paper also shows results for a corresponding model (that is not described in detail there) without spectrum renting. In order to repeat the corresponding experiments, we give in Appendix A a version of our PRISM model from which spectrum renting has been stripped but which is otherwise identical.

The experiments of this paper were performed with the execution script listed in Appendix C; it applies the parallel execution framework (command `parallel`) introduced in [9]. This command is implemented by a C program with the help of the POSIX multithreading API; it uses a manager/worker scheme to execute an arbitrary number of commands by a fixed number of processes. In detail, when started as `parallel` T, the program creates a pool of T worker threads and then starts an auxiliary thread that reads an arbitrary number of command lines from the standard input stream into an internal queue. At the same time the main thread acts as the manager and schedules the execution of the commands among the T worker threads: whenever a worker thread becomes idle, the manager removes one command from the queue and assigns it to the thread which then spawns (by the Unix `system` call) a process to execute that command, waits for its termination, and becomes free again to receive another command. Additionally the manager thread prints out status information for every command whose execution has terminated. If the command is executed on a multi-core/multi-processor system, the commands are thus executed by at most T processor cores.

The experiments were performed on an SGI Altix UltraViolet 1000 supercomputer installed at the Johannes Kepler University Linz. This machine is equipped with 256 Intel Xeon E78837 processors with 8 cores each which are distributed among 128 nodes with 2 processors (i.e. 16 cores) each; the system thus supports computations with up to 2048 cores. Access to this machine is possible via interactive login; by default every user may execute threads on 4 processors with 32 cores and 256 GB memory. Since PRISM is implemented in Java, we applied the execution script `prism-java` described in [9] which calls `java` with memory allocation and optimization optimized for execution on a NUMA system.

3. The analysis

With the help of our parallel execution framework, we have performed for our PRISM model all the experiments that were also reported in [4]; the results are depicted in Figures 2 to 10 (with references to the corresponding figures presented in [4]). The experiments shown in Figure 2 (corresponding to Figure 2 in [4]) are performed in the model without spectrum renting (see Appendix A); all other ones are performed in the model with spectrum renting (see Appendix B); in the later case appropriate variants of the model were used as required by the different sets of parameters with varying respectively fixed values.

From the 29 experiments (comprising in total 920 PRISM runs to produce the various data points of each experiment), 25 show results that are virtually identical to those presented in [4]. This correspondence strongly increases the confidence in both the original MOSEL-2 model and in our PRISM model. However, there are also four notable discrepancies:

- As already stated in [8], in Figure 3 (corresponding to Figure 3 of [4]) the two bottom diagrams show in our model (especially for low traffic intensity ρ_0) a lower mean number of rented blocks mB and a lower mean number of busy channels mC than originally reported (while the overall shape of the curves

P	t_p (s)			T_p (s)	S_p
1	3813	3788	3792	3798	1.0
2	1949	1950	1944	1948	1.9
4	1028	1025	1018	1024	3.7
8	559	561	560	560	6.8
16	334	328	329	330	11.5
32	252	252	244	249	15.3

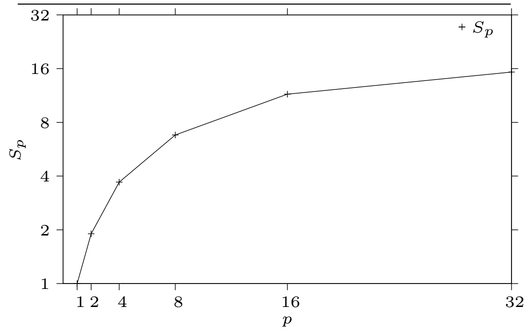


Figure 1: Execution Times and Speedups

are similar).

- Figures 9 and 10 (corresponding to Figures 9 and 10 of [4]) reporting on the impact of retrials on the average profit rate (APR) and on the average number of busy channels (mC) show for the first parameter set $\rho_0 = 0.4, p_{io} = 0.8$ the same results as originally reported; however for the two other parameter sets our experiments report significantly lower figures, i.e., the three lines are much farther apart than in [4].

Since in all other cases the results are identical to the other reports and we have both carefully checked our model and the deviating experiments, the possibility remains that the errors are in solvers of the MOSEL and PRISM. One should know that these details are hidden and we have no information about the solution methods.

As for the time needed for executing the analysis, Figure 1 lists the times (in seconds) for performing all the 920 PRISM runs illustrated in Figures 2–10 with P processes, $1 \leq P \leq 32$ (the maximum number of processor cores available to us for this experiment). The analysis was performed five times from which we have excluded the fastest and the slowest run. This leads to three values for the execution time t_p with average execution time T_p; the speedup for this average is reported as S_p.

We see that significant speedups up to a maximum of 15.3 can be achieved. The main reason that from $P = 16$ to $P = 32$ the speedup does not grow so much any more is that we have have attached to every Java thread that executes

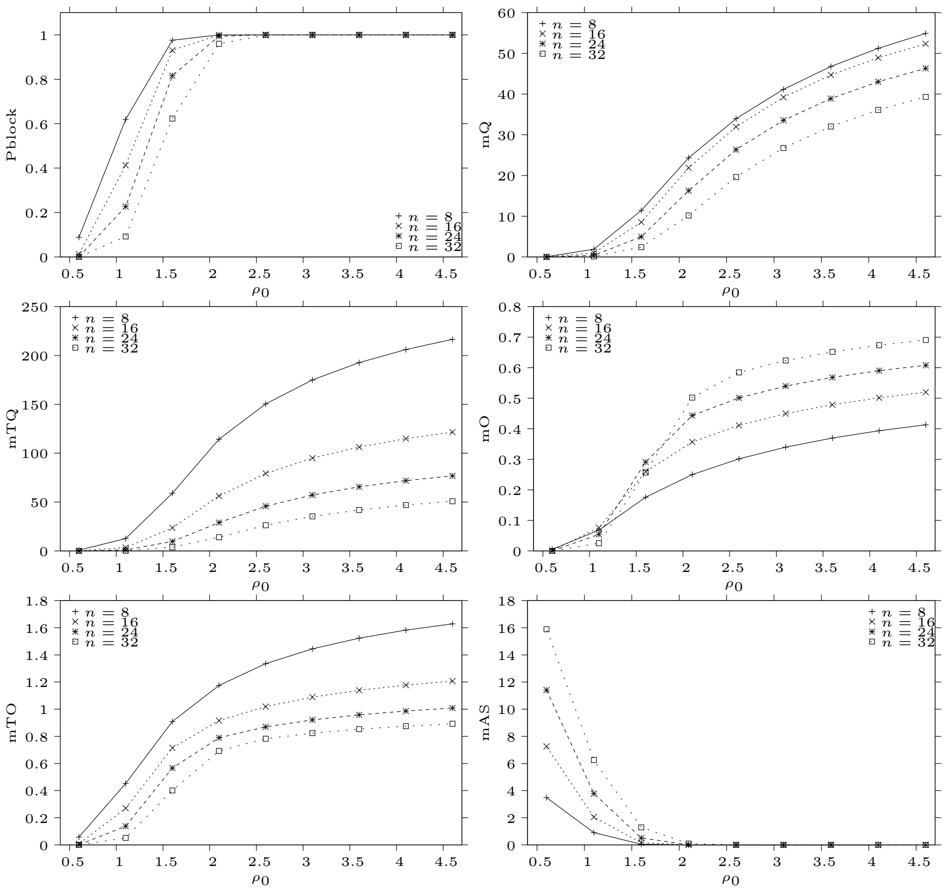


Figure 2: Performance Measures Without Renting (cf. Fig. 2 from [4])

one instance of PRISM by the command line option `-XX:ParallelGCThreads` a number of garbage collection threads that concurrently reclaim the memory of objects that are not accessible any more; since the experiment was performed on only 32 processor cores; the number of concurrently executing threads thus significantly exceeded the number of cores. With more cores available, we can expect also for $P = 32$ a considerably higher speedup.

4. Conclusions

We have shown in this report how the PRISM analysis of a non-trivial mobile cellular network can be efficiently performed on a modern high performance com-

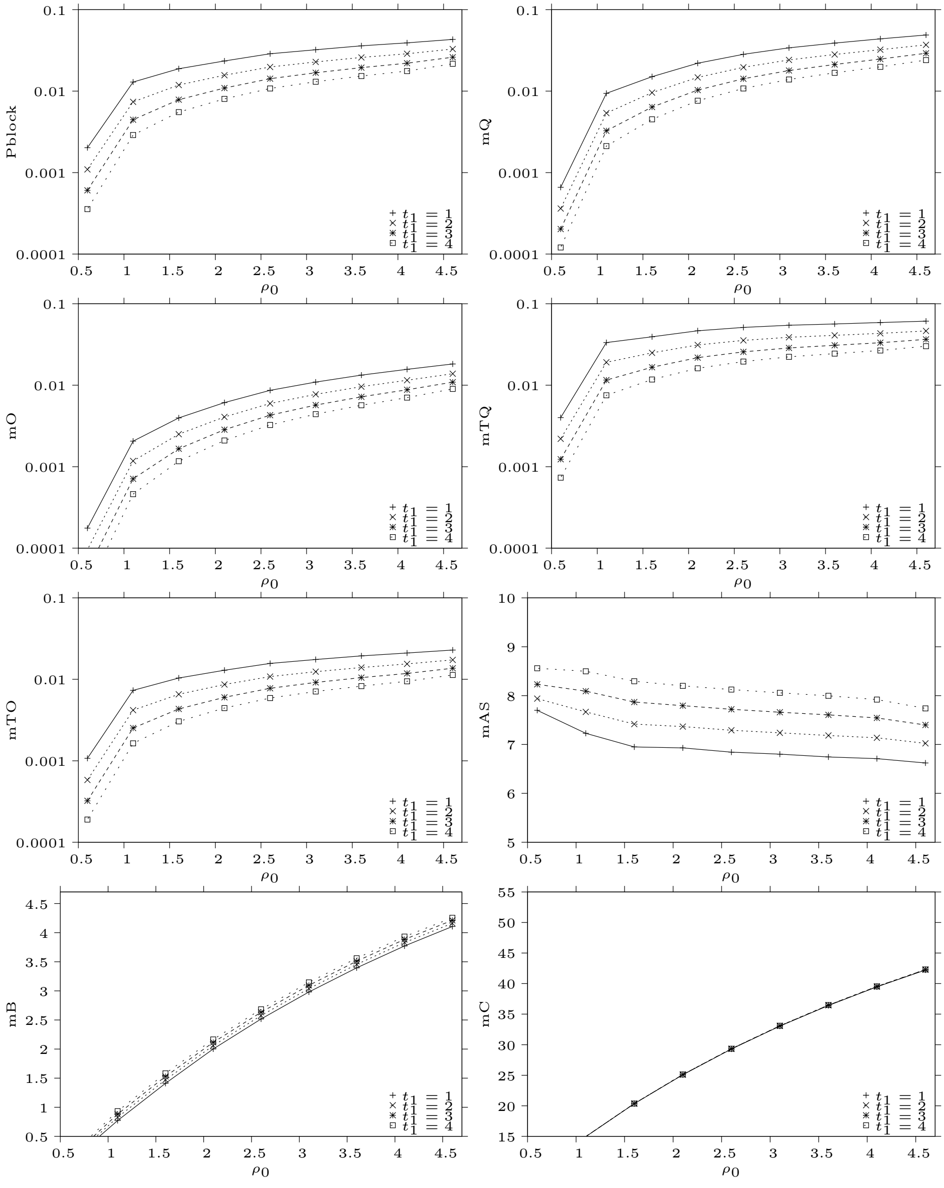


Figure 3: Performance Measures for $t_2 = 6$ (cf. Fig. 3 from [4])

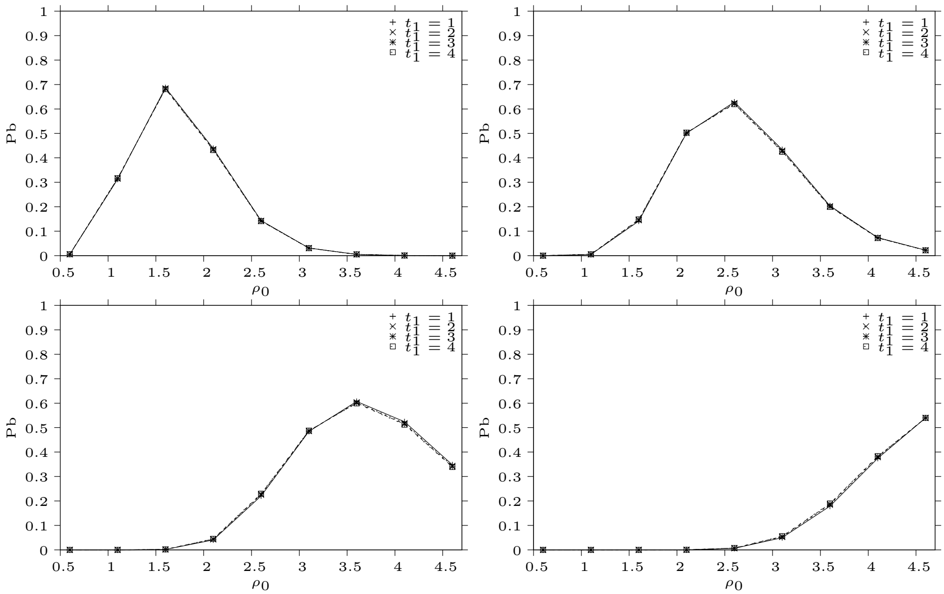


Figure 4: Further Performance Measures for $t_2 = 6$ (cf. Fig. 4 from [4])

puting system and how by this analysis the results performed with the older (and not any more supported) MOSEL-2 tool can be essentially confirmed. However, as already reported in [8], a crucial difference between MOSEL-2 and PRISM (the existence respectively lack of zero-time/infinite-rate transitions) makes the PRISM model somewhat more unhandy than originally thought; more efforts are needed in PRISM to express the desired models in an economical way.

Furthermore, while most of the originally reported results (25 of 29 experiments) could be confirmed, still some discrepancies (in 4 experiments) have to be resolved. While the error may well be in the PRISM model or its analysis, it might as well be true that there are errors in the originally reported results (we have asked one author of the original paper for a re-examination of these experiments). This demonstrates that the performance analysis of computing systems by analyzing system models alone cannot give full confidence in the correctness of the results: further verification (by comparison against measurements of the actual system) or validation (by comparison with the analysis of another model by another tool) is highly recommended.

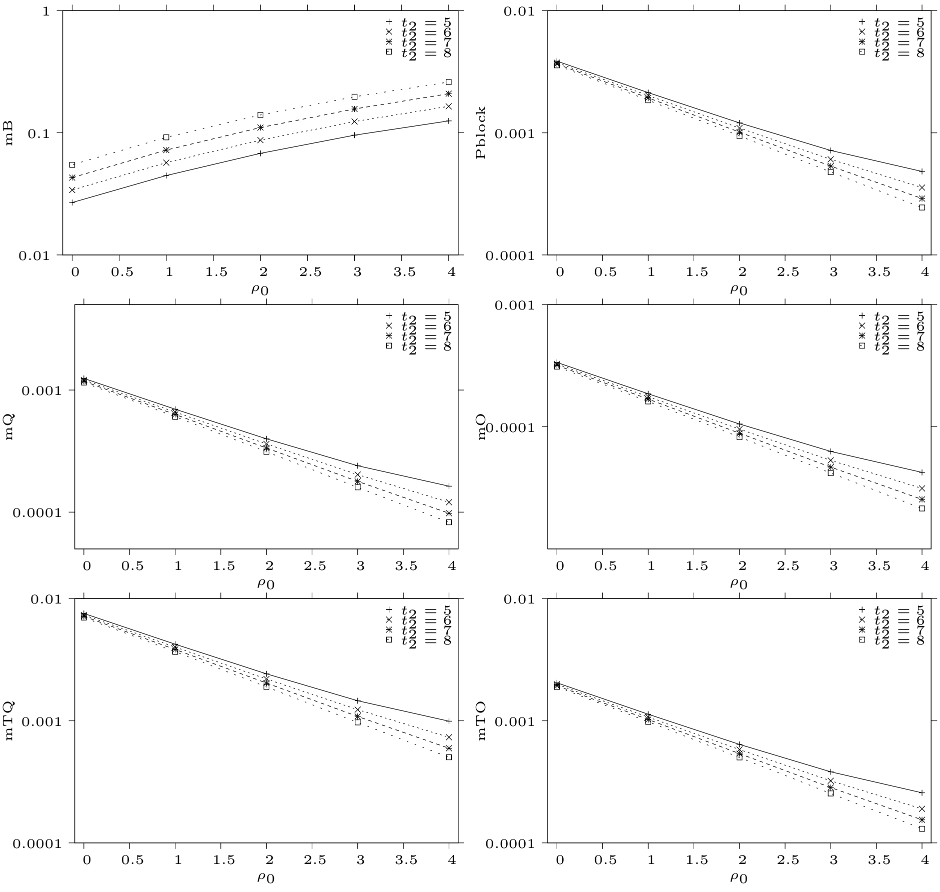


Figure 5: Performance Measures for $\rho_0 = 0.6$ (cf. Fig. 5 from [4])

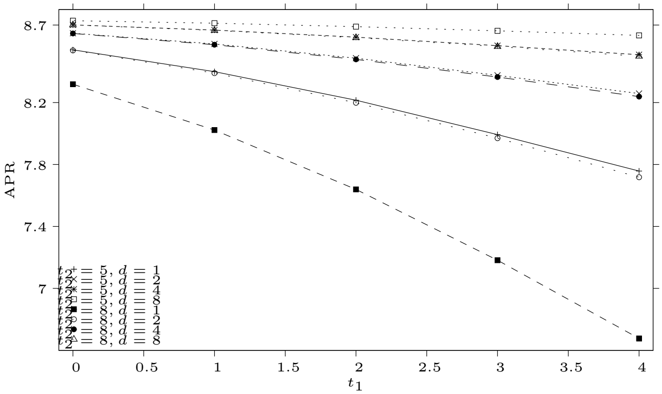


Figure 6: APR vs. t_1 and d for $\rho_0 = 0.6$ (cf. Fig. 6 from [4])

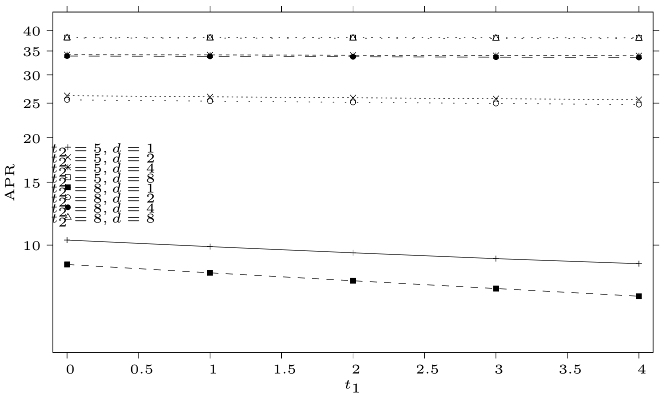


Figure 7: APR vs. t_1 and d for $\rho_0 = 4.6$ (cf. Fig. 7 from [4])

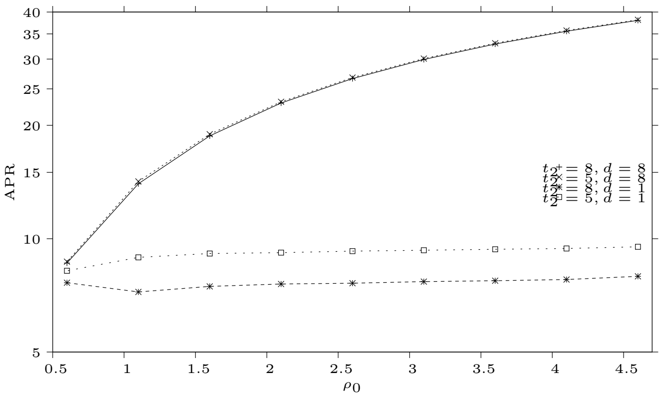


Figure 8: APR vs. ρ_0 and d (cf. Fig. 8 from [4])

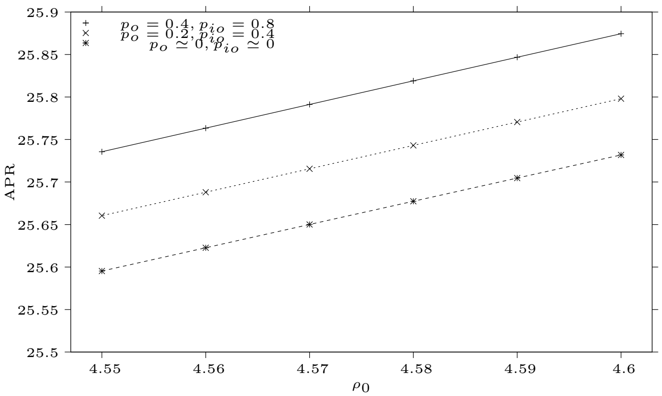


Figure 9: Impact of retrials on APR (cf. Fig. 9 from [4])

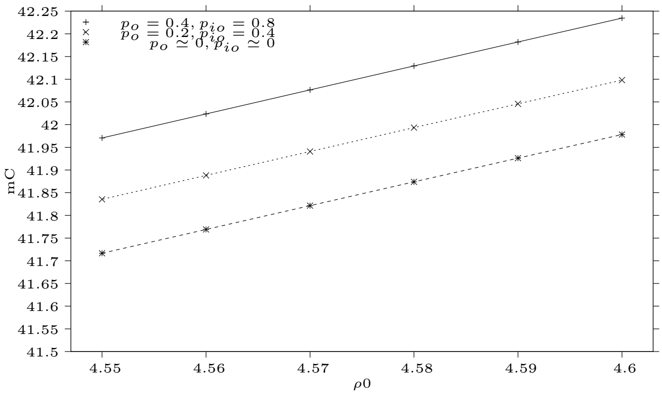


Figure 10: Impact of retrials on the average number of busy channels (cf. Fig. 9 from [4])

Acknowledgment

The publication was supported by the TÁMOP 4.2.2. C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

The authors are grateful to the reviewers for their comments and suggestions which improved the quality of the paper.

References

[1] K. Begain, G. Bolch, and Herold H. *Practical Performance Modeling Application of the MOSEL Language*. Kluwer Academic Publisher, 2012.

[2] T. V. Do, N. H. Do, and R. Chakka. A New Queueing Model for Spectrum Renting in Mobile Cellular Networks. *Computer Communications*, 35:1165–1171, 2012.

[3] T. V. Do, N.H. Do, and J. Zhang. An Enhanced Algorithm to Solve Multiserver Retrial Queueing Systems with Impatient Customers. *Computers & Industrial Engineering*, 65(4):719–728, 2013.

[4] T. V. Do, P. Wüchner, T. Bérczes, J. Sztrik, and H. de Meer. A New Finite-Source Queueing Model for Mobile Cellular Networks Applying Spectrum Renting. *Asia-Pacific Journal of Operational Research*, 31(2):14400004, 2014. 19 pages, DOI: 10.1142/S0217595914400041.

[5] N. Gharbi, B. Nemmouchi, L. Mokdad, and J. Ben-Othma. The Impact of Breakdowns Disciplines and Repeated Attempts on Performances of Small Cell Networks. *Journal of Computational Science*, 2014. http://dx.doi.org/10.1016/j.jocs.2014.02.011.

[6] M. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of Probabilistic Real-time Systems. In G. Gopalakrishnan and S. Qadeer, editors, *Proc. 23rd Interna-*

tional Conference on Computer Aided Verification (CAV'11), volume 6806 of *Lecture Notes in Computer Science*, pages 585–591. Springer, 2011.

[7] PRISM — Probabilistic Symbolic Model Checker, 2013. `http://www.prismmodelchecker.org`.

[8] W. Schreiner, N. Popov, T. Bérczes, J. Sztrik, and G. Kusper. Applying High Performance Computing to Analyzing by Probabilistic Model Checking Mobile Cellular Networks with Spectrum Renting. Technical report, Research Institute for Symbolic Computation (RISC), Johannes Kepler University Linz, Austria, July 2013.

[9] Wolfgang Schreiner. Initial Results on Modeling in PRISM Mobile Cellular Networks with Spectrum Renting. Technical report, Research Institute for Symbolic Computation (RISC), Johannes Kepler University Linz, Austria, March 2013.

A. The PRISM model without spectrum renting

```
// -------------------------------------------------------------------
// Spectrum0.prism
// A model for mobile cellular networks.
//
// The model serves as the comparison basis for the improvements
// introduced by the application of "spectrum renting" in
//
//   Tien v. Do, Patrick Wüchner, Tamas Berczes, Janos Sztrik,
//   Hermann de Meer: A New Finite-Source Queueing Model for
//   Mobile Cellular Networks Applying Spectrum Renting,
//   September 2012.
//
// Use for fastest checking the "Sparse" engine and the "Gauss-Seidel"
// solver and switch off "use compact schemes".
//
// Author: Wolfgang Schreiner <Wolfgang.Schreiner@risc.jku.at>
// Copyright (C) 2013, Research Institute for Symbolic Computation
// Johannes Kepler University, Linz, Austria, http://www.risc.jku.at
// -------------------------------------------------------------------

// continuous time markov chain (ctmc) model
ctmc

// -------------------------------------------------------------------
// system parameters
// -------------------------------------------------------------------

// bounds
const int K = 100;   // population size
const int n;         // number of servers/channels

// rates
const double rho;                 // normalized traffic intensity
const double mu     = 1/53.22;    // service rate
const double lambda = rho*n*mu/K; // call generation rate
const double nu     = 1;          // retrial rate
const double eta    = 1/300;      // rate of queueing users
                                  // getting impatient
```

```
// probabilities
const double p_b  = 0.1;       // prob. that user gives up
                               // (-> sources)
const double p_q  = 0.5;       // prob. that user presses button
                               // (-> queue)
const double p_o  = 1-p_b-p_q; // prob. that user retries later
                               // (-> orbit)
const double p_io = 0.8;       // prob. that impatient user retries
                               // later (-> orbit)
const double p_is = 1-p_io;    // prob. that impantient user gives up
                               // (-> sources)

// ------------------------------------------------------------------
// system model
// note that the order of the modules influences model checking time
// heuristically, this seems to be the best one
// ------------------------------------------------------------------

// available servers accept requests
module Servers
  servers: [0..n] init 0;
  [sservers] servers < n -> (servers' = servers+1);
  [oservers] servers < n -> (servers' = servers+1);
  [ssources1] servers > 0 & queue = 0 ->
    servers*mu : (servers' = servers-1);
  [ssources2] servers > 0 ->
     servers*mu : true ;
endmodule

// generate requests at rate sources*lambda
module Sources
  sources: [0..K] init K;
  [sservers] sources > 0 ->
    sources*lambda : (sources' = sources-1);
  [sorbit]   sources > 0 & servers = n ->
    sources*lambda*p_o : (sources' = sources-1);
  [squeue]   sources > 0 & servers = n ->
    sources*lambda*p_q : (sources' = sources-1);
  [ssources1] sources < K -> (sources' = sources+1);
  [ssources2] sources < K -> (sources' = sources+1);
  [qsources]  sources < K -> (sources' = sources+1);
  [osources]  sources < K -> (sources' = sources+1);
endmodule

// if no server is available, requests are redirected
// with probability p_o to the orbit
formula orbit = K-(sources+servers+queue); // make variable virtual
module Orbit
  // orbit: [0..K-n] init 0;
  [sorbit]   orbit < K-n -> true;
  [qorbit]   orbit < K-n -> true;
  [oservers] orbit > 0 -> orbit*nu : true;
  [oqueue]   orbit > 0 & servers = n -> orbit*nu*p_q : true;
  [osources] orbit > 0 & servers = n -> orbit*nu*p_b : true;
endmodule
```

```
// if no server is available, requests are redirected
// with probability p_q to the queue
module Queue
  queue: [0..K-n] init 0;
  [squeue]    queue < K-n -> (queue' = queue+1);
  [oqueue]    queue < K-n -> (queue' = queue+1);
  [qorbit]    queue > 0 & servers = n ->
    queue*eta*p_io : (queue' = queue-1);
  [qsources]  queue > 0 & servers = n ->
    queue*eta*p_is : (queue' = queue-1);
  [ssources2] queue > 0 -> (queue' = queue-1);
endmodule

// -----------------------------------------------------------------
// system rewards
// -----------------------------------------------------------------

// mean number of active requests
rewards "mM"
  true : max(0, orbit)+queue+servers;
endrewards

// mean number of calls in orbit
rewards "mO"
  true : max(0, orbit);
endrewards

// mean number of calls in queue
rewards "mQ"
  true: queue;
endrewards

// mean number of active calls
rewards "mC"
  true: servers;
endrewards


// -----------------------------------------------------------------
// Spectrum0.props
// -----------------------------------------------------------------

// mean number of active requests
"mM" : R{"mM"}=? [ S ] ;

// mean number of active sources
"mK" : K-"mM" ;

// mean throughput (served and unserved)
"m1" : "mK"*lambda ;

// mean number of active calls
"mC" : R{"mC"}=? [ S ] ;

// mean goodput
"m1good" : "mC"*mu ;
```

```
// probability that arriving customer gets served
"Pgood" : "m1good"/"m1" ;

// mean response time (served and unserved)
"mT" : "mM"/"m1" ;

// mean number of idle servers
"mAS" : n-"mC" ;

// utilization of available servers
"Sutil" : "mC"/n ;

// blocking probability
"Pblock" : S=? [ servers = n ] ;

// mean queue length
"mQ" : R{"mQ"}=? [ S ] ;

// mean time spent in queue
"mTQ" : "mQ" / "m1" ;

// mean orbit length
"mO" : R{"mO"}=? [ S ] ;

// mean time spent in orbit
"mTO" : "mO" / "m1" ;
```

B. The PRISM model with spectrum renting

```
// ----------------------------------------------------------------
// Spectrum.prism
// A model for mobile cellular networks applying spectrum renting.
//
// The model is described in
//
//   Tien v. Do, Patrick Wüchner, Tamas Berczes, Janos Sztrik,
//   Hermann de Meer: A New Finite-Source Queueing Model for
//   Mobile Cellular Networks Applying Spectrum Renting,
//   September 2012.
//
// Use for fastest checking the "Sparse" engine and the "Gauss-Seidel"
// solver and switch off "use compact schemes".
//
// Author: Wolfgang Schreiner <Wolfgang.Schreiner@risc.jku.at>
// Copyright (C) 2013, Research Institute for Symbolic Computation
// Johannes Kepler University, Linz, Austria, http://www.risc.jku.at
// ----------------------------------------------------------------

// continuous time markov chain (ctmc) model
ctmc

// ----------------------------------------------------------------
```

```
// system parameters
// ---------------------------------------------------------------

// renting tresholds
const int t1;      // block renting treshold
const int t2 = 6;  // block release treshold

// bounds
const int K = 100;   // population size
const int r = 8;     // number of servers/channels per block
const int m = 5;     // maximum number of blocks that can be rented
const int n = 2*r;   // minimum number of servers/channels
const int M = n+r*m; // maximum number of simultaneous calls

// rates
const double rho;                 // normalized traffic intensity
const double mu     = 1/53.22;    // service rate
const double lambda = rho*n*mu/K; // call generation rate
const double nu     = 1;          // retrial rate
const double eta    = 1/300;      // rate of queueing users getting
                                  // impatient
const double lam_r  = 1/5;        // block renting rate
const double nu_r   = 1/7;        // block rental retrial rate
const double mu_r   = 1;          // block release reate

// probabilities
const double p_b  = 0.1;       // prob. that user gives up (-> sources)
const double p_q  = 0.5;       // prob. that user presses button
                               // (-> queue)
const double p_o  = 1-p_b-p_q; // prob. that user retries later
                               // (-> orbit)
const double p_io = 0.8;       // prob. that impatient user retries
                               // later (-> orbit)
const double p_is = 1-p_io;    // prob. that impantient user gives up
                               // (-> sources)
const double p_r  = 0.8;       // block rental success probability
const double p_f  = 1-p_r;     // block rental failure probability

// ---------------------------------------------------------------
// system model
// note that the order of the modules influences model checking time
// heuristically, this seems to be the best one
// ---------------------------------------------------------------

// number of currently available servers/channels
formula servAvail = n+blocks*r;

// blocks are rented at rate lam_r and released at rate mu_r
// renting is successful with probability p_r and fails with
// probability p_f retrying a failed attempt is performed at rate nu_r
module Blocks
  blocks: [0..m] init 0;
  trial: [0..1] init 0;
  [success1] trial = 0 & servAvail-servers <= t1 & blocks < m & queue=0 ->
    lam_r*p_r: (blocks' = blocks+1);
  [success2] trial = 0 & servAvail-servers <= t1 & blocks < m ->
```

```
    lam_r*p_r: (blocks' = blocks+1);
  [failure] trial = 0 & servAvail-servers <= t1 & blocks < m ->
    lam_r*p_f: (trial' = 1);
  [retrial1] trial = 1 & servAvail-servers <= t1 & blocks < m & queue=0 ->
    nu_r*p_r : (trial' = 0) & (blocks' = blocks+1) ;
  [retrial2] trial = 1 & servAvail-servers <= t1 & blocks < m ->
    nu_r*p_r : (trial' = 0) & (blocks' = blocks+1) ;
  [interrupt] trial = 1 & servAvail-servers > t1  ->
    9999 : (trial' = 0); // "immediately"
  [release] servAvail-servers >= t2+r & blocks > 0 ->
    mu_r : (blocks' = blocks-1);
endmodule

// available servers accept requests
module Servers
  servers: [0..M] init 0;
  [sservers] servers < servAvail -> (servers' = servers+1);
  [oservers] servers < servAvail -> (servers' = servers+1);
  [success2] servers < M         -> (servers' = servers+1);
  [retrial2] servers < M         -> (servers' = servers+1);
  [ssources1] servers > 0 & queue = 0 ->
    servers*mu : (servers' = servers-1);
  [ssources2] servers > 0 ->
     servers*mu : true ;
endmodule

// generate requests at rate sources*lambda
module Sources
  sources: [0..K] init K;
  [sservers] sources > 0 ->
    sources*lambda : (sources' = sources-1);
  [sorbit]   sources > 0 & servers = servAvail ->
    sources*lambda*p_o : (sources' = sources-1);
  [squeue]   sources > 0 & servers = servAvail ->
    sources*lambda*p_q : (sources' = sources-1);
  [ssources1] sources < K -> (sources' = sources+1);
  [ssources2] sources < K -> (sources' = sources+1);
  [qsources]  sources < K -> (sources' = sources+1);
  [osources]  sources < K -> (sources' = sources+1);
endmodule

// if no server is available, requests are redirected
// with probability p_o to the orbit
formula orbit = K-(sources+servers+queue); // make variable virtual
module Orbit
  // orbit: [0..K-n] init 0;
  [sorbit]   orbit < K-n -> true;
  [qorbit]   orbit < K-n -> true;
  [oservers] orbit > 0 -> orbit*nu : true;
  [oqueue]   orbit > 0 & servers = servAvail -> orbit*nu*p_q : true;
  [osources] orbit > 0 & servers = servAvail -> orbit*nu*p_b : true;
endmodule

// if no server is available, requests are redirected
// with probability p_q to the queue
module Queue
```

```
  queue: [0..K-n] init 0;
  [squeue]    queue < K-n -> (queue' = queue+1);
  [oqueue]    queue < K-n -> (queue' = queue+1);
  [qorbit]    queue > 0 & servers = servAvail ->
    queue*eta*p_io : (queue' = queue-1);
  [qsources]  queue > 0 & servers = servAvail ->
    queue*eta*p_is : (queue' = queue-1);
  [ssources2] queue > 0 -> (queue' = queue-1);
  [success2]  queue > 0 -> (queue' = queue-1);
  [retrial2]  queue > 0 -> (queue' = queue-1);
endmodule

// ----------------------------------------------------------------------
// system rewards
// ----------------------------------------------------------------------

// mean number of active requests
rewards "mM"
  true : max(0, orbit)+queue+servers;
endrewards

// mean number of calls in orbit
rewards "mO"
  true : max(0, orbit);
endrewards

// mean number of calls in queue
rewards "mQ"
  true: queue;
endrewards

// mean number of active calls
rewards "mC"
  true: servers;
endrewards

// mean number of active blocks
rewards "mB"
  true: blocks;
endrewards


// ----------------------------------------------------------------------
// Spectrum.props
// ----------------------------------------------------------------------

// mean number of active requests
"mM" : R{"mM"}=? [ S ] ;

// mean number of active sources
"mK" : K-"mM" ;

// mean throughput (served and unserved)
"m1" : "mK"*lambda ;

// mean number of active calls
"mC" : R{"mC"}=? [ S ] ;
```

```
// mean goodput
"m1good" : "mC"*mu ;

// probability that arriving customer gets served
"Pgood" : "m1good"/"m1" ;

// mean response time (served and unserved)
"mT" : "mM"/"m1" ;

// mean number of rented blocks
"mB" : R{"mB"}=? [ S ] ;

// mean number of available servers
"mS" : n+"mB"*r ;

// mean number of idle servers
"mAS" : "mS"-"mC" ;

// utilization of available servers
"Sutil" : "mC"/"mS" ;

// blocking probability
"Pblock" : S=? [ servers = servAvail ] ;

const int B;

// probability that B blocks are partially utilized
"Pb" : S=? [ n+r*(B-1) < servers & servers <= n+r*B ] ;

// mean queue length
"mQ" : R{"mQ"}=? [ S ] ;

// mean time spent in queue
"mTQ" : "mQ" / "m1" ;

// mean orbit length
"mO" : R{"mO"}=? [ S ] ;

// mean time spent in orbit
"mTO" : "mO" / "m1" ;

const int d;

// average profit rate
"APR" : "mC" - (r/d) * "mB" ;
```

C. The parallel execution script

```
#!/bin/sh

# the program locations
export PRISM_JAVA="prism-java"
PRISM="prism"
```

```
PARALLEL="./parallel"
TIME="time"

# the input/output locations
MODELFILE="Spectrum.prism"
MODELFILE0="Spectrum0.prism"
MODELFILE2="Spectrum2.prism"
MODELFILE3="Spectrum3.prism"
PROPSFILE="Spectrum.props"
PROPSFILE0="Spectrum0.props"
RESULTDIR="Results"
LOGDIR="Logfiles"
LOGFILE="LOGFILE"

# the checker settings
PRISMOPTIONS="-sparse -gaussseidel -nocompact"

# the number of processes to be used
for PROC in 1 2 4 8 16 32 ; do

(

# the properties to be checked and the parameters for the experiment

# Figure 2
for PROPERTY in Pblock mO mTO mQ mTQ mAS ; do
  for RHO in $(seq 0.6 0.5 4.6) ; do
    for N in 8 16 24 32 ; do
      echo "$PRISM $PRISMOPTIONS $MODELFILE0 $PROPSFILE0 -prop $PROPERTY \
        -const rho=$RHO,n=$N \
        -exportresults $RESULTDIR/Fig2-$PROPERTY-$N-$RHO \
        > $LOGDIR/Fig2-$PROPERTY-$N-$RHO"
    done
  done
done

# Figure 3
for PROPERTY in Pblock mO mTO mB mQ mTQ mC mAS ; do
  for RHO in $(seq 0.6 0.5 4.6) ; do
    for T1 in $(seq 1 1 4) ; do
      echo "$PRISM $PRISMOPTIONS $MODELFILE $PROPSFILE -prop $PROPERTY \
        -const rho=$RHO,t1=$T1 \
        -exportresults $RESULTDIR/Fig3-$PROPERTY-$T1-$RHO \
        > $LOGDIR/Fig3-$PROPERTY-$T1-$RHO"
    done
  done
done

# Figure 4
PROPERTY="Pb"
for B in $(seq 1 1 4) ; do
  for RHO in $(seq 0.6 0.5 4.6) ; do
    for T1 in $(seq 1 1 4) ; do
      echo "$PRISM $PRISMOPTIONS $MODELFILE $PROPSFILE -prop $PROPERTY \
        -const B=$B,rho=$RHO,t1=$T1 \
        -exportresults $RESULTDIR/Fig4-$PROPERTY-$B-$T1-$RHO \
```

```
        > $LOGDIR/Fig4-$PROPERTY-$B-$T1-$RHO"
    done
  done
done

# Figure 5
RHO="0.6"
for PROPERTY in mB Pblock mQ mO mTQ mTO ; do
  for T1 in $(seq 0 1 4) ; do
    for T2 in $(seq 5 1 8) ; do
      echo "$PRISM $PRISMOPTIONS $MODELFILE2 $PROPSFILE -prop $PROPERTY \
        -const rho=$RHO,t1=$T1,t2=$T2 \
        -exportresults $RESULTDIR/Fig5-$PROPERTY-$T1-$T2 \
        > $LOGDIR/Fig5-$PROPERTY-$T1-$T2"
    done
  done
done

# Figures 6-7
PROPERTY="APR"
for RHO in 0.6 4.6 ; do
  for T1 in $(seq 0 1 4) ; do
    for T2 in 5 8 ; do
      for D in 1 2 4 8 ; do
      echo "$PRISM $PRISMOPTIONS $MODELFILE2 $PROPSFILE -prop $PROPERTY \
          -const rho=$RHO,t1=$T1,t2=$T2,d=$D \
          -exportresults $RESULTDIR/Fig67-$PROPERTY-$T1-$T2-$D-$RHO \
          > $LOGDIR/Fig67-$PROPERTY-$T1-$T2-$D-$RHO"
      done
    done
  done
done

# Figure 8, T1 apparently 2
PROPERTY="APR"
T1=2
for RHO in $(seq 0.6 0.5 4.6) ; do
  for T2 in 5 8 ; do
    for D in 1 8 ; do
      echo "$PRISM $PRISMOPTIONS $MODELFILE2 $PROPSFILE -prop $PROPERTY \
        -const rho=$RHO,t1=$T1,t2=$T2,d=$D \
        -exportresults $RESULTDIR/Fig8-$PROPERTY-$T2-$D-$RHO \
        > $LOGDIR/Fig8-$PROPERTY-$T2-$D-$RHO"
      done
  done
done

# Figures 9,10
PROPERTY="APR"
T1=2
T2=5
D=2
for PROPERTY in "APR" "mC" ; do
  PO=0.2
  PIO=0.4
  for RHO in $(seq 4.55 0.01 4.6) ; do
```

```
    echo "$PRISM $PRISMOPTIONS $MODELFILE3 $PROPSFILE -prop $PROPERTY \
      -const rho=$RHO,t1=$T1,t2=$T2,d=$D,p_o=$PO,p_io=$PIO \
      -exportresults $RESULTDIR/Fig910-$PROPERTY-$PO-$PIO-$RHO \
      > $LOGDIR/Fig910-$PROPERTY-$PO-$PIO-$RHO"
    done
  PO=0.4
  PIO=0.8
  for RHO in $(seq 4.55 0.01 4.6) ; do
    echo "$PRISM $PRISMOPTIONS $MODELFILE3 $PROPSFILE -prop $PROPERTY \
      -const rho=$RHO,t1=$T1,t2=$T2,d=$D,p_o=$PO,p_io=$PIO \
      -exportresults $RESULTDIR/Fig910-$PROPERTY-$PO-$PIO-$RHO \
      > $LOGDIR/Fig910-$PROPERTY-$PO-$PIO-$RHO"
  done
  PO=0.000000001
  PIO=0.000000001
  for RHO in $(seq 4.55 0.01 4.6) ; do
    echo "$PRISM $PRISMOPTIONS $MODELFILE3 $PROPSFILE -prop $PROPERTY \
      -const rho=$RHO,t1=$T1,t2=$T2,d=$D,p_o=$PO,p_io=$PIO \
      -exportresults $RESULTDIR/Fig910-$PROPERTY-$PO-$PIO-$RHO \
      > $LOGDIR/Fig910-$PROPERTY-$PO-$PIO-$RHO"
  done
done

# execute the experiments in parallel with PROC processes
) | $TIME -p $PARALLEL $PROC > $LOGDIR/$LOGFILE 2>&1

done
```

Annales Mathematicae et Informaticae
43 (2014) pp. 145–156
http://ami.ektf.hu

Constrained modification of the cubic trigonometric Bézier curve with two shape parameters

Ede Troll

University of Debrecen
ede.troll@gmail.com

Submitted September 13, 2014 — Accepted December 7, 2014

Abstract

A new type of cubic trigonometric Bézier curve has been introduced in [1]. This trigonometric curve has two global shape parameters λ and μ. We give a lower boundary to the shape parameters where the curve has lost the variation diminishing property. In this paper the relationship of the two shape parameters and their geometric effect on the curve is discussed. These shape parameters are independent and we prove that their geometric effect on the curve is linear. Because of the independence constrained modification is not unequivocal and it raises a number of problems which are also studied. These issues are generalized for surfaces with four shape parameters. We show that the geometric effect of the shape parameters on the surface is parabolic.

Keywords: trigonometric curve, spline curve, constrained modification

MSC: AMS classification numbers

1. Introduction

Although classical polynomial curves, such as Bézier curve and B-spline curve still play central role in computer aided geometric design, several new curves have been developed in the last decade. The basic principle of curve design is still valid: the curve is given by user-defined points (so-called control points) which are combined with predefined basis functions. Keeping this principle in mind, the generalizations show various directions of possible improvements in theory and practice as well, applying basis functions different from the polynomial ones. In several cases the reason for this is either to provide a curve description method which can exactly describe (and not only approximate) important classical curves, which cannot be done by polynomial basis functions, or to simplify the computation of the curves and their properties. The most well-known generalization is the rational Bézier and B-spline curve [5], where rational functions are applied as basis functions, but deriving these functions one can obtain high degree rational polynomials, which may cause stability problems computing higher order derivatives.

The other way to improve the abilities of such a curve is the application of trigonometric functions. Trigonometric spline curves can also represent important curves, such as circle, lemniscate, etc. exactly, which cannot be done by polynomial curves. The theoretical fundamentals for this kind of curves have been laid in [10]. C-Bézier and uniform CB-spline curves are defined by means of the basis $\{\sin t, \cos t, t, 1\}$, that was generalized to $\{\sin t, \cos t, t^{k-3}, t^{k-4}, \ldots, t, 1\}$ [11, 12, 13]. NUAT B-spline curves introduced by Wang et al. in [15], the non-uniform generalizations of CB-spline curves. The other basic type is the HB-spline curve, the basis of which is $\{\sinh t, \cosh t, t, 1\}$, and $\{\sinh t, \cosh t, t^{k-3}, t^{k-4}, \ldots, t, 1\}$ in higher order [14, 10].

Another, not necessarily independent direction of generalization is the incorporation of shape parameters to the basis functions in order to provide additional freedom in shape adjustment. One of the earliest methods in this way is β-spline curve with two global parameters [7, 8]. Further methods have been provided by direct generalization of B-spline curves as αB-splines in [9] and [6] and recently as GB-splines in [4]. A spline curve with exponential shape parameters is defined and studied in [28]. Some alternative spline curves with shape parameters can be found in [2, 3, 29]. HB-spline curves, CB-spline curves and the uniform B-spline curves have been unified under the name of FB-spline curves in [16, 17]. The evaluation of these trigonometric spline curves are more stable than that of NURBS curves [18, 19].

In the above mentioned papers the new curve types are defined and essential properties are proved, but the more detailed geometric analysis of the curve has not been provided, however, it is of great importance in applications. 'How the shape parameters influence the shape of the curve?' and 'How the curve can be applied for interpolation problems?' are just two of these questions. Constrained modification of the curves is also a central issue of applications. These questions are studied in several papers [20, 23, 24, 28, 26, 27].

The aim of this paper is to study the geometrical properties of a recently defined new curve type. Our methods for finding paths of the curve points (Section 3) and describing constrained modification (Section 4) follow the techniques developed in [21] and [22] for B-spline and NURBS curves.

In [1] the authors defined a new type of curves called T-Bézier curve as follows.

Definition 1.1. For two arbitrarily selected real values of λ and μ, where $\lambda, \mu \in [-2, 1]$, the following four functions of t ($t \in [0, 1]$) are defined as cubic trigonometric

Bézier (i.e. T-Bézier) basis functions with two shape parameters λ and μ:

$$\begin{aligned}
b_0(t) &= \left(1 - \sin\frac{\pi t}{2}\right)^2 \left(1 - \lambda \sin\frac{\pi t}{2}\right)\\
b_1(t) &= \sin\frac{\pi}{2}\left(1 - \sin\frac{\pi}{2}\right)\left(2 + \lambda - \lambda \sin\frac{\pi t}{2}\right)\\
b_2(t) &= \cos\frac{\pi}{2}\left(1 - \cos\frac{\pi}{2}\right)\left(2 + \mu - \mu \cos\frac{\pi t}{2}\right)\\
b_3(t) &= \left(1 - \cos\frac{\pi t}{2}\right)^2 \left(1 - \mu \cos\frac{\pi t}{2}\right)
\end{aligned}$$

In the following sections we study the variation diminishing property, the effect of the shape parameters for the curve and the constrained modification abilities of this curve. Finally, the curve type is generalized for surfaces.

2. Variation diminishing

In the following, we will discuss new properties of the T-Bézier curve. In CAD systems the variation diminishing property is necessary for a Bézier curve. When we define a control polygon for a curve then we expect that the number of intersections with the produced curve will be less than or equal to the number of intersections with the defined control polygon.

Theorem 2.1. *If $\lambda < -2$ or $\mu < -2$ than the T-Bézier curve has lost its variatonal diminishing property.*

Proof. The derivatives of the basis functions with respect to a variable t are

$$\begin{aligned}
\frac{\delta b_0}{\delta t} &= \frac{1}{2}\cos\frac{\pi t}{2}\left(3\cos\frac{\pi t}{2}^2 \lambda + 4\lambda \sin\frac{\pi t}{2} + 2\sin\frac{\pi t}{2} - 4\lambda - 2\right)\pi,\\
\frac{\delta b_1}{\delta t} &= -\frac{1}{2}\cos\frac{\pi t}{2}\left(3\cos\frac{\pi t}{2}^2 \lambda + 4\lambda \sin\frac{\pi t}{2} + 4\sin\frac{\pi t}{2} - 4\lambda - 2\right)\pi,\\
\frac{\delta b_2}{\delta t} &= -\frac{1}{2}\sin\frac{\pi t}{2}\left(3\cos\frac{\pi t}{2}^2 \mu - 4\mu \cos\frac{\pi t}{2} - 4\cos\frac{\pi t}{2} + \mu + 2\right)\pi,\\
\frac{\delta b_3}{\delta t} &= \frac{1}{2}\sin\frac{\pi t}{2}\left(3\cos\frac{\pi t}{2}^2 \mu - 2\mu \cos\frac{\pi t}{2} - 4\cos\frac{\pi t}{2} + \mu + 2\right)\pi,
\end{aligned}$$

and

$$\begin{aligned}
\frac{\delta b_0}{\delta t}(0) &= -\frac{1}{2}\pi\lambda - \pi, & \frac{\delta b_0}{\delta t}(1) &= 0,\\
\frac{\delta b_1}{\delta t}(0) &= \frac{1}{2}\pi\lambda + \pi, & \frac{\delta b_1}{\delta t}(1) &= 0,
\end{aligned}$$

$$\frac{\delta b_2}{\delta t}(0) = 0, \qquad \frac{\delta b_2}{\delta t}(1) = -\frac{1}{2}\pi\mu - \pi,$$
$$\frac{\delta b_3}{\delta t}(0) = 0, \qquad \frac{\delta b_3}{\delta t}(1) = \frac{1}{2}\pi\mu + \pi.$$

With the control points $\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ the tangent vectors at the points at $t = 0$ and $t = 1$ are $\pi\left(\frac{1}{2}\lambda + 1\right)(\mathbf{p}_1 - \mathbf{p}_0)$ and $\pi\left(\frac{1}{2}\mu + 1\right)(\mathbf{p}_3 - \mathbf{p}_2)$ respectively, therefore if $\lambda < -2$, than the direction of the tangent vector at the point at $t = 0$ is opposite to the vector $(\mathbf{p}_1 - \mathbf{p}_0)$, and if $\mu < -2$, then the direction of the tangent vector at the point at $t = 1$ is opposite to the vector $(\mathbf{p}_3 - \mathbf{p}_2)$. The theorem follows. $\square$

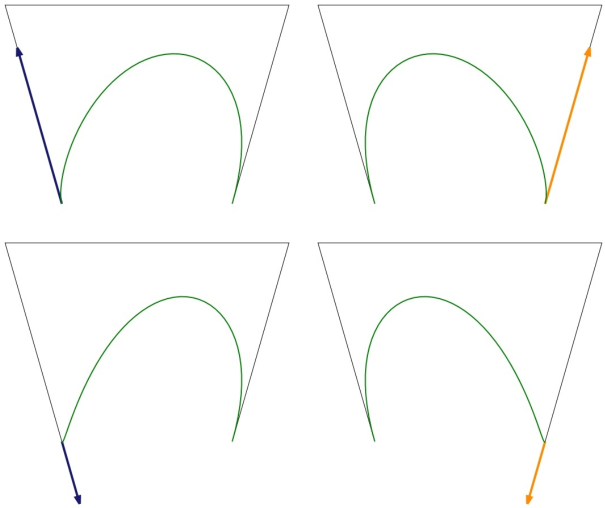


Figure 1: The tangent vectors at points at $t = 0$ with $\lambda = -1,5$ on the top left, $t = 0$ with $\lambda = -2,2$ on the bottom left, $t = 1$ with $\mu = -1,5$ on the top right and $t = 1$ with $\mu = -2,2$ on the bottom right

3. The geometric effect of the shape parameters

The basis functions of the cubic trigonometric Bézier curve are contain two arbitrarily selected real values λ and μ as shape parameters. When these parameters are changing the shape of the curve is altered too. If we have a given knot vector t_0 and one of the shape parameters is fix, than we can examine the path of the $\mathbf{T}(t, \lambda, \mu)$ point of the curve while the other shape parameter is varies between its boundaries.

Theorem 3.1. *If $t \in [0, 1]$ and $\mu \in [-2, 1]$ is a constant, then the geometric effect of the shape parameter λ is linear.*

Proof. Those basis function segments which does not include the shape parameter λ, we can expect as constants. Let

$$\begin{aligned}
c_{0_{\lambda,t}} &= \sin\frac{\pi}{2}t,\\
c_{1_{\lambda,t}} &= 1-\sin\frac{\pi}{2}t,\\
c_{2_{\lambda,t}} &= \cos\frac{\pi}{2}\left(1-\cos\frac{\pi}{2}\right)\left(2+\mu-\mu\cos\frac{\pi}{2}t\right),\\
c_{3_{\lambda,t}} &= \left(1-\cos\frac{\pi}{2}t\right)^2\left(1-\mu\cos\frac{\pi}{2}t\right).
\end{aligned}$$

With these constants we can express the basis functions of the quadratic trigonometric polynomial curve:

$$\begin{aligned}
b_0(t) &= c_{1_{\lambda,t}}^2(1-\lambda c_{0_{\lambda,t}}) = c_{1_{\lambda,t}}^2 - \lambda c_{0_{\lambda,t}} c_{1_{\lambda,t}}^2\\
b_1(t) &= c_{0_{\lambda,t}} c_{1_{\lambda,t}}(2+\lambda-\lambda c_{0_{\lambda,t}}) = 2c_{0_{\lambda,t}} c_{1_{\lambda,t}} + \lambda c_{0_{\lambda,t}} c_{1_{\lambda,t}} - \lambda c_{0_{\lambda,t}}^2 c_{1_{\lambda,t}}\\
b_2(t) &= c_{2_{\lambda,t}}\\
b_3(t) &= c_{3_{\lambda,t}}
\end{aligned}$$

The theorem follows. □

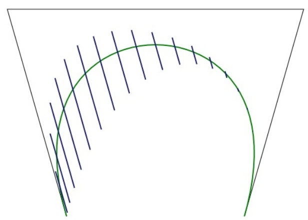


Figure 2: The geometric effect of the shape parameter λ

Theorem 3.2. *If $t \in [0,1]$ and $\lambda \in [-2,1]$ is a constant, then the geometric effect of the shape parameter μ is linear.*

Proof. Those basis function segments which does not include the shape parameter μ, we can expect as constants. Let

$$\begin{aligned}
c_{0_{\mu,t}} &= \cos\frac{\pi}{2}t,\\
c_{1_{\mu,t}} &= 1-\cos\frac{\pi}{2}t,\\
c_{2_{\mu,t}} &= \sin\frac{\pi}{2}\left(1-\sin\frac{\pi}{2}\right)\left(2+\lambda-\lambda\sin\frac{\pi}{2}t\right),\\
c_{3_{\mu,t}} &= \left(1-\sin\frac{\pi}{2}t\right)^2\left(1-\lambda\sin\frac{\pi}{2}t\right).
\end{aligned}$$

With these constants we can express the basis functions of the quadratic trigonometric polynomial curve:

$$\begin{aligned}
b_0(t) &= c_{3_{\mu,t}}\\
b_1(t) &= c_{2_{\mu,t}}\\
b_2(t) &= c_{0_{\mu,t}} c_{1_{\mu,t}}(2+\mu-\mu c_{0_{\mu,t}}) = 2c_{0_{\mu,t}} c_{1_{\mu,t}} + \mu c_{0_{\mu,t}} c_{1_{\mu,t}} - \mu c_{0_{\mu,t}}^2 c_{1_{\mu,t}}\\
b_3(t) &= c_{1_{\mu,t}}^2(1-\mu c_{0_{\mu,t}}) = c_{1_{\mu,t}}^2 - \mu c_{0_{\mu,t}} c_{1_{\mu,t}}^2
\end{aligned}$$

The theorem follows. □

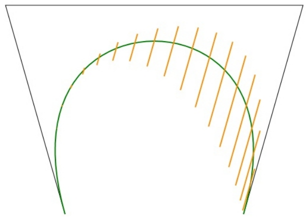


Figure 3: The geometric effect of the shape parameter μ

The two shape parameters are independent of each other so they modify the shape of the curve in separated ways. In a specific case we can examine the effect of the two parameters when they simultaneously changing their value.

Theorem 3.3. *If $t \in [0,1]$ is a constant and we run both of the shape parameters at the same time, then the geometric effect of the shape parameters is linear.*

Proof. Let $k \in R$, $\lambda, \mu \in [-2,1]$ and $\mu = k\lambda$, and consider the constants

$$c_{1_{\lambda,t}}, c_{2_{\lambda,t}}$$

from Theorem 3.1. With these constants we can express the basis functions of the quadratic trigonometric polynomial curve:

$$\begin{aligned}
b_0(t) &= c_{2_{\lambda,t}}^2(1-\lambda c_{1_{\lambda,t}}) = c_{2_{\lambda,t}}^2 - \lambda c_{1_{\lambda,t}} c_{2_{\lambda,t}}^2\\
b_1(t) &= c_{1_{\lambda,t}} c_{2_{\lambda,t}}(2+\lambda-\lambda c_{1_{\lambda,t}}) = 2c_{1_{\lambda,t}} c_{2_{\lambda,t}} + \lambda c_{1_{\lambda,t}} c_{2_{\lambda,t}} - \lambda c_{1_{\lambda,t}}^2 c_{2_{\lambda,t}}\\
b_2(t) &= c_{1_{\lambda,t}} c_{2_{\lambda,t}}(2+k\lambda-k\lambda c_{1_{\lambda,t}}) = 2c_{1_{\lambda,t}} c_{2_{\lambda,t}} + k\lambda c_{1_{\lambda,t}} c_{2_{\lambda,t}} - k\lambda c_{1_{\lambda,t}}^2 c_{2_{\lambda,t}}\\
b_3(t) &= c_{2_{\lambda,t}}^2(1-k\lambda c_{1_{\lambda,t}}) = c_{2_{\lambda,t}}^2 - k\lambda c_{1_{\lambda,t}} c_{2_{\lambda,t}}^2
\end{aligned}$$

The theorem follows. □

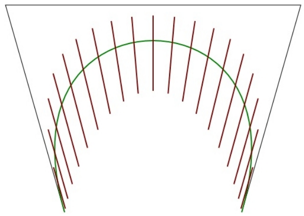


Figure 4: The shape parameters' geometric effect

4. Constrained modification

If a point $\mathbf{p}$ is given, than we need to find the values of the shape parameters $\lambda, \mu \in [-2, 1]$ with which the curve interpolates the point. Let the curve be

$$\mathbf{T}(t, \lambda, \mu) = \sum_{i=0}^{3} b_i(t)\mathbf{p}_i,$$

where $\mathbf{p}_i, i \in 0, 1, 2, 3$ are the control points.

From the inordinate case when μ is fixed and only the value of λ is changing to that case when λ is fixed every $\lambda = k\mu, k \in R$ paths are intersects the curve. Within the boundaries the curve can interpolate the point $\mathbf{p}$ and the appropriate value of the running parameter depends on the values of the shape parameters.

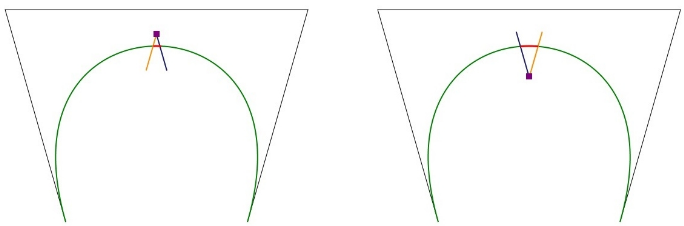


Figure 5: The appropriate section of the curve for interpolation

On the other hand, when we fix a point on the curve, then we can examine the paths we discussed above. In this case we can show the permissible area of the point of the curve.

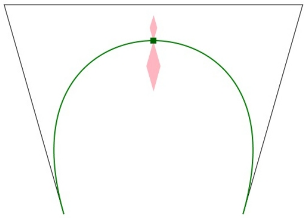


Figure 6: The permissible area of a point of the curve

If we consider the union of the permissible areas for every point of the curve, than we get the permissible area of the whole curve.

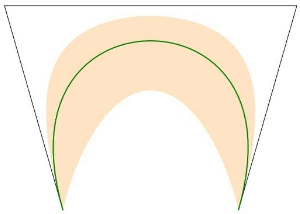


Figure 7: The admissible area of the curve

As regards the above the constrained modification is unequivocal only when $\lambda = \mu$. In this case with a numerical method we can produce the appropriate value of the running parameter t_0, whereby the produced line interpolate the given point $\mathbf{p}$. Finally, the value of the shape parameters are given from

$$\mathbf{T}(t_0, \lambda, \mu) = \mathbf{p},$$

where $\lambda = \mu$.

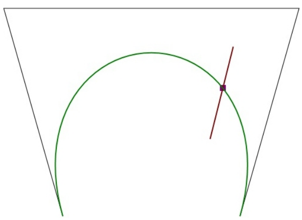


Figure 8: Constrained modification of the curve

While we discussed the algorithm of the constrained modification, we have assumed that the shape parameters are equivalent $\lambda = \mu$. This condition isn't necessary, but if $\lambda \neq \mu$, the count of the cases when the curve can interpolate a given point $\mathbf{p}$ is infinite.

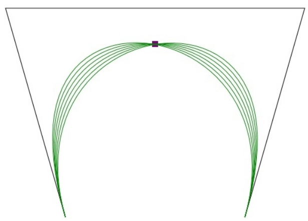


Figure 9: Constrained modification of the curve with different shape parameter values

5. Extension to surfaces

The cubic trigonometric Bézier surface made from the basis functions of the T-Bézier curve. The shape parameters also modify the face of the surface with their value, like in the case of the curve. In three dimension we expect only the case when one of the two shape parameter is fix. In the other case (when we change the other shape parameters) the proof is the same.

Theorem 5.1. *If* $t, u \in [0,1]$ *and* $\mu \in [-2,1]$ *is a constant, then the geometric effect of the shape parameter* λ *is parabolic.*

Proof. The cubic trigonometric Bézier surface is

$$\mathbf{T}(t,u) = \sum_{i,j=0}^{3} b_i(t) b_j(u) \mathbf{p}_{i,j},$$

where

$$\mathbf{p}_{i,j}, i, j \in 0,1,2,3$$

are the control points. Now we can express the coefficients.

$$\begin{aligned}
b_0(t)b_0(u) &= c_{1_{\lambda,t}}^2 c_{1_{\lambda,u}}^2 (c_{0_{\lambda,t}} c_{0_{\lambda,u}} \lambda^2 - (c_{0_{\lambda,t}} + c_{0_{\lambda,u}})\lambda + 1),\\
b_0(t)b_1(u) &= c_{0_{\lambda,u}} c_{1_{\lambda,t}}^2 c_{1_{\lambda,u}} (c_{0_{\lambda,t}}(c_{0_{\lambda,u}} - 1)\lambda^2 - (2c_{0_{\lambda,t}} + c_{0_{\lambda,u}} - 1)\lambda + 2),\\
b_0(t)b_2(u) &= c_{1_{\lambda,t}}^2 c_{2_{\lambda,u}} (-c_{0_{\lambda,t}} \lambda + 1),\\
b_0(t)b_3(u) &= c_{1_{\lambda,t}}^2 c_{3_{\lambda,u}} (-c_{0_{\lambda,t}} \lambda + 1),\\
b_1(t)b_0(u) &= c_{0_{\lambda,t}} c_{1_{\lambda,t}} c_{1_{\lambda,u}}^2 (c_{0_{\lambda,u}}(c_{0_{\lambda,t}} - 1)\lambda^2 - (c_{0_{\lambda,t}} + 2c_{0_{\lambda,u}} - 1)\lambda + 2),\\
b_1(t)b_1(u) &= c_{0_{\lambda,t}} c_{0_{\lambda,u}} c_{1_{\lambda,t}} c_{1_{\lambda,u}} ((c_{0_{\lambda,t}}(c_{0_{\lambda,u}} - 1) - c_{0_{\lambda,u}} + 1)\lambda^2\\
&\quad - (2c_{0_{\lambda,t}} + 2c_{0_{\lambda,u}} - 4)\lambda + 4),\\
b_1(t)b_2(u) &= c_{0_{\lambda,t}} c_{1_{\lambda,t}} c_{2_{\lambda,u}} ((1 - c_{0_{\lambda,t}})\lambda + 2),\\
b_1(t)b_3(u) &= c_{0_{\lambda,t}} c_{1_{\lambda,t}} c_{3_{\lambda,u}} ((1 - c_{0_{\lambda,t}})\lambda + 2),\\
b_2(t)b_0(u) &= c_{1_{\lambda,u}}^2 c_{2_{\lambda,t}} (-c_{0_{\lambda,u}} c_{2_{\lambda,t}} \lambda + 1),
\end{aligned}$$

$$b_2(t)b_1(u) = c_{0_{\lambda,u}} c_{1_{\lambda,u}} c_{2_{\lambda,t}}((1 - c_{0_{\lambda,u}})\lambda + 2),$$
$$b_2(t)b_2(u) = c_{2_{\lambda,t}} c_{2_{\lambda,u}},$$
$$b_2(t)b_3(u) = c_{2_{\lambda,t}} c_{3_{\lambda,u}},$$
$$b_3(t)b_0(u) = c^2_{1_{\lambda,u}} c_{3_{\lambda,t}}(-c_{0_{\lambda,u}} c_{3_{\lambda,t}}\lambda + 1),$$
$$b_3(t)b_1(u) = c_{0_{\lambda,u}} c_{1_{\lambda,u}} c_{3_{\lambda,t}}((1 - c_{0_{\lambda,u}})\lambda + 2),$$
$$b_3(t)b_2(u) = c_{3_{\lambda,t}} c_{2_{\lambda,u}},$$
$$b_3(t)b_3(u) = c_{3_{\lambda,t}} c_{3_{\lambda,u}},$$

where $c_{0_{\lambda,t}}, c_{1_{\lambda,t}}, c_{2_{\lambda,t}}, c_{3_{\lambda,t}}$ are presented in Theorem 3.1, and

$$c_{0_{\lambda,u}} = \sin\frac{\pi}{2}u,$$
$$c_{1_{\lambda,u}} = 1 - \sin\frac{\pi}{2}u,$$
$$c_{2_{\lambda,u}} = \cos\frac{\pi}{2}\Big(1 - \cos\frac{\pi}{2}\Big)\Big(2 + \mu - \mu\cos\frac{\pi}{2}u\Big),$$
$$c_{3_{\lambda,u}} = \Big(1 - \cos\frac{\pi}{2}u\Big)^2\Big(1 - \mu\cos\frac{\pi}{2}u\Big).$$

The theorem follows. $\square$

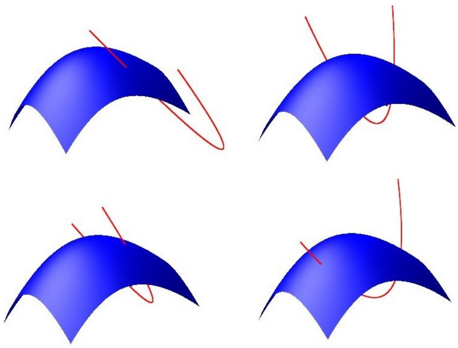


Figure 10: The geometric effect of the shape parameter λ

References

[1] Han, Xi-An, Ma, YiChen, Huang, XiLi, The cubic trigonometric Bézier curve with two shape parameters. *Applied Math. Letters* Vol. 22 (2009), 226–231.

[2] Habib Z., Sakai M. and Sarfraz M. Interactive Shape Control with Rational Cubic Splines, *International Journal of Computer-Aided Design & Applications* Vol. 1 (2004), 709–718.

[3] Habib Z., Sarfraz M. and Sakai M., Rational cubic spline interpolation with shape control, *Computers & Graphics* Vol. 29 (2005), 594–605.

[4] Guo, Q., Cubic GB-spline curves, *Journal of Information and Computational Science* Vol. 3 (2005), 465–471.

[5] Piegl, L., Tiller, W., The NURBS book. Springer Verlag (1995).

[6] Tai, C.L., Wang, G.J., Interpolation with slackness and continuity control and convexity-prservation using singular blending, *Journal of Computational and Applied Mathematics* Vol. 172 (2004), 337–361.

[7] Barsky, B.A., Beatty, J.C., Local control of bias and tension in $\beta-$ splines, *ACM Transactions on Graphics* Vol. 2 (1983), 109–134.

[8] Barsky, B.A., Computer graphics and geometric modeling using $\beta-$ splines. Springer-Verlag, Berlin (1988).

[9] Loe, K.F., αB-spline: a linear singular blending spline, *The Visual Computer* Vol. 12 (1996), 18–25.

[10] Pottmann, H., The geometry of Tchebycheffian splines, *Computer Aided Geometric Design* Vol. 10, (1993) 181–210.

[11] Chen, Q. and Wang, G., A class of Bézier-like curves, *Computer Aided Geometric Design* Vol. 20 (2003), 29–39.

[12] Zhang, J.W., C-curves, an extension of cubic curves, *Computer Aided Geometric Design* Vol. 13 (1996), 199–217.

[13] Zhang, J.W., C-Bézier curves and surfaces, *Graphical Models Image Processing* Vol. 61 (1999), 2–15.

[14] Lü, Y., Wang, G. and Yang, X., Uniform hyperbolic polynomial B-spline curves, *Computer Aided Geometric Design* Vol. 19 (2002), 379–393.

[15] Wang, G., Chen, Q. and Zhou, M., NUAT B-spline curves, *Computer Aided Geometric Design* Vol. 21 (2004), 193–205.

[16] Zhang, J.W. and Krause, F.-L., Extend cubic uniform B-splines by unified trigonometric and hyperbolic basis, *Graphical Models* Vol. 67 (2005), 100–119.

[17] Zhang, J.W. , Krause, F.-L. and Zhang, H., Unifying C-curves and H-curves by extending the calculation to complex numbers, *Computer Aided Geometric Design* Vol. 22 (2005), 865–883.

[18] Mainar, E. and Pena, J.M., A basis of C-Bézier splines with optimal properties, *Computer Aided Geometric Design* Vol. 19 (2002), 291–295.

[19] Mainar, E., Pena, J.M. and Sanchez-Reyes, J., Shape preserving alternatives to the rational Bézier model, *Computer Aided Geometric Design* Vol. 18 (2001), 37–60.

[20] Hoffmann, M., Juhász, I., Constrained shape control of bicubic B-spline surfaces by knots, in: Sarfraz, M,., Banissi, E. (eds.), *Geometric Modeling and Imaging, London, IEEE CS Press* (2006), 41–47.

[21] Juhász, I., Hoffmann, M., Modifying a knot of B-spline curves, *Computer Aided Geometric Design* Vol. 20 (2003), 243–245.

[22] Juhász, I., Hoffmann, M., Constrained shape modification of cubic B-spline curves by means of knots, *Computer Aided Design* Vol. 36 (2004), 437–445.

[23] Hoffmann, M., Y., Li, G., Wang, G.-Zh., Paths of C-Bézier and C-B-spline curves, *Computer Aided Geometric Design* Vol. 23 (2006), 463-475.

[24] Li, Y., Hoffmann, M., Wang, G-Zh., On the shape parameter and constrained modification of GB-spline curves, *Annales Mathematicae et Informaticae* Vol. 34 (2007), 51–59.

[25] Hoffmann, M., Juhász, I., Modifying the shape of FB-spline curves, *Journal of Applied Mathematics and Computing* Vol. 27 (2008), 257–269.

[26] Hoffmann, M., Juhász, I., On the quartic curve of Han, *Journal of Computational and Applied Mathematics* Vol. 223 (2009), 124–132.

[27] Troll, E., Hoffmann, M., Geometric properties and constrained modification of trigonometric spline curves of Han, *Annales Mathematicae et Informaticae* Vol. 37 (2010), 165-175.

[28] Hoffmann, M., Juhász, I., Károlyi Gy.: A control point based curve with two exponential shape parameters, *BIT Numerical Mathematics* (2014) (to appear).

[29] Papp, I., Hoffmann, M.: C^2 and G^2 continuous spline curves with shape parameters, *Journal for Geometry and Graphics* Vol. 11 (2007), 179–185.

Methodological papers

Annales Mathematicae et Informaticae
43 (2014) pp. 159–170
http://ami.ektf.hu

On intersections of the exponential and logarithmic curves

Nikola Koceić Bilan, Ivan Jelić

Faculty of Science, Department of Mathematics, University of Split
koceic@pmfst.hr

Submitted July 03, 2014 — Accepted November 08, 2014

Abstract

We consider the curves $y = a^x$ and $y = \log_a x$ and their intersecting points for various bases a. Although this problem belongs to the elementary calculus, it turns out that the problem of determining number of these points, for $a \in \langle 0, 1\rangle$, is overlooked, so far. We prove that this number can be $0, 1, 2$ or, even, 3, depending on the base a.

Keywords: exponential function, inflection, stationary point, homeomorphism

MSC: 26A06, 26A09.

1. Introduction

We consider the problem of determining the number of intersecting points of the graphs of the functions $f(x) = a^x$ and $g(x) = \log_a x$ depending on the base a. This problem is reduced to the study of solutions of the system

$$\begin{cases} y = a^x \\ y = \log_a x \end{cases} \tag{1.1}$$

which is equivalent to the equation

$$a^x = \log_a x \tag{1.2}$$

depending on a, $a \in \mathbb{R}^+ \setminus \{1\}$.

Although this problem belongs to the elementary calculus, usually, it was not considered in sufficient detail in the calculus courses on universities worldwide. Moreover, students of mathematics and many professional mathematicians are likely to think that these curves do not intersect, for $a > 1$, and meet at only one point, for $a \in \langle 0, 1\rangle$. This impression is caused by many calculus books, math teachers or professors who usually take nice bases $a = 2, e, 10..$ as standard examples for the exponential and logarithmic curves. However, in [1] and [2] can be found a solution of this problem for $a > 1$. However, for $a \in \langle 0, 1\rangle$, in [1] can be found an incorrect claim (Proposition 1) that the graphs $y = a^x$ and $y = \log_a x$ always meet at only one point. The author's conclusion seems correct at the first glance. Indeed, if we considered these curves for some standard bases $\frac{1}{2}, e^{-1} \ldots$ or if we try to make a sketch of the graphs of the functions $f(x) = a^x$ and $g(x) = \log_a x$, $a \in \langle 0, 1\rangle$, the inference, suggested by the picture, would be the same. Unexpectedly, this is not the case. Counterexample which was a motivation for this work is the base $a = \frac{1}{16}$. Namely, it holds

$$\log_{\frac{1}{16}} \frac{1}{4} = \frac{1}{2}, \quad \left(\frac{1}{16}\right)^{\frac{1}{4}} = \frac{1}{2},$$

$$\log_{\frac{1}{16}} \frac{1}{2} = \frac{1}{4}, \quad \left(\frac{1}{16}\right)^{\frac{1}{2}} = \frac{1}{4}.$$

This means that $\left(\frac{1}{4}, \frac{1}{2}\right)$ and $\left(\frac{1}{2}, \frac{1}{4}\right)$ are common points of the graphs of the functions $g(x) = \log_{\frac{1}{16}} x$ and $f(x) = \left(\frac{1}{16}\right)^x$. Since the both curves must meet the line $y = x$ at the same point we infer that there are (at least) 3 intersecting points.

The main goal of this paper is to prove:

Theorem 1.1. *The equation* (1.2)*:*
has no solutions, provided $a \in \langle \sqrt[e]{e}, +\infty\rangle$,
has exactly one solution, provided $a \in \left[\frac{1}{e^e}, 1\right\rangle \cup \{\sqrt[e]{e}\}$,
has exactly two solutions, provided $a \in \langle 1, \sqrt[e]{e}\rangle$,
has exactly three solutions, provided $a \in \left\langle 0, \frac{1}{e^e}\right\rangle$.

In order to eliminate any intuitive concluding and to avoid any possible ambiguity and incorrect inferences, which a shallow considering of the graphs might cause, we will conduct the proof of this theorem very strictly (in the mathematical sense). A necessary mathematical tool needed for the proof belongs to elementary calculus and to topology. We will split the proof of the theorem into two separate cases: $a > 1$ and $a < 1$. In the both cases we need the following corollary which is an immediate consequence of the Intermediate value theorem and some elementary facts of mathematical analysis (see e.g. [3]).

Corollary 1.2. *Let* $u\colon [c, d] \to \mathbb{R}$ *be a continuous function such that* $u(c)\, u(d) \leq 0$.

(i) *If* $u(c)\, u(d) < 0$, *then* u *has at least one zero* $x_0 \in \langle c, d\rangle$.

(ii) *If* u *is a strictly monotonic function, then* u *has exactly one zero* $x_0 \in [c, d]$.

Hereinafter, for a real function which is given by a formula we understand that the function domain is the (maximal) natural domain of that formula.

We will consider two (in)equations to be equivalent provided their solution sets coincide.

2. The case $a > 1$

The proof of this case can be given as an assignment to students of mathematics in some elementary courses. It is based on the following, several, auxiliary lemmata whose proofes we leave to the reader. Acctually, proving of these claims could be a good exercise for students in higher classes of a secondary school, providing they have sufficently ambitious math teacher.

Lemma 2.1. *If* (x_0, y_0) *is a solution of the system* (1.1), *for* $a > 1$, *then* $x_0 = y_0$.

Lemma 2.2. *If* $a > 1$, *the equation* (1.2) *is equivalent to the equation*

$$a^x = x, \tag{2.1}$$

and thus, the solution sets of (1.2) *and* (2.1) *coincide with the set of zeros of the function* $\chi_a(x) = a^x - x$.

Lemma 2.3. *If* $a > 1$, *the function* χ_a *is continuously differentiable. It is strictly decreasing on the interval* $\left\langle -\infty, \frac{1}{\ln a}\ln(\frac{1}{\ln a})\right\rangle$, *while it is strictly increasing on the interval* $\left\langle \frac{1}{\ln a}\ln(\frac{1}{\ln a}), +\infty\right\rangle$. *It reaches the global minimum at the point* $x_a^* = \frac{1}{\ln a}\ln(\frac{1}{\ln a})$.

Lemma 2.4. *Let* $a > 1$. *Then the equation* (1.2) *has: no zeros if and only if* $\chi_a(x_a^*) > 0$; *a unique zero if and only if* $\chi_a(x_a^*) = 0$; *exactly two zeros if and only if* $\chi_a(x_a^*) < 0$.

Let us interpret the previous result in term of the base a, i.e., how does a value $\chi_a(x_a^*)$ depend on a. Since the procedure is the same for all cases, it is sufficient to consider the case $\chi_a(x_a^*) < 0$. This is equivalent to $a^{x_a^*} < x_a^*$, which means

$$a^{\frac{1}{\ln a}\ln\left(\frac{1}{\ln a}\right)} < \frac{1}{\ln a}\ln\left(\frac{1}{\ln a}\right).$$

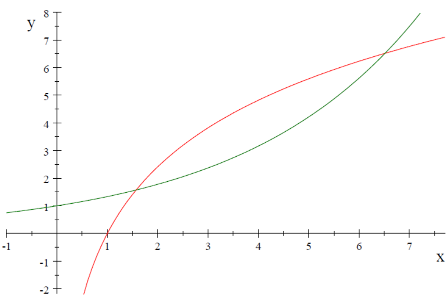
Now, one obtains, in several steps, the following mutually equivalent inequalities

$$\frac{1}{\ln a}\ln\left(\frac{1}{\ln a}\right)\ln a < \ln\left(\frac{1}{\ln a}\ln\left(\frac{1}{\ln a}\right)\right) \Leftrightarrow \ln\left(\frac{1}{\ln a}\right) < \ln\left(\frac{1}{\ln a}\ln\left(\frac{1}{\ln a}\right)\right)$$

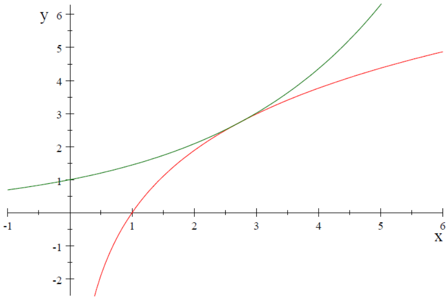
$$\frac{1}{\ln a} < \frac{1}{\ln a}\ln\left(\frac{1}{\ln a}\right) \Leftrightarrow 1 < \ln\left(\frac{1}{\ln a}\right) \Leftrightarrow \ln a < e^{-1} \Leftrightarrow a < e^{e^{-1}}.$$

Thus, the equation (1.2) has: exactly two solutions whenever $a \in \langle 1, \sqrt[e]{e}\rangle$, exactly one solution whenever $a = \sqrt[e]{e}$ (the solution is $x_0 = e$), no solutions whenever $a \in \langle \sqrt[e]{e}, +\infty\rangle$.

Example 2.5.



(a) $a = \frac{4}{3}$ $y = a^x$, $y = \log_a x$



(b) $a = \sqrt[e]{e}$ $y = a^x$, $y = \log_a x$

3. The case $0 < a < 1$

Unlike the previous case, the proof of this case is rather nontrivial. In order to make this proof easier to follow, we will split it into nine simpler claims.

Lemma 3.1. *Let* $0 < a < 1$. *then the curve* $y = a^x$ *(*$y = \log_a x$*) and the line* $y = x$ *meet at a single point* (ξ_a, ξ_a), $\xi_a \in \langle 0, 1\rangle$. *The point* ξ_a *is the solution of the equation* (1.2). *The function* $\zeta : \langle 0, 1\rangle \to \langle 0, 1\rangle$, $\zeta(a) = \xi_a$, *which assigns the point* ξ_a *to each base a, is an increasing homeomorphism whose inverse is given by the rule* $a_\xi = \zeta^{-1}(\xi) = \xi^{\frac{1}{\xi}}$.

Proof. Let λ be the real function given by $\lambda(x) = a^x - x$, for every $0 < a < 1$. Since $\lambda(0) = 1$ and $\lambda(1) < 0$, we may apply Corollary 1.2 on the function λ to infer that the curve $y = a^x$ intersects the line $y = x$. It remains to prove that they meet at exactly one point. Suppose that $(\xi, \xi = a^\xi)$ and $\left(\xi', \xi' = a^{\xi'}\right)$ are two different intersection points. There is no loss of generality in assuming $\xi < \xi'$. Since the function a^x is strictly decreasing ($a < 1$) it follows that $a^\xi = \zeta > \zeta' = a^{\xi'}$, which is an obvious contradiction.

Given an $x \in \langle 0, 1\rangle$, it is clear that, for $a = x^{\frac{1}{x}}$, it holds $a^x = x$. By examining limits $\lim\limits_{x \to 0^+} x^{\frac{1}{x}} = 0$, $\lim\limits_{x \to 1} x^{\frac{1}{x}} = 1$,and the first derivative $y' = x^{\frac{1}{x}} \frac{1 - \ln x}{x^2}$of the function

$$y(x) = \begin{cases} 0, & x = 0, \\ x^{\frac{1}{x}}, & 0 < a < 1, \end{cases}$$

one infers that it is a strictly increasing mapping on the interval $[0, 1]$ and it maps the interval $[0, 1]$ onto itself. Therefore, it is a homeomorphism and its inverse restricted to the interval $\langle 0, 1\rangle$ is the function $\zeta : \langle 0, 1\rangle \to \langle 0, 1\rangle$, $\zeta(a) = \xi_a$, exactly as asserted. □

Lemma 3.2. *If* $0 < a < 1$, *the solution set of the equation* (1.2) *is a nonempty subset of the interval* $\langle 0, 1\rangle$. *If that set is finite, then its cardinality is odd.*

Proof. Let $0 < a < 1$. Then, obviously, since the equation (1.2) is defined only for $x > 0$, it has no solution on the interval $\langle -\infty, 0]$. Further, it holds that $a^x > 0$, $\log_a x \le 0$, for every $x \in [1, +\infty\rangle$. Therefore, the equation (1.2) has no solution on the interval $[1, +\infty\rangle$. Consequently, by Lemma 3.1, the solution set of (1.2) is a nonempty subset of the interval $\langle 0, 1\rangle$. Let us assume that (x_0, y_0), $x_0 \neq y_0$, is an intersection point of the curves $y = a^x$ and $y = \log_a x$. Then, since these curves are mutually symmetric regarding the line $y = x$, they also meet at the point (y_0, x_0). Therefore, if there are only finitely many intersecting points of these curves, the number of those points which do not belong to the line $y = x$ is even. Now the statement follows by Lemma 3.1. □

Lemma 3.3. *If* $0 < a < 1$, *the solution set of the equation* (1.2) *coincides with the solution set of the equation*

$$a^{a^x} = x, \tag{3.1}$$

i.e., it coincides with the set of zeros of the real function $H_a(x) = a^{a^x} - x$.

Proof. Notice that there are no solution of (3.1) outside of the domain $\langle 0, \infty\rangle$ of the equation (1.2). Because of the injectivity of the exponential function, it is clear that (1.2) is equivalent to (3.1). □

Let us examine the functions $H_a(x) = a^{a^x} - x$ and $\varphi_a(x) = a^{a^x}$ which are, obviously, both continuously differentiable.

Lemma 3.4. *Let* $0 < a < 1$. *The function* φ_a *is strictly increasing, and the lines* $y = 1$ *and* $y = 0$ *are its horizontal asymptotes (from the right side and left side, respectively). The functions* φ_a *and* H_a *are convex on the interval* $\langle -\infty, \overline{x}_a\rangle$, *and the both are concave on the interval* $\langle \overline{x}_a, \infty\rangle$, *where*

$$\overline{x}_a = \log_a \log_a e^{-1}$$

is the common inflection point satisfying $\varphi_a(\overline{x}_a) = e^{-1}$.

Proof. Since $0 < a < 1$, it holds that $\lim\limits_{x\to+\infty} a^{a^x} = a^0 = 1$ and $\lim\limits_{x\to-\infty} a^{a^x} = a^{\infty} = 0$. Hence, the lines $y = 1$ and $y = 0$ are horizontal asymptotes of the function φ_a indeed.

Since, $\varphi_a'(x) = a^{a^x} a^x \ln^2 a > 0$, for every $x \in \mathbb{R}$,it follows that φ_a is strictly increasing. Further,

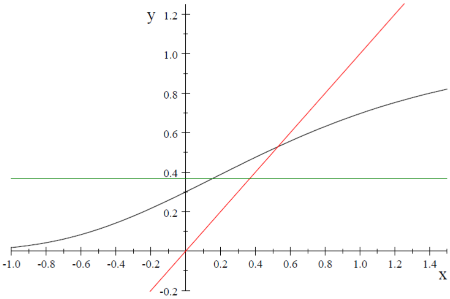
$$\begin{aligned} H_a''(x) = \varphi_a''(x) &= a^{a^x} a^x (\ln^2 a) a^x \ln^2 a + a^{a^x} a^x \ln^3 a = \\ &= \varphi_a''(x) = \underbrace{a^{a^x + x}}_{>0} \underbrace{\ln^3 a}_{<0} (a^x \ln a + 1)\,. \end{aligned}$$

Therefore, $H_a''(x) = 0$ ($\varphi_a''(x) = 0$) if and only if $a^x \ln a + 1 = 0$. Consequently,

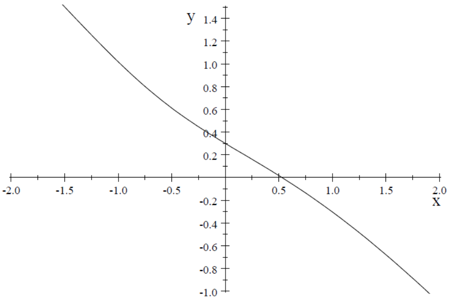
$$\overline{x}_a = \frac{1}{\ln a} \ln\left(\frac{-1}{\ln a}\right) = \log_a \log_a e^{-1} \text{ and } \varphi_a(\overline{x}_a) = a^{a^{\log_a \log_a e^{-1}}} = e^{-1}.$$

Now, it is trivial to check that $H_a''(x) = \varphi_a''(x) > 0$, for every $x \in \langle -\infty, \overline{x}_a\rangle$ and $H_a''(x) = \varphi_a''(x) < 0$, for every $x \in \langle \overline{x}_a, \infty\rangle$, which completes the proof. □

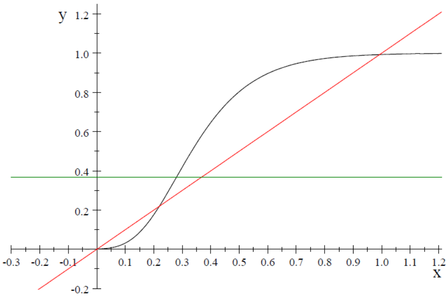
In the figures below, the graphs of the functions φ_a and H_a, for several bases a, $0 < a < 1$, are shown. In order to emphasize the inflection $\left(\overline{x}_a, e^{-1}\right)$ and solutions of the equation (3.1), the graph of the function φ_a is presented along with the lines $y = x$ and $y = e^{-1}$.



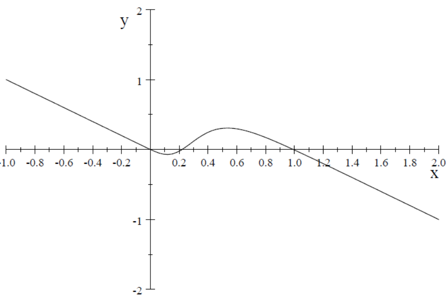
(a) $\varphi_{a=0.3}(x) = 0.3^{0.3^x}$



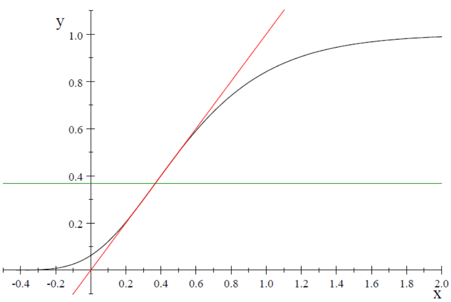
(b) $H_{a=0.3}(x) = 0.3^{0.3^x} - x$



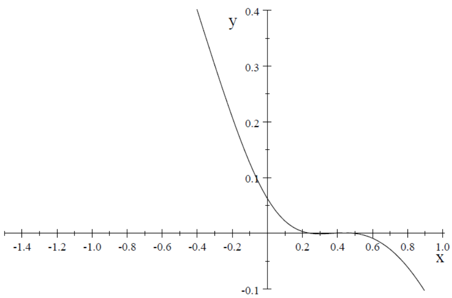
(a) $\varphi_{a=0.001}(x) = 0.001^{0.001^x}$



(b) $H_{a=0.001}(x) = 0.001^{0.001^x} - x$



(a) $\varphi_{a=\frac{1}{16}}(x) = \left(\frac{1}{16}\right)^{\left(\frac{1}{16}\right)^x}$



(b) $H_{a=\frac{1}{16}}(x) = \left(\frac{1}{16}\right)^{\left(\frac{1}{16}\right)^x} - x$

Lemma 3.5. *If $0 < a < 1$ the function H_a has at most two stationary points in* *the interval* $\langle 0,1\rangle$*, i.e., the equation*

$$H_a'(x) = 0$$

has 0*,* 1 *or* 2 *solutions in the interval* $\langle 0,1\rangle$*. If* $a < e^{-1}$*,* H_a *has at most two stationary points, while if* $a \geq e^{-1}$*,* H_a *has at most one stationary point.*

Proof. First,

$$H_a'(x) = 0 \text{ if and only if } a^{a^x} a^x \ln^2 a - 1 = 0.$$

Thus,

$$a^{a^x+x} = \frac{1}{\ln^2 a} \text{ if and only if } a^x + x = \frac{1}{\ln a}\ln(\frac{1}{\ln^2 a}).$$

We need to determine the number of solutions of the equation

$$a^x + x = \frac{1}{\ln a}\ln(\frac{1}{\ln^2 a}) \tag{3.2}$$

on the interval $\langle 0,1\rangle$. Given an $0 < a < 1$, let us define the real function u_a by $u_a(x) = a^x + x$. It holds $u_a'(x) = a^x \ln a + 1$. Now, one can easily verify that $u_a'(\overline{x}_a) = 0$, and conclude that the function u_a is strictly increasing on the interval $\langle \overline{x}_a, \infty\rangle$ and that it is strictly decreasing on the interval $\langle -\infty, \overline{x}_a\rangle$. Notice that

$$\overline{x}_a > 0 \ (\overline{x}_a < 0) \text{ if and only if } a < e^{-1} \ (a > e^{-1})$$

and that $\overline{x}_a = 0$ for $a = e^{-1}$. We infer that the function u_a reaches its global minimum at $\overline{x}_a$, and that

$$u_a(\overline{x}_a) = a^{\overline{x}_a} + \overline{x}_a = a^{\log_a \log_a e^{-1}} + \overline{x}_a = \frac{-1}{\ln a} + \overline{x}_a.$$

Hence, for $a = e^{-1}(\overline{x}_a = 0)$, we have $u(\overline{x}_a) = 1$.
Now, we infer that the number of intersection points of the curve $y = u_a(x)$, for $x \in \langle 0,1\rangle$, and the line $y = \frac{1}{\ln a}\ln(\frac{1}{\ln^2 a})$ coincide with the number of solution of the equation (3.2) in the interval $\langle 0,1\rangle$. Thus, by assuming $a \geq e^{-1}$, we obtain the strict monotonicity of the restriction of function u_a to the interval $\langle 0,1\rangle$, which implies that there are only 0 or 1 intersection points. Suppose that $a < e^{-1}$. Then, since the function u_a is strictly decreasing on the interval $\langle 0, \overline{x}_a]$ and strictly increasing on the interval $[\overline{x}_a, 1\rangle$, there are 0, 1 or 2 intersection points. □

Lemma 3.6. *If* $0 < a < 1$*, the equation* (1.2) *has either one or three solutions.*

Proof. If we assume that (1.2) has infinitely many solutions, then, by Lemmata 3.2 and 3.3, the function H_a has infinitely many zeros in the interval $\langle 0,1\rangle$. Now, by applying Rolle's theorem, one infers that H_a has infinitely many stationary points in $\langle 0,1\rangle$ which contradicts Lemma 3.5. Therefore, by Lemma 3.2, the number of solutions of the equation (1.2) is finite and odd. That number cannot exceed 3 because, by Rolle's theorem, in such a case the function H_a would have at least four stationary points in $\langle 0,1\rangle$ which is, according to Lemma 3.5, impossible. □

Lemma 3.7. *Let* $0 < a < 1$. *If the equation* (1.2) *has three solutions, then it holds* $a < e^{-e}$.

Proof. If the equation (1.2) has 3 solutions then, by Lemmata 3.2, 3.3 and Rolle's theorem, the function H_a has at least two stationary points in $\langle 0, 1\rangle$. Now, by Lemma 3.5, it follows that there are exactly two stationary points of the function H_a in $\langle 0, 1\rangle$. It implies that the equation (3.2) has two solutions in $\langle 0, 1\rangle$ and $a < e^{-1}$. Consequently, for $x \in \langle 0, 1\rangle$, the line $y = \frac{1}{\ln a}\ln(\frac{1}{\ln^2 a})$ meets the curve $y = u_a(x)$ at exactly two points, which is equivalent to

$$u_a(\overline{x}_a) < \frac{1}{\ln a}\ln\left(\frac{1}{\ln^2 a}\right) < 1, \quad a \in \left\langle 0, e^{-1}\right\rangle. \tag{3.3}$$

We propose to find solutions of this system of inequalities, i.e., to solve the system (3.3) in the terms of a. Let us put

$$t = -\frac{1}{\ln a}. \tag{3.4}$$

Notice that this substitution defines a bijective correspondence between $a \in \langle 0, e^{-1}\rangle$ and $t \in \langle 0, 1\rangle$. The replacement with t in (3.3) yields the system

$$t - t\ln t < -t\ln t^2 < 1, \quad t \in \langle 0, 1\rangle\,, \tag{3.5}$$

which we need to solve in terms of t. Now, from the first inequality $t - t\ln t < -t\ln t^2$, one obtains the following, mutually equivalent, inequalities

$$t < -t\ln t \Leftrightarrow t(1 + \ln t) < 0 \Leftrightarrow 1 + \ln t < 0 \Leftrightarrow t < e^{-1}.$$

Now, by (3.4), one infers that $-\frac{1}{\ln a} < e^{-1}$ which is equivalent to $\ln a < -e$. It follows that $a < e^{-e}$, which means that the solutions of the first inequality of the system (3.3) are all $a \in \langle 0, e^{-e}\rangle$.

Further, the second inequality $-t\ln t^2 < 1$ of the system (3.5) is equivalent to

$$-t\ln t < \frac{1}{2}, \tag{3.6}$$

which is fulfilled for every $t \in \langle 0, 1\rangle$. Indeed, by examining the function $w(t) = -t\ln t$ and its derivative $w'(t) = -\ln t - 1$, one can straightforwardly verify that w reaches the global maximum at the point $t_0 = e^{-1}$. Therefore, $w(e^{-1}) = e^{-1} < \frac{1}{2}$ implies (3.6), for every $t \in \langle 0, 1\rangle$. Consequently, the solutions of the second inequality of the system (3.3) are all $a \in \left\langle 0, e^{-1}\right\rangle$. Finally, the solution of the system (3.3) is the interval

$$\left\langle 0, e^{-e}\right\rangle = \left\langle 0, e^{-e}\right\rangle \cap \left\langle 0, e^{-1}\right\rangle. \qquad \square$$

Lemma 3.8. *For every* $a \in \left[e^{-e}, 1\right\rangle$, *the equation* (1.2) *has a unique solution. Especially, for* $a = e^{-e}$ *the solution is* e^{-1}.

Proof. By Lemma 3.6 and 3.7, it follows that, for $a \geq e^{-e}$, (1.2) has only one solution. According to Lemma 3.1, that solution is the point ξ_a such that $a^{\xi_a} = \xi_a = \log_a \xi_a$. Especially, for $a = e^{-e}$, it holds $\xi_a = e^{-1}$. Indeed, $(e^{-e})^{e^{-1}} = (e^{-e})^{\frac{1}{e}} = e^{-1}$. $\square$

Lemma 3.9. *If $a \in \langle 0, e^{-e} \rangle$, then the equation* (1.2) *has exactly three solutions.*

Proof. According to Lemma 3.6, for every $a \in \langle 0, e^{-e} \rangle$ the equation (1.2) has 1 or 3 solutions. Let us prove that the value of an inflection point $\overline{x}_a$ of the function H_a and φ_a ranges from 0 to e^{-1}, for every $a \in \langle 0, e^{-e} \rangle$. By using L'Hospital's rule, one easily evaluates the following limits:

$$\lim_{a \to 0^+} \overline{x}_a = \lim_{a \to 0^+} \frac{\ln(\frac{-1}{\ln a})}{\ln a} = \left[\frac{\infty}{-\infty} \right] = \lim_{a \to 0^+} \frac{-\ln(a)(\frac{1}{\ln^2 a})\frac{1}{a}}{\frac{1}{a}} = \lim_{a \to 0^+} \frac{-1}{\ln a} = 0,$$

$$\lim_{a \to e^{-e}} \overline{x}_a = \frac{1}{-e} \ln \frac{-1}{-e} = \frac{1}{e}.$$

We are claming that the function $\nu : \langle 0, e^{-e} \rangle \to \mathbb{R}$,

$$\nu(a) = \overline{x}_a = \frac{1}{\ln a} \ln \left(\frac{-1}{\ln a} \right),$$

is an increasing mapping. Indeed, from its first derivative

$$\nu'(a) = \frac{-1 - \ln(-\frac{a}{\ln a})}{a \ln^2 a},$$

one infers that

$$\nu'(a) > 0 \quad \text{if and only if} \quad -1 - \ln \left(-\frac{1}{\ln a} \right) > 0,$$

which is equivalent to

$$e^{-1} > -\frac{1}{\ln a} \Leftrightarrow \ln a < -e \Leftrightarrow a < e^{-e}.$$

Hence, $\nu'(a) > 0$, for every $a \in \langle 0, e^{-e} \rangle$. It follows that the function ν is an increasing and bijective mapping onto its image $\nu(\langle 0, e^{-e} \rangle) = \langle 0, e^{-1} \rangle$. Consequently, $\overline{x}_a < e^{-1}$, for every $a \in \langle 0, e^{-e} \rangle$. Now, by Lemma 3.4, it follows that $\overline{x}_a < \varphi_a(\overline{x}_a) = e^{-1}$, which implies that $H_a(\overline{x}_a) > 0$, for every $a \in \langle 0, e^{-e} \rangle$. On the other hand, it holds

$$H_a(1) = \varphi_a(1) - 1 = a^a - 1 < 0.$$

Therefore, by Corollary 1.2 and Lemma 3.3, there exists a solution x_1 of the equation (1.2), $a \in \langle 0, e^{-e} \rangle$, such that $x_1 \in \langle \overline{x}_a, 1 \rangle$. We propose to show that, beside x_1, there exists another solution x_0 of (1.2), $a \in \langle 0, e^{-e} \rangle$, such that $x_0 < \overline{x}_a$. It is sufficient to show that $\xi_a < \overline{x}_a$. First notice that $\xi_a < e^{-1}$. Indeed, since the function $\zeta : \langle 0,1\rangle \to \langle 0,1\rangle$ is an increasing bijection, and

$$\zeta^{-1}\left(e^{-1}\right) = \left(e^{-1}\right)^{\frac{1}{e^{-1}}} = e^{-e},$$

by Lemma 3.1, it follows that $\zeta\left(\langle 0, e^{-e}\rangle\right) = \left\langle 0, e^{-1}\right\rangle$. Now, from $a^{\xi_a} = \xi_a < \frac{1}{e}$, it follows that

$$\log_a a^{\xi_a} = \xi_a = a^{\xi_a} > \log_a e^{-1},$$

which implies that

$$\log_a a^{\xi_a} = \xi_a < \log_a \log_a e^{-1} = \overline{x}_a.$$

Hence, if $a \in \langle 0, e^{-e}\rangle$, the equation (1.2) has two different solutions x_1 and ξ_a. Therefore, by Lemma 3.6, (1.2) has exactly three solutions. $\square$

Remark 3.10. Notice that the point (ξ_a, ξ_a) is the common intersection point of the curves $y = \varphi_a(x)$, $y = a^x$ and $y = \log_a x$. It is interesting to consider what is happening with the inflection point $\left(\overline{x}_a, \frac{1}{e}\right)$ of φ_a and with the intersection point (ξ_a, ξ_a), and how $\overline{x}_a$ is related to the ξ_a and other solutions of (1.2), depending on a base $a \in \langle 0,1\rangle$. By the proof of Lemma 3.9, it is clear that, for $a \in \langle 0, e^{-e}\rangle$, there exist three different solutions x_2, ξ_a and x_1 of (1.2), such that

$$x_2 < \xi_a < \overline{x}_a < x_1.$$

By Lemma 3.1 and by the proof of Lemma 3.9, it follows that, while a ranges from 0 to e^{-e}, $\overline{x}_a$ and ξ_a tents from 0 to $\frac{1}{e}$. For $a = e^{-e}$, all the solutions and inflection merge into one point. Namely, $\xi_a = \overline{x}_a = \frac{1}{e}$ is the unique solution of (1.2), while the inflection point and intersection point coincide with the point $\left(\frac{1}{e}, \frac{1}{e}\right)$. "After that", for $a > e^{-e}$, they separate and $\overline{x}_a$ moves to the left and ξ_a moves to the right.

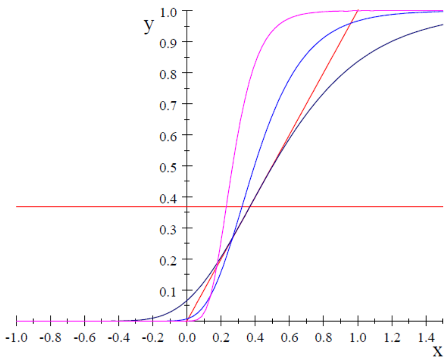


Figure 5: $y = x$, $y = \frac{1}{e}$, $y = \varphi_a(x)$, $a = e^{-10}, e^{-5}, e^{-e}$

If a ranges from e^{-e} to 1, since $\omega : \langle e^{-e}, 1\rangle \to \left\langle -\infty, e^{-1}\right\rangle$ $\omega(a) = \overline{x}_a = \frac{1}{\ln a}\ln(\frac{-1}{\ln a})$, is a decreasing bijective mapping, it follows that $\omega(a) = \overline{x}_a$ tends from e^{-1} to $-\infty$ and the unique solution ξ_a of (1.2), by Lemma 3.1, tends from e^{-1} to 1.

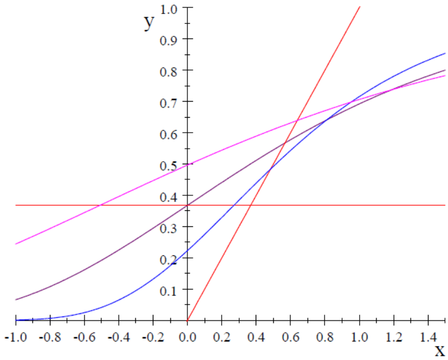


Figure 6: $y = x, y = \frac{1}{e}, y = \varphi_a(x), a = e^{-1.5}, e^{-1}, e^{-0.7}$

In the figures below an initial problem (1.1) is visualized for the bases $a = 0.3, \frac{1}{16}, 0.001$.

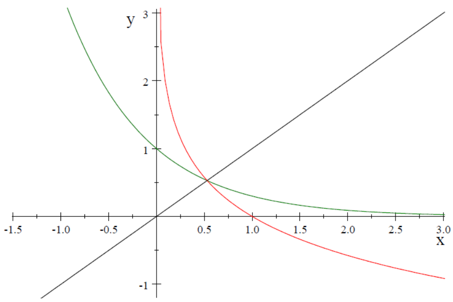


Figure 7: $a = 0.3$ $y = a^x, y = \log_a x$

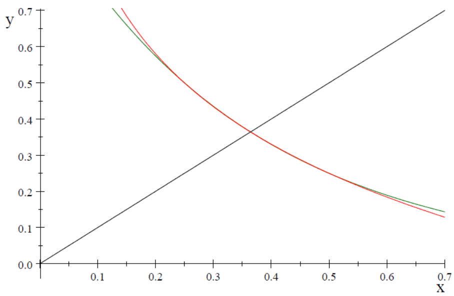


Figure 8: $a = \frac{1}{16}$ $y = a^x, y = \log_a x$

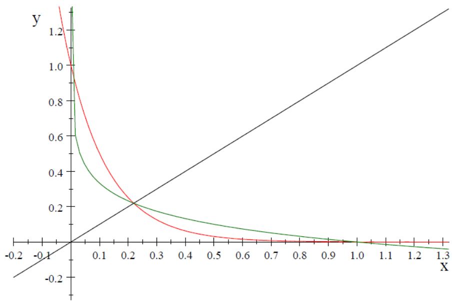


Figure 9: $a = 0.001$ $y = a^x, y = \log_a x$

The problem considered in this paper motivate us to study the equation

$$a^x = \log_b x,$$

for $a, b \in \langle 0, \infty\rangle \setminus \{1\}$ and to state the following problem:

Problem. *Determine the number of all intersecting points of the curves* $y = \log_b x$ *and* $y = a^x$ *depending on bases* a *and* b.

References

[1] A. Boukas, T. Valahas, *Intersection of the exponential and logarithmic curves*, Australian Senior Mathematics Journal 23 (1), (2009) 5–8.

[2] E. Couch, *An overlooked calculus question*, The College Mathematics Journal 33(5), (2002), 399–400.

[3] W. Rudin, *Principles of Mathematical Analysis,* McGraw-Hill,Inc., 1976.

Annales Mathematicae et Informaticae
43 (2014) pp. 171–181
http://ami.ektf.hu

Importance of spatial visualization skills in Hungary and Turkey: Comparative Studies

Rita Nagy-Kondor

University of Debrecen, Faculty of Engineering, Hungary
rita@eng.unideb.hu

Submitted October 5, 2014 — Accepted December 18, 2014

Abstract

The goal of this paper is to review research results and compare spatial abilities of prospective elementary mathematics teachers from Hungary and Turkey. The tests in a way that it contained the important components of the spatial ability (imaginary manipulation of the object, projection description and projection reading, reconstruction, transparency of the structure) were used; such as Mental Cutting Test, Purdue Spatial Visualization Test and Heinrich Spatial Visualization Test. By right of the curriculum of the two countries it can be said that for teaching the spatial geometry small time has left in both countries. The results of the survey verify that many students have problems with imagining a spatial figure and therefore to solve the spatial geometry exercises. As a future study, it is being planned to make special interactive worksheets to develop of spatial ability.

Keywords: Spatial ability, Mathematics education, Spatial tests

MSC: G20, G30, G40

1. Introduction

Spatial visualization skills are very important to success in many fields of science. Students with high scores on a mental rotation test systematically score higher on anatomy examinations [32]. According to previous studies spatial visualisation ability is a predictor for success in technical education, spatial ability development is importance in engineering training, especially for architects [1, 4, 12, 13, 20, 17, 18, 21, 22]. This ability is not determined genetically, but rather a result of a long learning process [23].

Spatial ability can be defined as the abilities of imagine the visualization of an object from different viewpoints, rotation of it and blend or integrate of the parts of the given object [9, 15, 20].

McGee [15] defines spatial ability as "the ability to mentally manipulate, rotate, twist or invert pictorially presented stimuli", McGee [15] and Maier [14] classify five components of spatial skills as

- Spatial perception: the vertical and horizontal fixation of direction regardless of troublesome information;
- Spatial visualization: it is the ability of depicting of situations when the components are moving compared to each other;
- Mental rotation: rotation of three dimensional solids mentally;
- Spatial relations: the ability of recognizing the relations between the parts of a solid;
- Spatial orientation: the ability of entering into a given spatial situation.

Spatial thinking has an important role in the teaching and learning of mathematics process. Studies showed that this ability has positive correlations with geometry and mathematics education [3, 29, 30]. Considering its important role in mathematics education, development of spatial ability by the aid of Information and Communication Technologies had great attention in the reviewed literature, especially with Dynamic Geometry Systems [6, 11; 16, 19, 26, 29].

The measurement of spatial abilities is standardized by international tests, among which the Mental Rotation Test (MRT) is introduced by Vanderberg and Kuse [31] and the Mental Cutting Test (MCT) are of greatest importance. MRT presents a criterion figure shown along with four candidate figures, two of which represent the criterion figure in a rotated position. MCT presents a 3D object with an imaginary cutting plane and five possible solutions for the cross-section shape. Heinrich Spatial Visualization Test (HSVT) and Purdue Spatial Visualization Test – Visualization of Rotation (PSVT-R) are widely used for testing the spatial ability.

Much work has been reported an analysis of MCT [1, 13, 21, 22, 25, 27], MRT [1, 4, 13, 18, 23], HSVT [5, 10, 28] and PSVT-R [1, 2, 7, 24] results of engineering students or prospective mathematics teachers, with emphasis on gender differences and attempted to find possible reasons of gender difference, concluding, that typical mistakes play central role in it. Most US researchers have used the PSVT-R to measure visualization skills; MRT and MCT are widely used in Europe and Japan.

One of paper-and-pencil test was selected to measure spatial visualization abili-ty of prospective elementary mathematics teachers: a reduced version of HSVT. This test was developed by Heinrich [10] to examine the spatial abilities of engineering graphics students. The original HSVT includes two major expert skills in spatial visualization: synthesis and decomposition. For each two basic skills she hypothesized that when mental rotation was added to these tasks at three hierarchical levels of complexity, this would render the spatial problem solving progressively more difficult [5, p.2].

The original HSVT consists of 48 items divided into 6 scales:

- synthesis without rotation;
- decomposition without rotation;
- synthesis with one-step rotation;
- decomposition with one-step rotation;
- synthesis with two-step rotation;
- decomposition with two-step rotation.

Example items of the test are given in the following figures ([28, p. 173]).

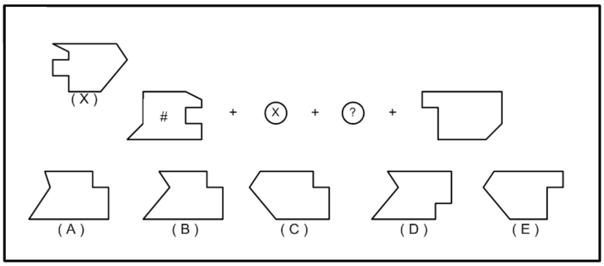


Figure 1: Example item for Synthesis section

Figure 1 expresses an example about the part of "Synthesis". Synthesize four pieces, adjusting Probe X to fit piece # and selecting one of 5 options A, B, C, D, E to replace the question mark [5, p. 3].

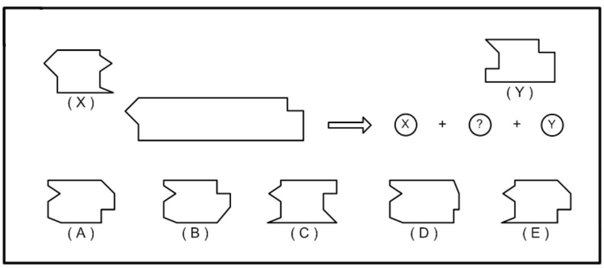


Figure 2: Example item for Decomposition section

In the Figure 2, decompose given pattern three pieces, X+?+Y, where probes X,Y may need to be adjusted, and after selecting one of 5 options, A, B, C, D, E to replace the question mark. The reduced test includes 15 items for the part of "synthesis", and 10 items for the part of "decomposition". In our work [28], following theoretical aspects of the HSVT, we used 25 items of it due to administering to the junior level prospective elementary mathematics teachers. Now it can be compared that HSVT performances [28] with new performances of PSVT-R.

Guay developed the PSVT in 1976 to determine student's ability to visualize, recognize orthograpic drawings. The PSVT includes three sections: developments, object rotations and views [8]. Most researchers and the autors use only the object rotations portion. The rotations section shows an object in two different positions. The first object is rotated on the X, Y or Z-axis, to show the rotation pattern. A second object is presented with five alternative views, one represents the second object subjected to te same rotation as the example. In our study coordinate axes were added to the first and second stimulus objects, but they were not added to the five solution choices [2]. The first stimulus object was shown in its new, rotated position. Figure 3 expresses an example about the part of PSVT-R (with coordinate axes).

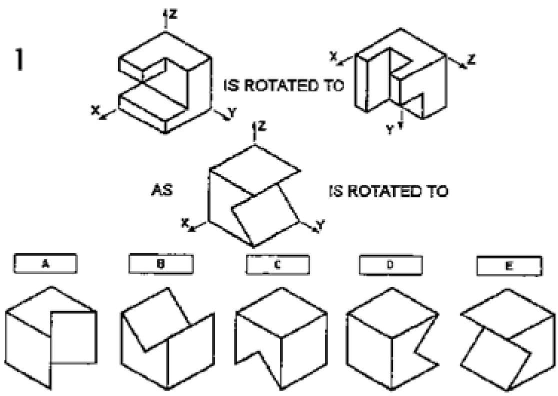


Figure 3: Example item for visualization of rotations test

In the light of the existing literature, we investigate and compare spatial visualization test PSVT-R and HSVT performances of two samples of junior level prospective elementary mathematics teachers from Hungary and Turkey enrolled to teacher training departments with the variables of gender and age. So, this work posed the following question:

- What are prospective Hungarian and Turkish elementary mathematics teachers' spatial visualization ability levels?
- Is there a significant relationship between prospective Hungarian and Turkish mathematics teachers' Heinrich and Prudue spatial visualization ability?

In the second section we report about the circumstances of the survey. The third section contains the results of the survey. The last section is the summary of the article and our experiences.

2. Background

Elementary mathematics teachers had been graduating from 3 years' education since the establishment of the Republic of Turkey (1923) to 1982. Secondary school mathematics teachers graduated from Teacher Traning Colleges and Universities.

After 1982, all education program areas came to be updated and standardized according to the framework programs of developed countries. In order to train major teachers national universities opened new teaching programs. From 1982 to 1997, mathematics teachers graduated from mathematics education programs of the Department of Science Education. Moreover, researchers graduating from the Faculty of Sciences got a mathematics teacher's degree by taking a pedagogical program organized by the education faculty.

After 1997, the mathematics education program (it was a part of science education) was divided into two teaching programs, namely elementary and secondary mathematics (four and five years long). Nowadays, it is also possible to be a secondary school mathematics teacher also for students who graduate from the faculty of sciences and get a teacher's degree after a pedagogical traning process (Figure 4).

Mathematics Teacher Training Program in Hungary can be found at the following website: http://www.math.unideb.hu/index.php?p=007&pa=000

3. The comparative survey

We made our comparative survey at the Eskişehir Osmangazi University and University of Debrecen, among prospective elementary mathematics teachers. 73 Hungarian students and 85 Turkish students took the test. All data were collected during the spring semester of 2012. Standard instructions were given to samples of junior level prospective elementary mathematics teachers from Hungary and Turkey. For the entire test, 25 minutes were given to the each sample to fill the whole HSVT test [28], and 20 minutes were given to the each sample to fill the PSVT-R test.

The tasks of PSVT-R focus on the imaginary manipulation of the solid. The task is to follow the phases of the objective activity that consist of the complex spatial transformation of the solid. The first task is the identification of the figure, and the second task is the manipulation of mental representations. Each problem is composed of a criterion figure, two one alternative and four incorrect alternatives. Correct alternative is structurally identical to the criterion, but shown in a rotated position. The subjects are asked to find the correct alternative. The PSVT-R (coordinate axes) contains 30 items of increasing level of difficulty.

Semester	Course	Credits L	Credits S
1	Computer I	2	2
	General Mathematics	4	2
2	Computer II	2	2
	Abstract Mathematics	3	0
	Geometry	3	0
3	Analysis I	4	2
	Linear Algebra I	3	0
	Physics I	4	0
	Graph Theory (Elective)	2	0
4	Analysis II	4	2
	Linear Algebra II	3	0
	Physics II	4	0
5	Analysis III	3	0
	Analtyic Geometry I	3	0
	Statistics and Probability I	2	2
	Introduction to Algebra	3	0
	Computer Based Mathematics Teaching	2	0
6	Differential Equations	4	0
	Analtyic Geometry II	3	0
	Statistics and Probability II	2	2

Figure 4: Mathematics Teacher Training Program in Turkey ([28, p. 178])

Data were analysed using the SPSS statistical analysis program. The performance of the students, correct responses given to each item of PSVT-R is presented in the Figure 5.

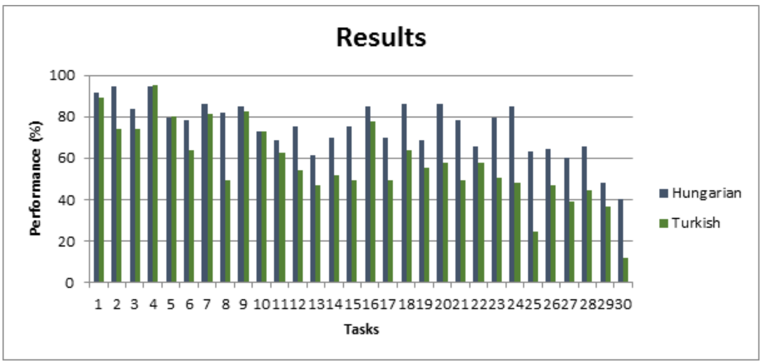


Figure 5: Students' performance

As it is shown, the most items are marked correctly: item1 with 92% (Hungarian students) and 89% (Turkish students) correct response rate; item4 with 95% (Hungarian students) and 95% (Turkish students) correct response rate; item2 with 95% (Hungarian students) and 74% (Turkish students) correct response rate and item9 with 85% correct response rate by Hungarian students and item4 with 82% correct response rate by Turkish students. The difference is remarkable between Hungarian students and Turkish students: item25 with 38%; item24 with 37% difference rate. Turkish sample performed better than Hungarian sample in items 4, 5 (the biggest difference rate is 1%).

Hungarian sample's performance is better than Turkish sample do is the distribution of the PSVT-R scores. Figure 6 gives us the results with respect to distribution of the scores.

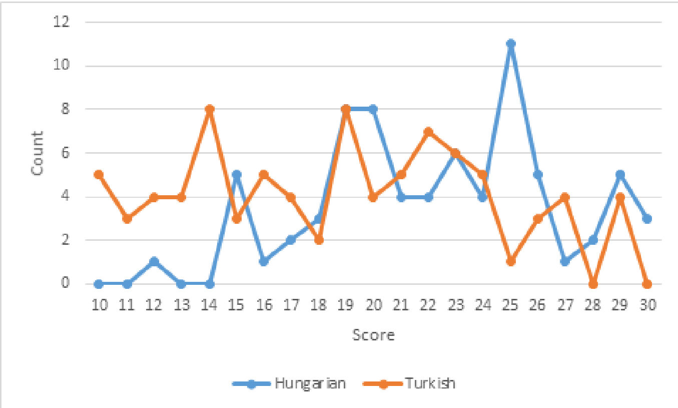


Figure 6: Distribution of the scores PSVT-R

Figure 6 shows that, while there are 27 Hungarian junior prospective elementary mathematics teachers performed 25 and greater scores, in Turkish sample there are 12 prospective teachers. And none of Turkish prospective teachers did give correct responses for the whole PSVT-R while 3 Hungarian did.

Means, standard deviations of spatial visualization ability and statistical differences of each group are analyzed in terms of descriptive statistics. The results appear in the Figure 7.

Group	f	All	Group Difference
	n	73	
Hungarian	M	22.26	
	SD	4.33	$p<.05$
	n	85	
Turkish	M	18.82	
	SD	5.36	

Figure 7: Mean scores of each sample and statistical differences

Investigating of each sample's and all subjects' means and standard deviations, we find that Hungarian junior level prospective elementary mathematics teachers mean score of PSVT-R is 22.26 (SD=4.33), and mean score of Turkish sample is 18.82 (SD=5.36). According to these results, it can be said that Hungarian and Turkish prospective elementary mathematics teachers have adequate spatial visualization ability.

Figure 7 also shows that there is a significant difference ($p<.05$) between mean scores of spatial visualization of Hungarian and Turkish junior level prospective elementary mathematics teachers. Hungarian sample performed better than those Turkish did at PSVT-R.

Additionally, there is a significant relationship between elementary mathematics teachers' scores in PSVT-R and HSVT ($p<.01$).

We found significant difference between prospective Hungarian and Turkish elementary mathematics teachers' scores of PSVT-R in favor of Hungarian sample. We think that one of the reasons may be teacher training programs. We give mathematics area courses in Figure 4. We compared the syllabi, one can see that Hungarian sample takes more lectures related to computer and geometry which may develop their geometrical and spatial reasoning. Similar findings are also observed in our previous study interpreted by HSVT [28] and MCT [27]. Moreover, results of the present study support the related literature. It is well known that to develop spatial ability, researchers suggest activities including isometric and technical drawings, computer applications and use of geometric manipulative in the teaching process [11, 16, 19, 20, 26]. In the mentioned courses there are a lot of applications need the use of spatial thinking. Therefore, suffice it to say that the related literature supports results of the present study.

4. Conclusion

In this work, we compared PSVT-R and HSVT performances of prospective Hungarian and Turkish mathematics teachers. There was a significant difference between mean scores of spatial visualization of Hungarian and Turkish students. As a consequence, the following conclusions were obtained.

The results of the survey verify that many students have problems with imagining a spatial figure and therefore to solve the spatial geometry, PSVT-R exercises. So it would be very useful in the high schools and in the university training as well, if we devote more time for spatial ability, for summarizing the spatial geometry knowledge, for solving spatial geometry tasks. According to these results, mathematics teacher curriculums may be updated.

In the related literature there are various factors effecting spatial ability. In order to make further interpretations about the obtained results, we will analyse each group's data qualitatively in terms of prepared spatial visualization, mental rotation and spatial orientation tasks. Studies suggest that interactive animation and virtual solids are promising tools for training spatial thinking in undergraduates [11, 16, 19, 24, 26]. Similar studies were conducted and concluded that students' education of preschool, primary, middle and secondary school are also important in the development of spatial ability [3, 26, 30]. Future work will be comparing of curriculums from preschool to university level and deal with another variables such as preschool education and spatial experience. It would be useful to focus on task based student interviews to reveal the student's spatial problems. Moreover, some comparative studies with self-report measures [26] may yield concrete elements to evaluate the overviewed results.

Acknowledgements. I would to thank Dr Melih Turgut (Eskişehir Osmangazi University, Faculty of Education, Turkey) his valuable cooperation in this important topic and useful helps in the data gathering process.

References

[1] Ault, H. K., John, S., Assessing and Enhancing Visualization Skills of Engineering Students in Africa: A Comprehensive Study, *Engineering Design Graphics Journal* Vol. 74/2 (2010), 12–20.

[2] Branoff, T., Connolly, P., The Addition of Coordinate Axes to the Purdue Spatial Visualization Test – Visualization of Rotations: A Study at Two Universities, *Proceedings of the American Society for Engineering Education Annual Conference*, (1999).

[3] Bosnyák, Á., Nagy-Kondor, R., The spatial ability and spatial geometrical knowledge of university students majored in mathematics, *Acta Didactica Universitatis Comenianae* Vol. 8 (2008), 1–25.

[4] Bölcskei, A., Kovács, A. Z., Kušar, D., New Ideas in Scoring the Mental Rotation Test, *YBL Journal of Built Environment* Vol. 1/1 (2013), 59–69.

[5] Chen, K. H., Validity studies of the Heinrich spatial visualization test, *Doctoral Dissertation, Ohio State University, Ohio, USA* (1995).

[6] Cohen, C. A., Hegarty, M., Visualizing cross sections: Training spatial thinking using interactive animations and virtual objects, *Learning and Individual Differences* Vol. 22/6 (2012), 868–874.

[7] Ferguson, C., Ball, A., McDaniel, W., Anderson, R., A Comparison of Instructional Methods for Improving the Spatial Visualization Ability of Freshman Technology Seminar Students, *Proceedings, IAJC-IJME International Conference* (2008).

[8] Guay, R. B., Purdue Spatial Visualisation Test: Rotations, *West Lafayette: Purdue Research Foundation* (1977).

[9] Haanstra, F. H., Effects of art education on visual-spatial and aesthetic perception: two meta-analysis, *Rijksuniversiteit Groningen* (1994).

[10] Heinrich, V. L. S., The development and validation of a spatial perception test for selection purpopes, *Master Science Dissertation, Ohio State University, Columbus, Ohio, USA* (1989).

[11] Kurtulus, A., The effects of web–based interactive virtual tours on the development of prospective mathematics teachers' spatial skills, *Computers & Education* Vol. 63 (2013), 141–150.

[12] Langley, D., Zadok, Y., Arieli, R., Exploring spatial relationships: a strategy for guiding technological problem solving, *Journal of Automation Mobile Robotics and Intelligent Systems* Vol. 8 (2014), 30–36.

[13] Leopold, C., Górska, R. A., Sorby, S. A., International Experiences in Developing the Spatial Visualization Abilities of Engineering Students, *Journal for Geometry and Graphics* Vol. 5/1 (2001), 81–91.

[14] Maier, P. H., Spatial geometry and spatial ability – How to make solid geometry solid? *In Elmar Cohors-Fresenborg, K. Reiss, G. Toener, and H.-G. Weigand, editors, Selected Papers from the Annual Conference of Didactics of Mathematics 1996, Osnabrueck* (1998), 63–75.

[15] McGee, M. G., Human Spatial Abilities: Psychometric studies and environmental, genetic, hormonal and neurological influences, *Psychological Bulletin* Vol. 86 (1979), 899–918.

[16] Nagy-Kondor, R., Spatial Ability, Descriptive Geometry and Dynamic Geometry Systems, *Annales Mathematicae et Informaticae* Vol. 37 (2010), 199–210.

[17] Nagy-Kondor, R., Technical Mathematics in the University of Debrecen, *Annales Mathematicae et Informaticae* Vol. 38 (2011), 157–167.

[18] Nagy-Kondor, R., Sörös, C., Engineering students' Spatial Abilities in Budapest and Debrecen, *Annales Mathematicae et Informaticae* Vol. 40 (2012), 187–201.

[19] Nagy-Kondor, R., Using dynamic geometry software at technical college, *Mathematics and Computer Education* Fall (2008), 249–257.

[20] Olkun, S., Making Connections: Improving Spatial Abilities with Engineering Drawing Activities, *International Journal of Mathematics Teaching and Learning* (2003), `http://www.ex.uk/cimt/ijmtl/ijabout.htm`

[21] Németh, B., Hoffmann, M., Gender differences in spatial visualization among engineering students, *Annales Mathematicae et Informaticae* Vol. 33 (2006), 169–174.

[22] Németh, B., Sörös, C., Hoffmann, M., Typical mistakes in Mental Cutting Test and their consequences in gender differences, *Teaching Mathematics and Computer Science* (2007), 1–8.

[23] Shiina, K., Short, D. R., Miller, C. L., Suzuki, K., Development of Software to Record Solving Process of a Mental Rotations Test, *Journal for Geometry and Graphics* Vol. 5 /2 (2001), 193–202.

[24] Sorby, S., A New and Improved Course for Developing Spatial Visualization Skills, *Proceedings, ASEE Annual Conference* (2001).

[25] Tsutsumi, E., A mental cutting test using drawings of intersections, *Journal for Geometry and Graphics* Vol. 8/1 (2004), 117–126.

[26] Turgut, M., Development of the spatial ability self-report scale (SASRS): reliability and validity studies, *Quality & Quantity: International Journal of Methodology* (2014, in press), Doi: 10.1007/s11135-014-0086-8.

[27] Turgut, M., Nagy-Kondor, R., Comparison of Hungarian and Turkish prospective mathematics teachers' Mental Cutting performances, *Acta Didactica Universitatis Comenianae* Vol. 13 (2013), 47–58. ISBN 978-80-223-3507-2

[28] Turgut, M., Nagy-Kondor, R., Spatial Visualisation Skills of Hungarian and Turkish prospective mathematics teachers, *International Journal for Studies in Mathematics Education* Vol. 6/1 (2013), 168–183.

[29] Turgut, M., Uygan, C., Spatial ability training via 3D modelling software. In E. Faggiano & A. Montone (Eds), *Proceedings of the 11th International Conference on Technology in Mathematics Teaching–ICTMT11, University of Bari, 9-12 July 2013, Italy: Università degli Studi di Bari Aldo Moro* (2013), 292—297.

[30] Turgut, M., Yenilmez, K., Spatial visualization abilities of preservice mathematics teachers, *Journal of Research in Education and Teaching* Vol. 1/2 (2012), 243–252.

[31] Vanderberg, S. G., Kuse, A. R., Mental Rotations, a group test of three dimensional spatial visualization, *Perceptual and Motor Skills* Vol. 47 (1978), 599—604.

[32] Vorstenbosch, M. A., Klaassen, T. P., Donders, A. R. T., Kooloos, J. G., Bolhuis, S. M., Laan, R. F., Learning anatomy enhances spatial ability, *Anatomical sciences education* Vol. 6/4 (2013), 257–262.

Annales Mathematicae et Informaticae
43 (2014) pp. 183–199
http://ami.ektf.hu

IT Competences: Modelling the Real World

Péter Szlávi, László Zsakó

Eötvös Lóránd University
szlavip@elte.hu
zsako@caesar.elte.hu

Submitted January 23, 2014 — Accepted December 30, 2014

Abstract

We often understand real-world phenomena through their models. To be able to do this, we need to be aware of the basics and the operation of modelling and the methods of its application. Modelling should be used not only for understanding but also for predicting real-world phenomena. The IT-specific peculiarity of modelling is that even the operation of models is a complex creative process.

Keywords: Model, Simulation, IT Competences, Elementary and Secondary School

MSC: 68U06

Having IT competences means that one is able to apply the basic IT tools and methods for getting information and solving problems in one's everyday life, both at home and at the workplace. People with such skills are able to use their knowledge practically, to learn and operate new technologies and methods, to solve problems, to reach individual and social goals, and to make informed decisions in our information society. [1, 2]

The important IT competences are the following[1]:

- Algorithmic thinking [3]
- Data modelling
- Modelling real-world phenomena
- Problem-solving

[1]The majority of these competences appear in other disciplines as well.

- Communication skills
- Application skills
- Teamwork, cooperation skills
- Creativity
- Information literacy
- Systemic thinking

Modelling is of course not an IT competence only. Models are used in several disciplines of science, such as mathematics (for phenomena like the point, which has no dimensions, the line, which has no thickness, or infinity), physics (for phenomena like friction or entities moving without air resistance).

Blum, referring to modelling as a mathematical competence, writes: *"Modelling competence includes the following: to structure, to mathematize, to interpret and to solve problems and it includes as well the ability to work with mathematical models: to validate the model, to analyze it critically and to assess the model and its results, to communicate the model and to observe and to control selfadjustingly the modelling process."* [4]

Nevertheless, three basic differences need to be pointed out that make IT special among the school subjects:

- In IT students can make models based on real-world systems (this is possible in mathematics, and only in very rare cases in physics, too [4, 5, 6, 7][2]);
- In IT models can even be made real with the computer (and often the realization is more important than the usage of the model);
- The very operation of IT models is a complex creative activity.

In other words, IT has a tool (the computer) for making models. As a consequence, the IT models are based on data (determined by the features of real-world objects), with which the computer algorithm performs calculations, and these results inform us about the specific real-world system. [8]

One of the most efficient tools of cognition is modelling. Those who are able to clearly describe a process they experience using abstract terms are on the right track to say: "We have understood the phenomenon." By operating the model, they can gather abstract experiences, which can serve as bases for real-world experiments. Those experiments then enable them to polish the model, thus, their knowledge.

Modelling is a "schematic process," which calls for the consideration of the following. First, we need to define the abstract objects of the model, which are the

[2]Gabriella Ambrus wrote the following about mathematical modelling: *"When solving a modelling task, the focus is put on the process and procedure(s) the student has to find and implement in order to create a relation between a non-mathematical problem and some mathematical content."*

"metaphors" of the objects (or classes of objects) in the real-world system. Later, we need to determine the set of states of the objects, (one of) which will characterize them during their operation. Finally, we set the rules of state change with the appropriate algorithms. [9, 10]

Their levels:

- Defining the models of reality
- Operating the models of reality
- Comprehending reality through the models
- Making the models of reality
- Learning model making
- Predicting real-world phenomena through the use of the models
- Analyzing the models

The competence levels of mathematical modelling is more thoroughly described by Herbert Henning and Mike Keune [11]:

Level 1 – Recognize and understand modelling – is characterized by the ability:
to recognize and
to describe the modelling process,
to characterize, to distinguish and to localize phases of the modelling process.

Level 2 – Independent modelling – is characterized by the ability:
to analyze and to structure problems and to abstract quantities,
to adopt different perspectives,
to set up mathematical models,
to work on models,
to interpret results and statements of models,
to validate models and the whole process.

Level 3 – Meta-reflection on modelling – is characterized by the ability:
to critically analyze modelling,
to characterize the criteria of model evaluation,
to reflect on the cause of modelling,
to reflect on the application of mathematics.

1. Where Do We Find IT Modelling in Public Education?

Modelling can occur in many forms.

1.1. Data Modelling

"Data modelling is an abstraction process, in which the real-(micro)world facts and the data about the relations between these facts are collected and are converted in a format applicable for computer adaptation, that is, in so-called data models. Data modelling is concerned with the internal structure and relations of the data, but not with their specific values." [12]

1.2. Problem-Solving: The Models of Problems

During the process of problem–solving, we need to define the model that describes the problem, as the initial step. [13] Let us demonstrate the series of steps of modelling through a task made for a qualifier for the student olympiad (in 2012).

A farmer had three cans of milk of different capacity measures. When they are full, they have the capacity of A, B, and C liter of milk. The farmer knows that by pouring milk from one can to the other, it is possible to measure a specific quantity of milk. All the farmer needs to do is keep in mind how much milk is left in the one used for pouring and how much milk is filled in the one used to contain the pouring. In the beginning all the cans contain some milk, but the farmer wants to have all the milk for sale to be in can A. In order to minimize the time used, the farmer needs to know that the time spent on filling equals the amount of milk poured. Make a program that calculates the time and the amount of milk, the measurement of which takes the most time. [14]

In the beginning the three cans contain (a, b, c) liter of milk. We can only pour from a can that is not empty and pour into a can that is not full. After every pouring either the can, used for pouring, becomes empty, or the can, used for containing, becomes full. That is, a state (a, b, c) can turn into a $(0, b + a, c)$ or $(a - (B - b), B, c)$ state, for example. The model of the problem, therefore, is a graph, whose points are the states describing the current states of the cans and the edges are the regular pourings.

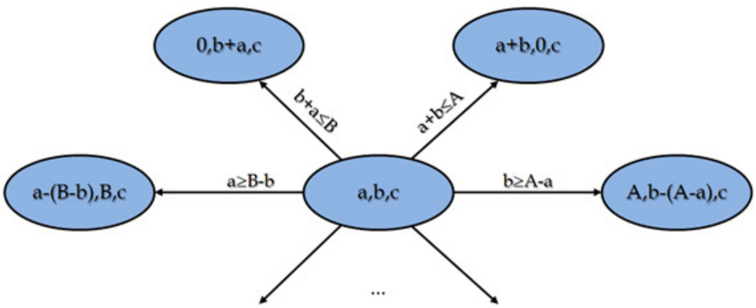


Figure 1: The graph model of the problem

The model of the problem is a data model. It is this onto which we can build a calculation model, that is, an algorithm, with whose implementation the problem can be solved in the case of any initial state.

1.3. Robots as the Models of Technical Systems

It can be a task even in grades 1 to 4 to move robots (cars, for example). These cars can resemble real cars, as they can move forward and backwards, or turn left and right. Nevertheless, they can be different as well: often their turning is not executed by manipulating a steering wheel but by switching the engine of the side wheels on and off.

In the case of managing toy trains and traffic lights, a very similar technical system model is in operation. As the main focus is on their operation, not on their structure, the robots can be models not only of technical but biological systems as well.

This issue is relevant in IT when we want to program robots. In such a case, a sensor of the robot converts the signs perceived from the environment into data, then our program makes a calculation, and the results will be used to control the robot.

1.4. Simulation Models

To understand the operation of real-world systems, we are often able to use simulation models. [9] Simulation models operate very similarly to real-world systems, but they function within a virtual world, not in reality, as opposed to the previous type. While technical models require us to build the model, simulation models require us to write a program.

What do we mean by computer simulation? Its essence, in short, is the following: it is the model of an examined universe (may that be biological, chemical, or relating to any other scientific discipline, including economic micro-worlds), which lies on the stochastic state changes of a discrete object and gets embodied in a program. In other words, the program is the tool which the users, that is, the experimenters hold in their hands to compare the ideas of the modeller with the facts of reality.

It is worth to note that simulation has a significant "advantage" compared to the regular, abstract tool of mathematical modelling. In the latter, a serious abstraction process is involved to obtain the mathematical variables, among which some kind of formal mathematical relation needs to be found, which is another abstraction process. (In addition, based on the number and relation of the different parameters and variables we need to recognize which specific mathematical subfield is to be utilized, starting from linear equations and equations systems, through the non-linear, to the ordinary and partial differential equation, or the tools of stochastic processes.) Contrary to this, simulation modelling "only" requires us to "copy" real-world relations. A further difference is that the model keeps the dynamics and naturalness of the real-world system; it is interaction that expands the possibilities of the model through the temporal changeability of the context.

1.5. Computer Model

A model that models the computer itself is useful because it can demonstrate and help understand its operation, of course if it adapts to the level of the audience. One such model is exemplified by [15]. Two "peculiarities" can be spotted in its adaptation:

1. The model in reality is a series of models, as it aims to demonstrate the real "evolution" of the computer, following the principle of historicity in the enumeration of the different computer models.

2. The models go beyond technical "novelties"; they touch upon several seemingly "secondary" features. They mention, for example, the programming process, the relevance and role of operation systems, and some of the problems of parallel calculation (in a level still comprehensible for kids).

1.6. Network Models

The physical manifestation of computer networks are partly covered by network topologies (ring, hub, etc.), which are basically graph models of the network structure. [16]

1.7. Communication Models

A classical network communication model is the ISO OSI (Open System Interconnection) model, which describes the relations of the specific machines within the computer systems on several levels. Even though today the protocols used are not directly connected to this model anymore, it is still very useful for understanding communication and reviewing tasks. [16]

1.8. Computer Games as Real-World Models

While in the case of simulation models, as analyzed in chapter 1.4, we can control the model from an omnipotent position, changing the parameters that characterize the modelled world, the genre of computer games grants us with a more or less cooperational position in shaping the modelled events. When talking about computer games, we mean both online, multi-peered applications and "regular" programs. Among the participants we can find real people, next to a compulsory artificial intelligence, who is either only a mediator of the "messages" of the human players or functions as a real playmate, fighting for the same goals.

These games are models of the real world, idealized for a certain goal. This goal can be to manage within a human community, but it can also be to survive within an abstract world organized around certain rules. To quote examples to the former type, we could list the myriads of online role-play games, for instance the well-known WoW (World of Warcraft). As the other extreme of computer games, we could mention programs simulating natural or social processes (SimEarth, SimCity, and Civilization, among others) or the computerized versions of "classical" board and card games.

Note that the above listed "social" games do resemble reality, but often times they lack the intention, characteristic of modelling, to mimic real conditions. In this sense, they are not models of reality. They are simulations, but the world we build for the simulation to take place is virtual (where events and objects are called the same as in the real world, and they might even look alike). Since the context is virtual, the results do not have to be related to real conditions. Despite this, the experiences are still useful.

Computer games are worth mentioning for two reasons. On the one hand, they are playful and as such highly motivating; thus, they can serve as specifically effective educational tools. On the other hand, they can be used as programming tasks. Adjusted to the programming competence of the students, we can set tasks related to the game, which students will be enthusiastic to solve given its playful nature.

2. The Methodology of Modelling

The **simulation models** and the **models of technical systems** (which also function to simulate operation) can be used within the subject of IT as the models of real-world processes. Therefore, we are going to deal with them when modelling the real world.

A **model** is a schematic notion made generally for understanding the operation of a complicated, not thoroughly known system, from which new correlations can be drawn, or which enables us to describe the phenomena of this system mathematically. The model usually reflects only the main features of the real-world system, in a simplified way. Which counts as the main features depends always on the purpose of the model. Therefore, by models we mean – mathematical – constructions that describe the observed phenomena. Such – mathematical – constructions are justified exclusively and exactly by their operation.

In the following we are going to use the notions of **modelling** and **simulation** always according to what aspects we want to refer to modelling from.

Modelling: the process of making a model. **Simulation**: the process of using a model.

The modern "model method" relates very closely to reality. The first usable and successful models were made in physics; such was the ideal gas, the perfect liquid, the point-like particle, the mathematical pendulum, or the atomic model. Physics was followed a lot later by chemistry, biology, and earth sciences.

The steps of modern model method [17]:

- *Gathering experiences through observation.*
- *Making a model to understand the experiences.*
- *Predicting the unknown phenomena with the help of the model.*

- *Checking the validity of the prediction with experiments and determining the validity limit of the model.*
- *Solving practical tasks with the help of the model within its validity limit.*
- *Developing, modifying, or replacing the model for understanding the phenomena beyond the validity limit.*

The model, naturally, needs to be monitored. What is even more important than this, however, is that it needs to be adjusted and changed if new aspects come up.

In the digital world, it is not surprising if we add the making of the computer model (that is, the program) to this and perform the prediction with the help of operating this computer model. Our activity is characterized by a process which consists of the following objects and operations [9]:

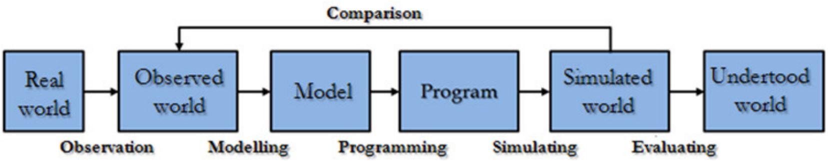


Figure 2: The process of understanding through simulation

Note: In a general sense the whole process can be called simulation. We can find a similar figure in the Werner Blum's (1996) article [18], which has been widely quoted ever since:

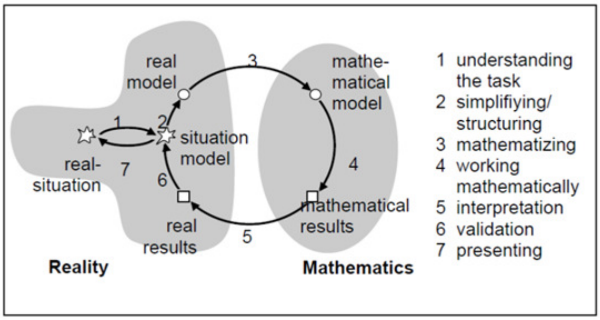


Figure 3: The process of mathematical modelling according to Blum

In every step of our activity we have to make sure that the features of the object created match those they refer to (for example, if the program has all the characteristics we expect from the model), and that the given objects themselves are valid. The final decision can be made after the comparison, which is when we can evaluate our entire activity. If the results of the simulation meet our expectations, we have reached our goal. If not, we have to check in which step we made a mistake, and correcting it, we go through the above described process again.

During modelling we use the observed data to describe the system with the help of – generally mathematical – methods, from which we can draw conclusions about the characteristics of the real-world system; we can even predict its behavior. There exist numerous examples to such formalisms (like ordinary or partial differential equations, difference equations, finite state machines, or Markov chains), but as they involve complicated mathematical apparatuses, we avoid using them in public education.

The model itself can never be the goal of modelling. The model needs to be suitable for the analysis of the modelled; thus, it must be functional and operational. Before moving on to the next activity, we always need to make sure the model fits our goal. If we apply some mathematical method, we need to check, by solving the mathematical task or providing a partial solution, if we get the expected results (which is why it was important to set the expected results in advance).

If the goal of our model is **demonstration**, for the sake of the result it is acceptable for the internal structure of the model to have some relations that are certainly different from the real phenomenon, because in this case only the result matters. (Naturally, the relations must not concern the essential mechanism of the model's operation; that is, we cannot suggest false ideas.) If, however, the goal is some kind of **analysis**, the model must follow our conceptions of the real-world system, avoiding "false analogies" (which is how the term of *horror vacui*, suggesting that bodies fall downwards due to a fear of the empty, was born).

Before making the model, we observe and gather information, formulating hypotheses about the modelled system, regarding its objects, their relations, their states, their changes, the external forces, and the overall state of the modelled system.

In the case of regular mathematical models, creating the model means defining the mathematical correlations between the parameters and state changes of the system, while for constructive methods there are other – easier – ways.

As the first step of model making, we need to define the abstract **objects** of the model, which correspond to the objects (or classes of objects) in the real-world system. This correspondence usually involves the correspondence of their states as well. To be able to speak about them on their own (like in the real-world system), they require individual existence, which is replaced by setting their state. As the next step, we need to make the algorithm describing the **state changes** (change in number, change in state) of the objects.

The significant difference from mathematical models lies in the circumstance that it is not the mathematical variables, defined as a result of serious abstraction efforts, that we need to find a mathematical relation for (which is another abstraction process). Our task is to "copy" the relations of the much more easily understandable real-world system. A further difference is that the model keeps the dynamic structure, results, and naturalness of the real world.

It is a very frequent mistake in model making that our observations are inadequate, inaccurate, and not goal-oriented enough. This actually is an inevitable obstacle in every case; therefore, we need to address it. If we refuse to acknowledge these inadequacies and inaccuracies, we can cause great damage to ourselves and – if the goal of modelling is demonstration – to others too.

The algorithm correctly describes the operation of the real-world system if:

- we take a random state of the system,
- we do the correspondences in the model,
- we create a future state of the model with the help of the algorithm,
- we find the real-world correspondent of the model state (result),
- and the result "matches" the real-world state (in deterministic cases it means equation, while in stochastic cases it is equal distribution).

The model–modelled relation is similarity: the model is similar to the modelled only from a certain aspect. The existence of such a similarity is crucial, since this guarantees that by knowing the state of the model we can define the state of the modelled too (certainly or with great probability – statistic models).

Of "whole models," resembling the modelled completely, there exists only one; the modelled itself. As a consequence, we always need to determine from which aspect the model needs to resemble the modelled. It can also occur that modelling is possible only through a very rough approximation, but this, as Hans-Wolfgang Henn points out, is not a problem. "*Models for a real problem can be more or less suitable. One should never talk about 'right' or 'wrong'. For example, it does not make sense to call Newton's model of physics 'incorrect' and Einstein's model 'correct'.*" [19]

It is an important quality that the above similarity is equivalence relation, mathematically speaking. As a consequence, the model of the model is the model of the modelled too, provided that we followed the same principles when making it. Like this, we can guarantee, even in the case of a long abstraction process, that we are still talking about reality. Another expected quality is that the modelled is the model of its own model, which means that the relation is valid only if the events of the model can take place in the real-world system as well.

3. Modelling for Different Age Groups

Development of the models: the older the students are, the more realistic, the more profound, and the more structured the models in use will be.

3.1. Tasks for Grades 1 to 4

The models of the first four grades come from the world of games: Lego, toy cars, and model railways are natural models for children.

In other words, models of technical systems (like Lego robots) can be used for this age group. These models can be grouped into two major types: moving robots or robots parking objects.

Both types can be controlled, first without any perception of the external world. As a second step, we can equip them with sensors, so we make their operation dependent of some external condition (for example, the car should not bump into the wall; the car should follow a path on the floor; the robot that parks cars should load objects of certain height on top of each other).

The shared characteristic of these models is that they are all programmable. Programmability, however, can mean a simple puzzle game, in which we graphically edit the program units together in order to form the appropriate program. [20]

3.2. Tasks for Grades 5 to 6

The modelling of reality, suitable for this age group, can be of two kinds. On the one hand, we can continue the application of robots as the models of real-world systems, within more complex tasks, possibly involving programming solutions as well. Our model (in this case, the robot) operates within a real-world environment, reacting to all changes and events of it. Nevertheless, the robot can be perceived like a limited automaton: it is exposed to the effects of the environment, to which it reacts with the change in its state components; this, in turn, affects its environment.

On the other hand, computer simulation models, more distant from reality, can also appear; their function is to simulate some kind of phenomenon with the help of the computer. To be able to do this, the application of random number generators is indispensable. The computer modelling of simple random phenomena (like dice, screen-walk, and random figures) becomes possible.

Note that we consider it important that the simulation of random phenomena, or even more random games, precedes the simulation of non-random events.

Here the simulation model does not relate to a real-world system, at least not directly.

3.3. Tasks for Grades 7 to 8

The models, used for understanding the operation of real-world systems, not for the joy of playing, can come to the foreground with this age group. We can even begin to introduce the computer modelling of simple random – and non-random – natural phenomena.

For this age group, the model becomes the simplified copy of reality (as we try to mimic reality as closely as possible); we deal primarily with defining the model and realizing it with the help of the computer.

Extending the tasks with robots requires us to introduce o new terminology, namely, the "model of the model." It is difficult – not only time-consuming but occasionally even dangerous – to test complicated programs with robots. This is when it is useful to simulate robots with the help of the computer. Only when we tested several types of robot control and eliminated the occurring mistakes can we begin to test the finalized program with the robot. In other words, the robot is a simplified model of a real-world object, while the robot simulation program is the model of the model.

Side-note: In the case of both controlling robots and simulating natural phenomena, the structure of the algorithm (program) varies significantly from what we are first introduced to in the classical programming education. Traditionally, we assume a deterministic, sequential implementation, whereas robots and interactive simulations react to the effects coming from the environment. According to the classical program structure, the program is composed of three elements: scanning data, calculating results, and showing results. Both robot control and simulation are based on online algorithms, which means that we obtain and communicate new data during the calculation.

According to Eigen and Winkler's book, we can interpret the computer simulation of natural phenomena like a board game: *"Play is a natural phenomenon that has guided the course of the world from its beginnings. It is evident in the shaping of matter, in the organisation of matter into living structures, and in the social behaviour of human beings.... Every game has its rules that set it apart from the surrounding world of reality and establish its own standards of value."* [21] The elements of the real-world systems we want to model can be grouped into classes, and the number of these elements (and sometimes the spatial distribution and pattern of the classes) determines the state of the system.

The specific units of the system can be in a finite state (discrete state model), and the number of units can also be only finite. The future (next moment) state of each unit is determined by the current state of the elements and the parameters of the system. We need to store information about the specific units, for which tables would be appropriate. If the spatial location of the units is irrelevant, we can use vectors; if the physical proximity of the units is relevant, however, two-dimensional tables (matrices) would be the better choice. We need to fill out the table based on some initial distribution, generally randomly. After this the changes, that is, the events can start.

We can define the future state deterministically, provided that we have a direct rule for calculation, or randomly. In the latter case, several calculation rules can be assigned to the specific units, coupled with a calculation rule about how likely its application should be.

The challenge of translating the parallelism of real-world systems into the sequential von Neumann computer can be overcome in two ways:

- Event stepping: we constantly monitor the on-going events. In a given moment one (or more) event(s) can occur. The question is how we choose that one event.

- Time stepping: we monitor the events with discrete intervals. In a given moment every unit of the system is "in an action." The challenge is to set an adequate order to the events.

It could be interesting for students of the Logo programming languages to get acquainted with the simulation environment of NetLogo. [21]

3.4. Tasks for Grades 9 to 10

For this age group, the scientific simulation can extend and cover all subjects of science, including biology, chemistry, physics, and geography. The obvious question arises: Which subject should deal with such simulations? Our answer is that both IT and the specific science subject. To explain IT's part we should note that simulation applies general patterns which are independent from the specific discipline, such is Eigen and Winkler's classical book [22]. In essence, its realization is primarily a programming task.

In addition to the simulation of random phenomena, we can also introduce the simulation of deterministic events too. As a start, it is wise to choose phenomena whose temporal display is spectacular. A possible example for a very simple task is to move a point on the screen with a given speed in a given direction, make the point bounce off at the edges of the screen, due to friction make it get slower and slower, apply the downward force of gravitation... and the instructions can go on.

The idea behind such simulations is that we have a given number of elements, with a given set of characteristics about their motion. We break up the time into smaller units and we calculate each element's state in the next time unit. We delete the previous position of the elements from the screen and draw them in their new one. And then comes the next time unit. (This approach can lead us also to simulation games, where, for example, a player has to catch the point, described in the previous task, using some kind of tool.)

The operation of models can serve the goal not only of understanding but also of experimenting ("what if"). [23] This type of simulation, however, belongs more to the specific discipline, not IT.

The realization of the model, with the help of the computer, is perhaps more important for this age group than experimenting with the help of the model. Experimentation becomes much more enjoyable with simulation games.

3.5. Tasks for Grades 11 to 12

It is worth continuing with subject-specific simulations but primarily in classes where students specialize in the given subject, studying it in higher weekly hours.

For this age group, two subjects, linked much more closely to real-world applications, can appear in the curriculum: Simulation of transportation systems; EconoNumbermic simulation.

Interestingly enough, it is very often the case that these grow out of the world of computer games (skill-based games, car race, strategic games).

Furthermore, prediction can be introduced as another goal of simulation for 11^{th} and 12^{nd} graders. An example for this is the demographic simulation based on the Leslie matrix we can see below.

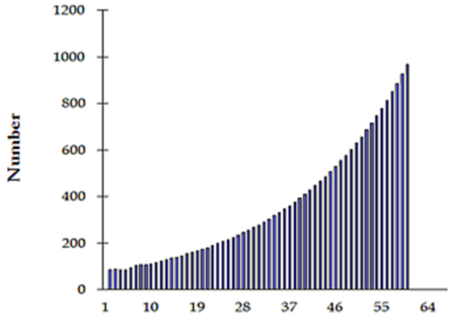


Figure 4: Demographic prediction with the help of the Leslie matrix

The above figure, just like the one below, shows that we can skip illustrating the "participants" of the simulation.

A very interesting experiment can be shown to this age group: the use of special software systems, not necessarily developed but applicable for simulation (take GeoGebra [25, 26] and spreadsheet [27] for example).

The following task, the simulation of the vibration of an object with M weight, on an "ideal" spring of 0 weight and D elastic modulus, can be easily solved with spreadsheet. We know the initial deflection (s_0), the acceleration of gravity, and the drag coefficient (d). The simulation's principle is that if we choose short dT intervals (assuming that the changes of the parameters are negligible during these dT intervals), the important state variables can be easily calculated and (for example) a state diagram can be drawn:

F **force**: $F = F_s + F_d + G$, $F_s = D \cdot s$ (*spring*), $F_d = k \cdot v$ (*drag coefficient*), $G = M \cdot g$

a **acceleration**: $a = \dfrac{F}{M}$

v **speed**: $v = a \cdot dT$

s **deflection**: $s = v \cdot dT$

s_0 **initial deflection**: $s_0 = \dfrac{M \cdot g}{D} \Leftarrow M \cdot g = s_0 \cdot D$

The relevance of modelling is well demonstrated by the fact that there have been robot programming competitions, organized for high school students, for years now,

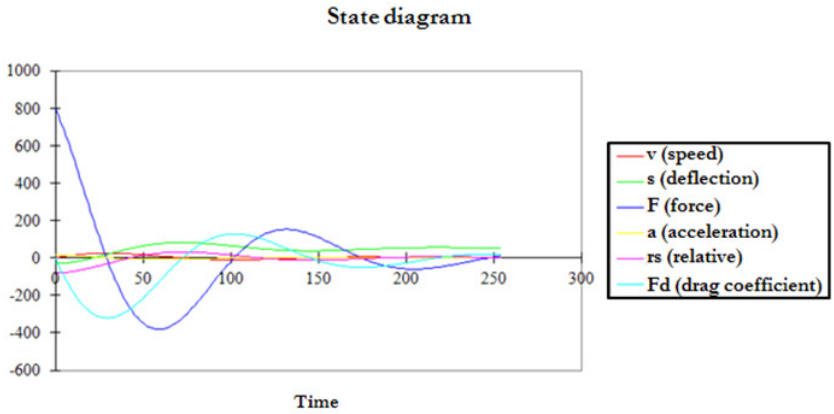


Figure 5: State diagram about spring motion

and even traditional programming competitions embrace more and more simulation tasks.

At the Imre Gyula Izsák Mathematics–Physics–Informatics Competition, initiated in 1992, it has been a practice to include, as part of the IT assignment, a simulation task connected to physics (for example, motion in the gravitational field – 1995, refraction and reflection – 1996, and so on). [28]

In 2012 even the two major national competitions in IT (Nemes Tihamér OITV, OKTV) introduced the first simulation tasks (transportation simulations - intersection and pedestrian crossing), which, despite their novelty, not only became the favorites of the competitors, but they were also solved successfully by many). [29]

4. Conclusions

The question might arise why it is the IT classroom that has to make room for all these, and why it is IT teachers, not physics, biology, literature, etc. teachers, that have to teach the above described skills. We may answer this question, partially, with a cultural historical analogy, which we attribute to Győző Kovács:

The Christian religion spread not all by itself, and it is not the Roman Pope or the 20–30 bishops whose role was the most significant in its dissemination. Instead, it was thanks to the small chapels and missionaries that the religion, with the related technological and cultural knowledge, reached every village. The "missionaries" of the IT applications are the IT teachers; only they can be able to convince the other 95% of the teachers about the possibility and the need to incorporate IT in the wide range of school subjects.

On the other hand, the majority of the models can be schematized. It means that we can make model frames (in a more trendy word, templates) that help shorten and simplify the modelling process. In addition to this, templates can enable us to categorize models according to their qualitative behavior; we can distinguish, for example, specific basic models and basic growth models. [9, 21] We believe model making, especially computer-based model making, is an important and clearly IT field; consequently, it belongs in IT education.

Computer *simulation* can lead to monumental tasks, which often require serious discipline-related knowledge as well. As a consequence, it facilitates project work in larger groups, where both IT competence and discipline-specific knowledge are needed.

References

[1] Vass, V. (ed.); Kompetenciaháló (Competence Web), *Nemzeti Tankönyvkiadó*, (2009).

[2] Szlávi, P.; Zsakó, L.; ICT competences – Algorithmic thinking, *Acta Didactica Napocensia* Vol. 5. No. 2. (2012), pp. 49–58.

[3] Horváth, Gy.; Szlávi, P. and Zsakó, L.; Informatics (ICT) competencies, *ICAI 2010 – 8th International Conference on Applied Informatics Eger, Hungary, January 27–30*, (2010).

[4] Blum, W.; ICMI study 14: Application and Modelling in Mathematics Education–Discussion Document, *Educational Studies in Mathematics* 51, (2002), pp. 149–171.

[5] Ambrus, G.; Modellezési feladatok a matematikaórán (Modelling Tasks for Maths Classes), *Matematika–Tanári Kincsestár, B 1.2, RAABE Tanácsadó és Kiadó Kft., Budapest*, (2007).

[6] Tóth,B.; Modellezési feladatok a matematikában (Modelling Tasks in Maths), *ELTE TTK*, (2010).

[7] Vancsó, Ö.; A matematikai modellezés nehézségei egy 2009-es OKTV feladat kapcsán (The Difficulties of Mathematical Modelling Based on a Task in the National IT Competition), *A Matematika Tanítása*, 2009/9, *Mozaik Oktatási Stúdió, Szeged*, (2009), pp. 30–34.

[8] Abramovich, S.; Spreadsheet-Enhanced Problem Solving in Context as Modeling, *Spreadsheets in Education (eJSiE)*, Vol. 1 Iss. 1 Article 1, (2003).

[9] Horváth, L.; Szlávi, P. and Zsakó, L.; Modellezés és szimuláció. (Modelling and Simulation), *Mikrológia 1. ELTE IK*, (2004).

[10] Szlávi, P.; Zsakó, L.; Az informatika alkalmazási típusai a közoktatásban (The Application Types of IT in Public Education), *Informatika a Felsőoktatásban'96 – Networkshop'96, 1996 August 27–30*, (1996), pp. 534–543.

[11] Henning, H.; Keune, M.; Levels of modelling competence – Modelling and Applications in Mathematics Education, *New ICMI Study Series*, Vol. 10, (2007), pp. 225–232.

[12] Bércesné Novák, Á.; Az adatmodellezés szintjei (The Levels of Data Modelling), (2013)
Available at `http://www.inf.u-szeged.hu/oktatas/Tempus/abkr2.doc` (accessed 31 December 2013).

[13] Geda, G.; Hernyák, Z.; Algoritmizálás és adatmodellek (Algorithms and Data Models), *Kempelen Farkas Hallgatói Információs Központ*, (2013)
Available at `http://www.tankonyvtar.hu/hu/tartalom/tamop425/0038_informatika_Geda_Gabor_Hernyak_Zoltan-Algoritmizalas_es_adatmodellek/ch03.html` (accessed 31 December 2013).

[14] Informatikai Diákolimpiák válogató versenye. (Qualifiers for the Olymiads in Informatics), (2013)
Available at `http://tehetseg.inf.elte.hu/valogatok/valogatok_main.html` (accessed 31 December 2013).

[15] Szlávi, P.; A számítógépről népszerűsítő stílusban (About the Computer in Promotional Style), *Mikrológia 5. ELTE TTK Informatikai Tanszékcsoport*, (1988).

[16] Tanenbaum, A. S.; Computer Networks, *Pearson Education, Inc. Publishing as Prentice Hall PTR Upper Saddle River, New Jersey*, (2003).

[17] Dr. Huzsvai, L.; Kutatói pályára felkészítő akadémiai ismeretkörön alapuló tananyagfejlesztés – Környezet- és természetvédelem ismeretkörben. (Curriculum Development for Researchers – Nature and Environment Protection), (2013)
Available at `http://www.tankonyvtar.hu/hu/tartalom/tamop425/0032_kornyezetved_termved_kutatoi/ch02.html` (accessed 31 December 2013).

[18] Blum, W.; Anwendungsbezüge im Mathematikunterricht – Trends und Perspektiven, *Schriftenreihe Didaktik der Mathematik*, Vol. 23, (1996), pp. 15–38.

[19] Hans-Wolfgang Henn; Modelling in school – chances and obstacles, *The Montana Mathematics Enthusiast*, ISSN 1551-3440, Monograph 3, (2007), pp. 125–138.

[20] van Lith, P.; Teaching Robotics in Primary and Secondary schools, *ComLab Conference 2007, November 30 – December 1, Radovljica, Slovenia*, (2007)

[21] Wilensky,U.; NetLogo, (1999)
Available at `http://ccl.northwestern.edu/netlogo` (accessed 31 December 2013).

[22] Eigen, M.; Winkler, R.; A játék (The Game), *Gondolat*, (1981).

[23] Interactive Science Simulations, *University of Colorado at Boulder, PhET project*, (2013)
Available at `http://phet.colorado.edu/` (accessed 31 December 2013).

[24] Age Structured Leslie Matrix, *Virtual Amrita Laboratories Universalizing Education. Amrita Vishwa Vidyapeetham University*, (2013)
Available at `http://amrita.vlab.co.in/?sub=3\&brch=65\&sim=183\&cnt=1` (accessed 31 December 2013).

[25] Geda, G.; Bíró, Cs. and Kovács, E.; Számítógépes szimuláció GeoGebrával (Computer Simulation with GeoGebra), *INFODIDACT 2011, Szombathely, 2011. 03.31–04.01.*, (2011).

[26] Geda, G.; Bíró, Cs.; Tánczos, T.; Számítógépes szimuláció lehetőségei (The Possibilities of Computer Simulation), *Agria Media 2011, Eger, 2011.10–12.*, (2011), pp. 426–431.

[27] Szlávi, P.; Zsakó, L.; Szimulációs modellek táblázatkezelővel (Simulation Models with Spreadsheets), *INF.O.'97 Informatika és számítástechnika tanárok konferenciája, Békéscsaba, 1997. November 20–22.*, (1997).

[28] Izsák Imre Gyula komplex természettudományi verseny (Imre Gyula Izsák Complex Science Competition), (2013)
Available at `http://www.zmgzeg.sulinet.hu/izsak/` (accessed 31 December 2013).

[29] Nemes Tihamér Országos Informatikai Tanulmányi Verseny – Programozás kategória. (Tihamér Nemes National Competition in Informatics – Category of Programming), (2013)
Available at `http://tehetseg.inf.elte.hu/nemes/index.html` (accessed 31 December 2013).