Contents

Methodological papers

ANNALES MATHEMATICAE ET INFORMATICAE 40. (2012)

ANNALES MATHEMATICAE ET INFORMATICAE

TOMUS 40. (2012)

HUNGARIA, EGER

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 Only new and unpublished material can be accepted.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 40. (2012)

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: Kis-Tóth Lajos
Felelős szerkesztő: Zimányi Árpád
Műszaki szerkesztő: Tómács Tibor
Megjelent: 2012. december Példányszám: 30

Készítette az
Eszterházy Károly Főiskola nyomdája
Felelős vezető: Kérészy László

Annales Mathematicae et Informaticae
***40** (2012) pp. 3–11*
http://ami.ektf.hu

Cofinite derivations in rings

O. D. Artemovych

Institute of Mathematics, Cracow University of Technology, ul. Cracow, Poland
artemo@usk.pk.edu.pl

Submitted December 11, 2011 — Accepted April 19, 2012

Abstract

A derivation $d : R \to R$ is called cofinite if its image $\operatorname{Im} d$ is a subgroup of finite index in the additive group R^+ of an associative ring R. We characterize left Artinian (respectively semiprime) rings with all non-zero inner derivations to be cofinite.

Keywords: Derivation, Artinian ring, semiprime ring

MSC: 16W25, 16P20, 16N60

1. Introduction

Throughout this paper R will always be an associative ring with identity. A derivation $d : R \to R$ is said to be *cofinite* if its image $\operatorname{Im} d$ is a subgroup of finite index in the additive group R^+ of R. Obviously, in a finite ring every derivation is cofinite. As noted in [3], only a few results are known concerning images of derivations.

We study properties of rings with cofinite non-zero derivations and prove the following

Proposition 1.1. Let R be a left Artinian ring. Then every non-zero inner derivation of R is cofinite if and only if it satisfies one of the following conditions:

(1) R is finite ring;

(2) R is a commutative ring;

(3) $R = F \oplus D$ is a ring direct sum of a finite commutative ring F and a skew field D with cofinite non-zero inner derivations.

Recall that a ring R with 1 is called *semiprime* if it does not contains non-zero nilpotent ideals. A ring R with an identity in which every non-zero ideal has a finite index is called *residually finite* (see [2] and [10]).

Theorem 1.2. *Let R be a semiprime ring. Then all non-zero inner derivations are cofinite in R if and only if it satisfies one of the following conditions:*

(1) *R is finite ring;*

(2) *R is a commutative ring;*

(3) *$R = F \oplus B$ is a ring direct sum, where F is a finite commutative semiprime ring and B is a residually finite domain generated by all commutators $xa - ax$, where $a, x \in B$.*

Throughout this paper for any ring R, $Z(R)$ will always denote the center, $Z_0 = Z_0(R)$ the ideal generated by all central ideals of R, $N(R)$ the set of all nilpotent elements of R, $\mathrm{Der}R$ the set of all derivations of R, $\operatorname{Im} d = d(R)$ the image and $\operatorname{Ker} d$ the kernel of $d \in \operatorname{Der} R$, $U(R)$ the unit group of R, $|R : I|$ the index of a subring I in the additive group R^+, $\partial_x(a) = xa - ax = [x, a]$ the commutator of $a, x \in R$ and $C(R)$ the commutator ideal of R (i.e., generated by all $[x, a]$). If $|R : I| < \infty$, then we say that I has a finite index in R.

Any unexplained terminology is standard as in [6], [4], [5], [8] and [11].

2. Some examples

We begin with some examples of derivations in associative rings.

Example 2.1. Let D be an infinite (skew) field,

$$A = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix},\ X = \begin{pmatrix} x & y \\ z & t \end{pmatrix} \in M_2(D).$$

Then we obtain that

$$\partial_A(X) = AX - XA = \begin{pmatrix} ax - xa & ay \\ -za & 0 \end{pmatrix},$$

and so the image $\operatorname{Im} \partial_A$ has an infinite index in $M_2(D)^+$.

Recall that a ring R having no non-zero derivations is called *differentially trivial* [1].

Example 2.2. Let $F[X]$ be a commutative polynomial ring over a differentially trivial field F. Assume that d is any derivation of $F[X]$. Then for every polynomial

$$f = \sum_{i=0}^{n} a_i X^{n-i} \in F[X]$$

we have

$$d(f) = (\sum_{i=0}^{n-1}(n-i)a_i X^{n-i-1})d(X) \in d(X)F[X],$$

where $d(X)$ is some element from $F[X]$. This means that the image $\operatorname{Im} d \subseteq d(X)F[X]$.

a) Let F be a field of characteristic 0. If we have

$$g = \left(\sum_{i=0}^{m} b_i X^{m-i}\right) \cdot d(X) \in d(X)F[X],$$

then the following system

$$\begin{cases} (1+m)d_0 & = b_0, \\ md_1 & = b_1, \\ & \vdots \\ 2d_{m-1} & = b_{m-1}, \\ d_m & = b_m, \end{cases}$$

has a solution in F, i.e., there exists such polynomial

$$h = \sum_{i=0}^{m+1} d_i X^{m+1-i} \in F[X],$$

that $d(h) = g$. This gives that $\operatorname{Im} d = d(X)F[X]$. If d is non-zero, then the additive quotient group

$$G = F[X]/d(X)F[X]$$

is infinite and every non-zero derivation d of a commutative Noetherian ring $F[X]$ is not cofinite.

b) Now assume that F has a prime characteristic p and $d(X) = X$. If $X^{p^l} - X^{p^s} \in \operatorname{Im} d$ for some positive integer l, s, where $l > s$, then

$$X^{p^l} - X^{p^s} = d(t)$$

for some polynomial $t = d_0 X^m + d_1 X^{m-1} + \cdots + d_{m-1}X + d_m \in F[X]$ and consequently

$$X^{p^l} - X^{p^s} = md_0 X^m + (m-1)d_1 X^{m-1} + \cdots + 2d_{m-1}X^2 + d_{m-1}X.$$

Let k be the smallest non-negative integer such that

$$(m-k)d_k \neq 0.$$

Then $p^l = m - k$, a contradiction. This means that $|F[X] : \operatorname{Im} d| = \infty$.

Example 2.3. Let

$$\mathbb{H} = \{\alpha + \beta\mathbf{i} + \gamma\mathbf{j} + \delta\mathbf{k} \mid \alpha, \beta, \gamma, \delta \in \mathbb{R},\\ \mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1,\ \mathbf{ij} = -\mathbf{ji} = \mathbf{k},\ \mathbf{jk} = -\mathbf{kj} = \mathbf{i},\ \mathbf{ki} = -\mathbf{ik} = \mathbf{j}\}$$

be the skew field of quaternions over the field $\mathbb{R}$ of real numbers. Then

$$\partial_i(\mathbb{H}) = \{\gamma\mathbf{j} + \delta\mathbf{k} \mid \gamma, \delta \in \mathbb{R}\}$$

and so the index $|\mathbb{H} : \operatorname{Im}\partial_i|$ is infinite. Hence the inner derivation ∂_i is not cofinite in $\mathbb{H}$.

Example 2.4. Let $D = F(y)$ be the rational functions field in a variable y over a field F and $\sigma : D \to D$ be an automorphism of the F-algebra D such that

$$\sigma(y) = y + 1.$$

By

$$R = D((X;\sigma)) = \{\sum_{i=n}^{\infty} a_i X^i \mid a_i \in D \text{ for all } i \geq n,\ n \in \mathbb{Z}\}$$

we denote the ring of skew Laurent power series with a multiplication induced by the rule

$$(aX^k)(bX^l) = a\sigma^k(b)X^{k+l}$$

for any elements $a, b \in D$. Then we compute the commutator

$$\begin{aligned}\left[\sum_{i=n}^{\infty} a_i X^i, y\right] &= \sum_{i=n}^{\infty} a_i X^i y - y\sum_{i=n}^{\infty} a_i X^i\\ &= \sum_{i=n}^{\infty} a_i \sigma^i(y) X^i - \sum_{i=n}^{\infty} a_i y X^i\\ &= \sum_{i=n}^{\infty} a_i(\sigma^i(y) - y)X^i = \sum_{i=n}^{\infty} i a_i X^i.\end{aligned}$$

If now

$$f = \sum_{i=n}^{\infty} b_i X^i \in R,$$

then there exist elements $a_i \in D$ such that

$$b_i = i a_i$$

for any $i \geq n$. This implies that the image $\operatorname{Im}\partial_y = R$ and ∂_y is a cofinite derivation of R.

Lemma 2.5. *Let $R = F[X, Y]$ be a commutative polynomial ring in two variables X and Y over a field F. Then R has a non-zero derivation that is not confinite.*

Proof. Let us $f = \sum \alpha_{ij} X^i Y^j \in R$ and $d : R \to R$ be a derivation defined by the rules

$$\begin{aligned} d(X) &= X, \\ d(Y) &= 0, \\ d(f) &= \sum i\alpha_{ij} X^{i-1} Y^j d(X). \end{aligned}$$

It is clear that $\operatorname{Im} d \subseteq XR$ and $|R : XR| = \infty$. □

In the same way we can prove the following

Lemma 2.6. *Let $R = F[\{X_\alpha\}_{\alpha\in\Lambda}]$ be a commutative polynomial ring in variables $\{X_\alpha\}_{\alpha\in\Lambda}$ over a field F. If $\operatorname{card}\Lambda \geq 2$, then R has a non-zero derivation that is not cofinite.*

3. Cofinite inner derivations

Lemma 3.1. *If every non-zero inner derivation of a ring R is cofinite, then for each ideal I of R it holds that $I \subseteq Z(R)$ or $|R : I| < \infty$.*

Proof. Indeed, if I is a non-zero ideal of R and $0 \neq a \in I$, then the image $\operatorname{Im}\partial_a \subseteq I$. □

Remark 3.2. If δ is a cofinite derivation of an infinite ring R, then $|R : \operatorname{Ker}\delta| = \infty$.

In fact, if the kernel $\operatorname{Ker}\delta = \{a \in R \mid \delta(a) = 0\}$ has a finite index in R, in view of the group isomorphism

$$R^+ / \operatorname{Ker}\delta \cong \operatorname{Im}\delta,$$

we conclude that $\operatorname{Im}\delta$ is a finite group.

Lemma 3.3. *If I is a central ideal of a ring R, then $C(R)I = (0)$.*

Proof. For any elements $t, r \in R$ and $i \in I$ we have

$$(rt)i = r(ti) = (ti)r = t(ir) = t(ri) = (tr)i,$$

and therefore

$$(rt - tr)i = 0.$$

Hence $C(R)I = (0)$. □

Lemma 3.4. *Let R be a non-simple ring with all non-zero inner derivations to be cofinite. If all ideals of R are central, then R is commutative or finite.*

Proof. *a*) If a ring R is not local, then $R = M_1 + M_2 \subseteq Z(R)$ for any two different maximal ideals M_1 and M_2 of R.

b) Suppose that R is a local ring and $J(R) \neq (0)$, where $J(R)$ is the Jacobson ideal of R. Then $J(R)C(R) = (0)$, $C(R) \neq R$ and, consequently,

$$C(R)^2 = (0).$$

If we assume that R is not commutative, then

$$(0) \neq C(R) < R,$$

and so there exists an element $x \in R \setminus Z(R)$ such that

$$\{0\} \neq \operatorname{Im} \partial_x \subseteq C(R).$$

Then $|R : C(R)| < \infty$. Since $C(R) \subseteq Z(R)$, we deduce that the index $|R : Z(R)|$ is finite. By Proposition 1 of [7], the commutator ideal $C(R)$ is finite and R is also finite. □

Lemma 3.5. *If $N(R) \subseteq Z(R)$, then every idempotent is central in a ring R.*

Proof. If $d \in \operatorname{Der} R$ and $e = e^2 \in R$, then we obtain $d(e) = d(e)e + ed(e)$, and this implies that

$$ed(e)e = 0 \text{ and } d(e)e, ed(e) \in N(R).$$

Then $ed(e) = e^2 d(e) = ed(e)e = 0$ and $d(e)e = 0$. As a consequence, $d(e) = 0$ and so $e \in Z(R)$. □

Lemma 3.6. *Let R be a ring with all non-zero inner derivations to be cofinite. Then one of the following conditions holds:*

(1) *R is a finite ring;*

(2) *R is a commutative ring;*

(3) *R contains a finite central ideal Z_0 such that R/Z_0 is an infinite residually finite ring (and, consequently, R/Z_0 is a prime ring with the ascending chain condition on ideals).*

Proof. Assume that R is an infinite ring which is not commutative and its every non-zero inner derivation is cofinite. Then $|R : C(R)| < \infty$ and every non-zero ideal of the quotient ring $B = R/Z_0$ has a finite index. If B is finite (or respectively $C(R) \subseteq Z_0$), then $|R : Z(R)| < \infty$ and, by Proposition 1 of [7], the commutator ideal $C(R)$ is finite. From this it follows that a ring R is finite, a contradiction. Hence B is an infinite ring and $C(R)$ is not contained in Z_0. Since $Z_0 C(R) = (0)$, we deduce that Z_0 is finite. By Corollary 2.2 and Theorem 2.3 from [2], B is a prime ring with the ascending chain condition on ideals. □

Let $D(R)$ be the subgroup of R^+ generated by all subgroups $d(R)$, where $d \in \operatorname{Der} R$.

Corollary 3.7. *Let R be an infinite ring that is not commutative and with all non-zero derivations (respectively inner derivations) to be cofinite. Then either R is a prime ring with the ascending chain condition on ideals or Z_0 is non-zero finite, $Z_0D(R) = (0)$, $D(R) \cap U(R) = \emptyset$ and $D(R)$ is a subgroup of finite index in R^+ (respectively $Z_0C(R) = (0)$, $C(R) \cap U(R) = \emptyset$ and $|R : C(R)| < \infty$).*

Proof. We have $Z_0 \neq R$, $Z_0C(R) = (0)$ and the quotient R/Z_0 is an infinite prime ring with the ascending chain condition on ideals by Corollary 2.2 and Theorem 2.3 from [2]. By Lemma 3.6, Z_0 is finite. Assume that $Z_0 \neq (0)$. If d is a non-zero derivation of R, then $Z_0d(R) \subseteq Z_0$ and so $Z_0d(R) = (0)$.

If we assume that $A = \operatorname{ann}_l d(R)$ is infinite, then A/Z_0 is an infinite left ideal of B with a non-zero annihilator, a contradiction with Lemma 2.1.1 from [6]. This gives that A is finite and, consequently, $A = Z_0$.

Finally, if $u \in D(R) \cap U(R)$, then $Z_0 = uZ_0 = (0)$, a contradiction. □

Corollary 3.8. *Let R be a ring that is not prime. If R contains an infinite subfield, then it has a non-zero derivation that is not cofinite.*

Proof of Proposition 1.1. ($\Leftarrow$) It is clear.

($\Rightarrow$) Assume that R is an infinite ring which is not commutative and its every non-zero inner derivation is cofinite. Then $Z_0 \neq R$ and R/Z_0 is an infinite prime ring by Lemma 3.6. Then $J(R) \subseteq Z_0$. Then

$$R/Z_0 = \sum_{i=1}^{m \ \oplus} M_{n_i}(D_i)$$

is a ring direct sum of finitely many full matrix rings $M_{n_i}(D_i)$ over skew fields D_i $(i = 1, \ldots, m)$ and so by applying Example 2.1 and Remark 3.2, we have that $R/Z_0 = F_1 \oplus D_1$ is a ring direct sum of a finite commutative ring F_1 and an infinite skew field D_1 that is not commutative. As a consequence of Proposition 1 from [8, §3.6] and Lemma 3.5,

$$R = F \oplus D$$

is a ring direct sum of a finite ring F and an infinite ring D. Then $F = Z_0$. □

4. Semiprime rings with cofinite inner derivations

Lemma 4.1. *Let R be a prime ring. If R contains a non-zero proper commutative ideal I, then R is commutative.*

Proof. Assume that $C(R) \neq (0)$. Then for any elements $u \in R$ and $a, b \in I$ we have

$$abu = a(bu) = (bu)a = b(ua) = uab$$

and so $ab \in Z(R)$. This gives that

$$I^2 \subseteq Z(R)$$

and therefore

$$I^2C(R) = (0).$$

Since $I^2 \neq (0)$, we obtain a contradiction with Lemma 2.1.1 of [6]. Hence R is commutative. □

Lemma 4.2. *Let R be a reduced ring (i.e. R has no non-zero nilpotent elements). If R contains a non-zero proper commutative ideal I such that the quotient ring R/I is commutative, then R is commutative.*

Proof. Obviously, $C(R) \leq I$ and $I^2 \neq (0)$. If $C(R) \neq (0)$, then, as in the proof of Lemma 4.1,

$$C(R)^3 \leq I^2C(R) = (0)$$

and thus $C(R) = (0)$. □

Lemma 4.3. *If a ring R contains an infinite commutative ideal I, then R is commutative or it has a non-zero derivation that is not cofinite.*

Proof. Suppose that R is not commutative. If all non-zero derivations are cofinite in R, then $B = R/Z_0$ is a prime ring by Lemma 3.6 and $C(B) \neq (0)$. Therefore $I^2C(R) \subseteq Z_0$ and, consequently, $I \subseteq Z_0$, a contradiction. □

Proof of Theorem 1.2. ($\Leftarrow$) It is obviously.

($\Rightarrow$) Suppose that R is an infinite ring which is not commutative and its every non-zero inner derivation is cofinite. Then $B = R/Z_0$ is a prime ring satisfying the ascending chain condition on ideals.

Assume that B is not a domain. By Proposition 2.2.14 of [11],

$$\operatorname{ann}_l b = \operatorname{ann}_r b = \operatorname{ann} b$$

is a two-sided ideal for any $b \in B$, and by Lemma 2.3.2 from [11], each maximal right annihilator in B has the form $\operatorname{ann}_r a$ for some $0 \neq a \in B$. Then $\operatorname{ann}_r a$ is a prime ideal. Since $|B : \operatorname{ann}_r a|$ is finite, left and right ideals Ba, aB are finite and this gives a contradiction. Hence B is a domain.

Now assume that $Z_0 \neq (0)$. In view of Corollary to Proposition 5 from [8, §3.5] we conclude that Z_0 is not nilpotent. As a consequence of Lemma 3 from [9] and Lemma 3.5,

$$R = Z_0 \oplus B_1$$

is a ring direct sum with a ring B_1 isomorphic to B. □

Remark 4.4. If R is a ring with all non-zero inner derivations to be cofinite and R/Z_0 is an infinite simple ring, then $R = Z_0 \oplus B$ is a ring direct sum of a finite central ideal Z_0 and a simple non-commutative ring B.

Problem 4.5. Characterize domains and, in particular, skew fields with all non-zero derivations (respectively inner derivations) to be cofinite.

Acknowledgements. The author is grateful to the referee whose remarks helped to improve the exposition of this paper.

References

[1] Artemovych, O. D., Differentially trivial and rigid rings of finite rank, *Periodica Math. Hungarica*, 36(1998) 1–16.

[2] Chew, K. L., Lawn, S., Residually finite rings, *Can. J. Math.*, 22(1970) 92–101.

[3] van den Essen, A., Wright, D., Zhao, W., Images of locally finite derivations of polynomial algebras in two variables, *J. Pure Appl. Algebra*, 215(2011) 2130–2134.

[4] Fucks, L., Infinite abelian groups, Vol. I. Pure and Applied Mathematics, Vol. 36. Academic Press, New York London, 1970.

[5] Fucks, L., Infinite abelian groups, Vol. II. Pure and Applied Mathematics, Vol. 36-II. Academic Press, New York London, 1973.

[6] Herstein, I. N., Noncommutative rings, The Carus Mathematical Monographs, No 15. Published by The Mathematical Association of America; distributed by J. Wiley & Sons, Inc., New York, 1968.

[7] Hirano, Y., On a problem of Szász, *Bull. Austral Math. Soc.*, 40(1989) 363–364.

[8] Lambek, J., Lectures notes on rings and modules, Blaisdell Publ. Co., Ginn and Co, Waltham, Mass. Toronto London, 1966.

[9] Lanski, C., Rings with few nilpotents, *Houston J. Math.*, 18(1992) 577–590.

[10] Levitz, K. B., Mott, J. L., Rings with finite norm property, *Can. J. Math.*, 24(1972) 557–562.

[11] McConnell, J. C., Robson, J. C., Noncommutative Noetherian rings, Pure and Applied Mathematics, J. Wiley & Sons, Ltd., Chichester, 1987.

Annales Mathematicae et Informaticae
***40** (2012) pp. 13–23*
http://ami.ektf.hu

A characterization of Symmetric group S_r, where r is prime number

**Alireza Khalili Asboei[a], Seyed Sadegh Salehi Amiri[a]
Ali Iranmanesh[b], Abolfazl Tehranian[a]**

[a]Department of Mathematics, Science and Research Branch
Islamic Azad University, Tehran, Iran
khaliliasbo@yahoo.com, salehisss@yahoo.com, tehranian1340@yahoo.com

[b]Department of Mathematics, Faculty of Mathematical Sciences
Tarbiat Modares University, Tehran, Iran
iranmanesh@modares.ac.ir

Submitted November 9, 2011 — Accepted April 11, 2012

Abstract

Let G be a finite group and $\pi_e(G)$ be the set of element orders of G. Let $k \in \pi_e(G)$ and m_k be the number of elements of order k in G. Set $\mathrm{nse}(G) := \{m_k \mid k \in \pi_e(G)\}$. In this paper, we prove the following results:

1. If G is a group such that $\mathrm{nse}(G) = \mathrm{nse}(S_r)$, where r is prime number and $|G| = |S_r|$, then $G \cong S_r$.
2. If G is a group such that $\mathrm{nse}(G) = \mathrm{nse}(S_r)$, where $r < 5 \times 10^8$ and $r-2$ are prime numbers and r is a prime divisor of $|G|$, then $G \cong S_r$.

Keywords: Element order, set of the numbers of elements of the same order, Symmetric group

MSC: 20D06, 20D20, 20D60

1. Introduction

If n is an integer, then we denote by $\pi(n)$ the set of all prime divisors of n. Let G be a finite group. Denote by $\pi(G)$ the set of primes p such that G contains an element of order p. Also the set of element orders of G is denoted by $\pi_e(G)$. A

finite group G is called a simple K_n-group, if G is a simple group with $|\pi(G)| = n$. Set $m_i = m_i(G) := |\{g \in G|$ the order of g is $i\}|$ and $\mathrm{nse}(G) := \{m_i | i \in \pi_e(G)\}$. In fact, m_i is the number of elements of order i in G and $\mathrm{nse}(G)$ is the set of sizes of elements with the same order in G. Throughout this paper, we denote by ϕ the Euler's totient function. If G is a finite group, then we denote by P_q a Sylow q-subgroup of G and by $n_q(G)$ the number of Sylow q-subgroup of G, that is, $n_q(G) = |Syl_q(G)|$. Also we say $p^k \parallel m$ if $p^k \mid m$ and $p^{k+1} \nmid m$. For a real number x, let $\varphi(x)$ denote the number of primes which are not greater than x, and $[x]$ the greatest integer not exceeding x. For positive integers n and k, let $t_n(k) = \prod_{i=1}^{k}(\prod_{n/(i+1)<p\leq n/i} p)^i$, where p is a prime. Denote by $\gcd(a, b)$ the greatest common divisor of positive integers a and b, and by $\exp_m(a)$ the exponent of a modulo m for the relatively prime integers a and m with $m > 1$. If m is a positive integer and p is a prime, let $|m|_p$ denote the p-part of m; in the other words, $|m|_p = p^k$ if $p^k \mid m$ but $p^{k+1} \nmid m$. For a finite group H, $|H|_p$ denotes the p-part of $|H|$. All further unexplained notations are standard and refer to [1], for example. In [2] and [3], it is proved that all simple K_4-groups and Mathieu groups can be uniquely determined by $\mathrm{nse}(G)$ and the order of G. In [4], it is proved that the groups A_4, A_5 and A_6 are uniquely determined only by $\mathrm{nse}(G)$. In [5], the authors show that the simple group $PSL(2, q)$ is characterizable by $\mathrm{nse}(G)$ for each prime power $4 \leq q \leq 13$. In this work it is proved that the Symmetric group S_r, where r is a prime number is characterizable by $\mathrm{nse}(G)$ and the order of G. In fact the main theorems of our paper are as follow:

Theorem 1. *Let G be a group such that nse(G)=nse(S_r), where r is a prime number and $|G| = |S_r|$. Then $G \cong S_r$.*

Theorem 2. *Let G be a group such that nse(G)=nse(S_r), where $r < 5 \times 10^8$ and $r - 2$ are prime numbers and $r \in \pi(G)$. Then $G \cong S_r$.*

In this paper, we use from [6] for proof some Lemmas, but since some part of the proof is different, we were forced to prove details get'em. We note that there are finite groups which are not characterizable by $\mathrm{nse}(G)$ and $|G|$. For example see the Remark in [2].

2. Preliminary Results

We first quote some lemmas that are used in deducing the main theorems of this paper.

Let $\alpha \in S_n$ be a permutation and let α have t_i cycles of length i, $i = 1, 2, \ldots, l$, in its cycle decomposition. The cycle structure of α is denote by $1^{t_1}2^{t_2}\ldots l^{t_l}$, where $1t_1 + 2t_2 \cdots + lt_l = n$. One can easily show that two permutations in S_n are conjugate if and only if they have the same cycle structure.

Lemma 2.1 ([6]).
(i) $\varphi(x) - \varphi(x/2) \geq 7$ for $x \geq 59$.

(ii) $\varphi(x) - \varphi(x/4) \geq 12$ *for* $x \geq 61$.
(iii) $\varphi(x) - \varphi(6x/7) \geq 1$ *for* $x \geq 37$.

Lemma 2.2 ([6]). *If* $n \geq 402$, *then* $(2/n)t_n(6) > e^{1.201n}$. *If* $n \geq 83$, *then* $(2/n)t_n(6) > e^{0.775n}$.

Lemma 2.3 ([6]). *Let* p *be a prime and* k *a positive integer.*
(i) If $|n!|_p = p^k$, *then* $(n-1)/(p-1) \geq k \geq n/(p-1) - 1 - [\log_p n]$.
(ii) If $|n!/m!|_p = p^k$ *and* $0 \leq m < n$, *then* $k \leq (n-m-1)/(p-1) + [\log_p n]$.

Lemma 2.4 ([7]). *Let* $\alpha \in S_n$ *and assume that the cycle decomposition of* α *contains* t_1 *cycles of length* 1, t_2 *cycles of length* $2, \ldots, t_l$ *cycles of length* l. *Then the order of conjugacy class of* α *in* S_n *is* $n!/1^{t_1}2^{t_2}\ldots l^{t_l}t_1!t_2!\ldots t_l!$.

Lemma 2.5 ([8]). *Let* G *be a finite group and* m *be a positive integer dividing* $|G|$. *If* $L_m(G) = \{g \in G | g^m = 1\}$, *then* $m \mid |L_m(G)|$.

Lemma 2.6 ([9]). *Let* G *be a finite group and* $p \in \pi(G)$ *be odd. Suppose that* P *is a Sylow* p*-subgroup of* G *and* $n = p^s m$, *where* $(p, m) = 1$. *If* P *is not cyclic and* $s > 1$, *then the number of elements of order* n *in* G *is always a multiple of* p^s.

Lemma 2.7 ([4]). *Let* G *be a group containing more than two elements. Let* $k \in \pi_e(G)$ *and* m_k *be the number of elements of order* k *in* G. *If* $s = \sup\{m_k | k \in \pi_e(G)\}$ *is finite, then* G *is finite and* $|G| \leq s(s^2 - 1)$.

Let m_n be the number of elements of order n. We note that $m_n = k\phi(n)$, where k is the number of cyclic subgroups of order n in G. Also we note that if $n > 2$, then $\phi(n)$ is even. If $n \in \pi_e(G)$, then by Lemma 2.2 and the above notation we have

$$\begin{cases} \phi(n) \mid m_n \\ n \mid \sum_{d|n} m_d \end{cases} \tag{2.1}$$

In the proof of the main theorem, we often apply (2.1) and the above comments.

3. Proof of the Main Theorem 1

We now prove the theorem 1 stated in the introduction. Let G be a group such that $\mathrm{nse}(G) = \mathrm{nse}(S_r)$, where r is a prime number and $|G| = |S_r|$. The following Lemmas reduce the problem to a study of groups with the same order with S_r.

Lemma 3.1. $m_r(G) = m_r(S_r) = (r-1)!$ *and if* $S \in Syl_r(G)$, $R \in Syl_r(S_r)$, *then* $|N_G(S)| = |N_{S_r}(R)|$.

Proof. Since $m_r(G) \in \mathrm{nse}(G)$ and $\mathrm{nse}(G) = \mathrm{nse}(S_r)$, then by (2.1) there exists $k \in \pi_e(S_r)$ such that $p \mid 1 + m_k(S_r)$. We know that $m_k(S_r) = \sum |cl_{S_r}(x_i)|$ such that $|x_i| = k$. Since $r \mid 1 + m_k(S_r)$, then $(r, m_k(S_r)) = 1$. If the cyclic structure of x_i for any i is $1^{t_1}2^{t_2}\ldots l^{t_l}$ such that $t_1, t_2, \ldots, t_l$ and $1, 2, \ldots, l$ are not equal

to r, then $r \mid r!/1^{t_1}2^{t_2}\cdots l^{t_l}t_1!t_2!\ldots t_l!$, that is $r \mid |cl_{S_r}(x_i)|$ for any i. Therefore $(r, m_k(S_r)) \neq 1$, which is a contradiction. Thus there exist $i \in \mathbb{N}$ such that $t_i = r$ or one of the numbers $1, 2, \ldots$ or l is equal to r. If there exist $i \in \mathbb{N}$ such that $t_i = r$, then the cyclic structure of x_i is 1^r. Hence $|x_i| = 1$, which is a contradiction. If one of the numbers $1, 2, \ldots$ or l is equal to r, then the cyclic structure of x_i is r^1. Hence $|x_i| = r$ and $k = r$. Therefore $m_r(G) = m_r(S_r)$, since $|G| = |S_r|$, then $n_r(G) = n_r(S_r) = m_r(G)/(r-1) = (r-2)!$. Hence if $S \in Syl_r(G)$, $R \in Syl_r(S_r)$, then $|N_G(S)| = |N_{S_r}(R)| = r(r-1)$. □

Lemma 3.2. *G has a normal series $1 \leq N < H \leq G$ such that $r \mid |H/N|$ and H/N is a minimal normal subgroup of G/N.*

Proof. Suppose $1 = N_0 < N_1 < \cdots < N_m = G$ is a chief series of G. Then there exists i such that $p \mid |N_i/N_{i-1}|$. Let $H = N_i$ and $N = N_{i-1}$. Then $1 \leq N < H \leq G$ is a normal series of G, H/N is a minimal normal subgroup of G/N, and $r \mid |H/N|$. Clearly, H/N is a simple group. □

Lemma 3.3. *Let $r \geq 5$ and let $1 \leq N < H \leq G$ be a normal series of G, where H/N is a simple group and $r \mid |H/N|$. Let $R \in Syl_r(G)$ and $Q \in Syl_r(G/N)$.*
(i) $|N_{G/N}(Q)| = |N_{H/N}(Q)||G/H|$ and $|N_N(R)||N_{G/N}(Q)| = |N_G(R)| = r(r-1)$.
(ii) If $P \in Syl_p(N)$ with $|P| = p^k$, where p is a prime and $k \geq 1$, then either $|H/N| \mid \prod_{i=0}^{k-1}(p^k - p^i)$ or $p^k|N_{G/N}(Q)| \mid r(r-1)$.

Proof. (i) By Frattini's argument, $G/N = N_{G/N}(Q)(H/N)$. Thus

$$G/H \cong N_{G/N}(Q)/N_{H/N}(Q).$$

So the first equality holds. Since we have

$$N_{G/N}(Q) \cong N_G(R)N/N \cong N_G(R)/N_N(R),$$

the second equality is also true.

(ii) By Frattini's argument again, $H = N_H(P)N$. Thus, we have $H/N \cong N_H(P)/N_N(P)$. Since H/N is a simple group, $C_H(P)N_N(P) = N_H(P)$ or $N_N(P)$. If $C_H(P)N_N(P) = N_H(P)$, then $r \mid |C_H(P)|$. Without loss of generality, we may assume $R \leq C_H(P)$. It means that $N_N(R) \geq P$. Then $p^k|N_{G/N}(Q)| \mid r(r-1)$ by (i). If $C_H(P)N_N(P) = N_N(P)$, then $C_H(P) \leq N_N(P)$. Thus $|N_H(P)/N_N(P)| \mid |N_H(P)/C_H(P)|$. Since $|H/N| = |N_H(P)/N_N(P)|$ and $N_H(P)/C_H(P)$ is isomorphic to a subgroup of $Aut(P)$, $|H/N| \mid |Aut(P)|$. Since $|Aut(P)| \mid \prod_{i=0}^{k-1}(p^k - p^i)$, $|H/N| \mid \prod_{i=0}^{k-1}(p^k - p^i)$. □

Lemma 3.4. *Let $r \geq 5$ and let $1 \leq N < H \leq G$ be a normal series of G with H/N simple and $r \mid |H/N|$. If $|N|_p|G/H|_p = p^k$ with $k \geq 1$ and $|H/N|$ not dividing $\Pi_{i=0}^{k-1}(p^k - p^i)$, then $p^k \mid (r-1)$.*

Proof. Assume $|N|_p = p^k$. If $t = 0$, then $p^k \mid |G/H|$. By Lemma 3.3 (i), $p^k \mid r(r-1)$. If $t \geq 1$, since $|H/N|$ does not divide $\Pi_{i=0}^{k-1}(p^k - p^i)$ and $\prod_{i=k-t}^{k-1}(p^k - p^i) = p^{t(k-t)}\prod_{j=0}^{t-1}(p^t - p^j)$, we have that $|H/N|$ does not divide $\prod_{j=0}^{t-1}(p^t - p^j)$. By Lemma 3.3 (ii), $p^t|N_{G/N}(Q)| \mid r(r-1)$, where $Q \in Syl_r(G/N)$. By Lemma 3.3 (i), $|N_{G/N}(Q)| = |N_{H/N}(Q)||G/H|$, so we have $p^t|G/H| \mid r(r-1)$. Since $|N|_p|G/H|_p = p^k$ and $|N|_p = p^k$, we obtain $|G/H|_p = p^{k-t}$. Thus $p^k \mid r(r-1)$. Since $r \mid |H/N|$, it is easy to know $p \neq r$. Therefore, $p^k \mid (r-1)$. □

Lemma 3.5. *Let $r \geq 5$ and let $1 \leq N < H \leq G$ be a normal series of G with H/N simple. If $r \mid |H/N|$, then $t_r(1) \mid |H/N|$ and H/N is a non-ablian simple group and G is not solvable group.*

Proof. We first prove that $t_r(1) \mid |H/N|$. If $t_r(1) \nmid |H/N|$, then there exists a prime p satisfying $r/2 < p < r$ such that $p \mid |N||G/H|$. Since $r \mid |H/N|$, $|H/N| \nmid (p-1)$. Hence $p \mid (r-1)$ by Lemma 3.4. But $(r-1)/2 < r/2$, contrary to $r/2 < p$. Since the number of prime factors of $t_r(1)$ is greater that 1, then H/N is a non-ablian simple group. Clearly G is not solvable group. □

Lemma 3.6. *If $r \geq 59$ and let $1 \leq N < H \leq G$ be a normal series of G with H/N simple and $r \mid |H/N|$,*
(i) If $\gcd(t_r(6), r-1) = 1$, then $t_r(6) \mid |H/N|$.
(ii) If $\gcd(t_r(6), r-1)$ is a prime p, then $(t_r(6)/p) \mid |H/N|$.

Proof. By Lemma 3.5, $t_r(1) \mid |H/N|$. Suppose $t_r(6) \nmid |H/N|$. There exists a prime q with $r/7 < q \leq r/2$ such that $q \mid |N||G/H|$. Let $|N|_q|G/H|_q = q^k$. If $|H/N| \mid \prod_{i=0}^{k-1}(q^k - q^i)$ with $1 \leq k \leq 6$, then $t_r(1) \mid \prod_{i=1}^{k}(q^i - 1)$. By Lemma 2.1, the number of prime factors of $t_r(1)$ is greater than 6. But the number of primes p with $p \mid \prod_{i=1}^{6}(q^i - 1)$ and $r/2 < p$ is less than or equal to 6, a contradiction. By Lemma 3.4, $q^k \mid (r-1)$. If $\gcd(t_r(6), r-1) = 1$, then $k = 0$, contrary to $q \mid |N||G/H|$. Hence, (i) is true. If $\gcd(t_r(6), r-1) = p$, then $k = 1$ and $q = p$. It follows that $(t_r(6)/p) \mid |H/N|$. This proves (ii). □

Lemma 3.7. *Let $r \geq 5$. If $1 \leq N < H \leq G$ is a normal series of G, $t_r(1) \mid |H/N|$, and H/N is a non-abelian simple group, then $H/N \cong A_r$.*

Proof. We consider the following cases:

Case 1. $r = 5$. In this case, we have $|H/N| = 2^a 3 \cdot 5$ with $a \leq 3$. It is clear that $H/N \cong A_5$.

Case 2. $r = 7$. In this case, we have $|H/N| = 2^a 3^b 5 \cdot 7$ with $a \leq 4$ and $b \leq 2$. It is clear that $H/N \cong A_7$.

Case 3. $11 \leq r \leq 19$. Note that $|G| < 10^{25}$ for $11 \leq r \leq 19$. If H/N is not isomorphic to any alternating group, since $t_r(1) \mid |H/N|$, by [1, pp. 239–241], H/N is isomorphic to one of the following groups:

M_{22} (for $r = 11$),	$L_2(q)$ of order $\geq 10^6$,	$G_2(q)$ of order $\geq 10^{20}$,
Suz (for $r = 13$),	$L_3(q)$ of order $\geq 10^{12}$,	
HS (for $r = 11$),	$U_3(q)$ of order $\geq 10^{12}$,	
McL (for $r = 11$),	$L_4(q)$ of order $\geq 10^{16}$,	
Fi_{22} (for $r = 13$),	$U_4(q)$ of order $\geq 10^{16}$,	
$U_6(2)$ (for $r = 11$),	$S_4(q)$ of order $\geq 10^{16}$,	

If H/N is isomorphic to one of the six groups on the left side, by $|H/N| \mid |G|$, we have $H/N \cong M_{22}$ and $r = 11$. So $|N|_3|G/H|_3 = 3$ by $|S_r|_3/|M_{22}|_3 = 3$. Since $|M_{22}| \nmid (3^2-3)(3^2-1)$, we have $3 \mid 10$ by Lemma 3.4, a contradiction. Suppose H/N is isomorphic to a simple group of Lie type in characteristic p. Let $|H/N|_p = p^t$. If H/N is isomorphic to $L_4(q)$, $U_4(q)$, $S_4(q)$, or $G_2(q)$ of order $\geq 10^{16}$, then $p^t \geq 10^6$ by Lemma 4 in [10]. When $p \geq 3$, by Lemma 2.3, $10^6 \leq p^t \leq p^{(r-1)/(p-1)} \leq 3^{(r-1)/2} < 3^{11}$, a contradiction. When $p = 2$, since $2^{19} \nmid |G|$, we have $10^6 \leq p^t \leq 2^{18}$, a contradiction. If $H/N \cong U_3(q)$ ($q = p^k$), then $p \neq 11$ by $p^{3k} \mid |U_3(q)|$ and $11^3 \nmid |G|$. Thus $11 \mid p^{2k} - 1$ or $11 \mid p^{3k} + 1$. Since $11 \nmid p^2 - 1$, we have $\exp_{11}(p) = 5$ or 10. Therefore, $5 \mid k$. Thus $p^{3k} + 1$ has a prime factor ≥ 31 (see Lemma 2 in [11]), contrary to $r \leq 19$. Similarly, we derive a contradiction if $H/N \cong L_2(q)$ or $L_3(q)$.

Case 4. $23 \leq r \leq 43$. Since $t_r(1) \mid |H/N|$, it is easy to prove that H/N is not isomorphic to any sporadic simple group. If H/N is isomorphic to a simple group of Lie type in characteristic 23, we have $H/N \cong L_2(23)$ or $L_2(23^2)$. If $H/N \cong L_2(23)$, we have $r = 23$, since $29 \nmid |L_2(23)|$. But $19 \nmid |L_2(23)|$, contrary to $t_r(1) \mid |H/N|$. If $H/N \cong L_2(23^2)$, then $r = 43$. But $43 \nmid |L_2(23^2)|$, again contrary to $r \mid |H/N|$. If $H/N \cong^3 D_4(p^k)$ with $p \neq 23$, then $23 \mid p^{8k} + p^{4k} + 1$ or $23 \mid p^{6k} - 1$. Moreover, $23 \mid p^{12k} - 1$. We have $\exp_{23}(p) = 11$ or 22 since $23 \nmid p^2 - 1$. Thus, $11 \mid k$. Then $p^{132} \mid |{}^3D_4(p^k)|$, contrary to $p^{132} \nmid |G|$. If H/N is isomorphic to a simple group of Lie type in characteristic p except ${}^3D_4(p^k)$ with $p \neq 23$, let $|H/N|_p = p^s$. By examining the orders of simple groups of Lie type, we know that there exists a positive integer $t \leq s$ such that $23 \mid p^t + 1$ and $(p^t + 1) \mid |H/N|$, or $23 \mid p^t - 1$ and $(p^t - 1) \mid |H/N|$. As above, we can prove $11 \mid t$. Thus, $s \geq t \geq 11$. Since $p^{11} \nmid |G|$ for $r \leq 43$ and $p \geq 5$, we have $p = 2$ or 3. Since 2 and 3 are not primitive roots, we have $(2^{11} - 1) \mid |H/N|$ or $(3^{11} - 1) \mid |H/N|$. But $2^{11} - 1$ and $3^{11} - 1$ have a prime factor > 43, contrary to $r \leq 43$.

Case 5. $47 \leq r \leq 79$. In this case, $47 \mid |H/N|$. It can be proved that H/N is isomorphic to an alternating group as above.

Case 6. $r \geq 83$. Clearly, H/N is not isomorphic to any sporadic simple group for $r \geq 83$. If H/N is isomorphic to a simple group of Lie type in characteristic p and $|H/N|_p = p^t$, then $|H/N| < p^{3t}$ by Lemma 4 in [10]. In particular, if H/N is not isomorphic to $L_2(p^t)$, then $|H/N| < p^{8t/3}$. We first prove $p \leq r/7$. If $r/2 < p \leq r$, then we have $H/N \cong L_2(p)$. Since $|L_2(p)| = p(p^2-1)/2$, the number of prime factors of $t_r(1)$ is not greater than 2, contrary to Lemma 2.1. If $r/(s+1) < p \leq r/s$ with $s = 2$ or 3, then $t_r(1) < |H/N|/p^t < p^{2t} \leq p^{2s} \leq p^6 \leq (r/2)^6$. But $t_r(1) > (r/2)^7$ by Lemma 2.1, a contradiction. If $r/(s+1) < p \leq r/s$ with $4 \leq s \leq 6$, by Lemma 3.6, we have $(2/r)t_r(3) < |H/N|/p^t < p^{2t} \leq p^{2s} \leq p^{12} \leq (r/4)^{12}$. By

Lemma 2.1, we have $(2/r)t_r(3) > (2/r)(r/4)^{12}t_{[r/2]}(1) > (r/4)^{12}$, a contradiction. Now we prove that $p \leq r/7$ is impossible.

(i) If $r \geq 409$ and $p \geq 3$, by Lemmas 2.2, 2.3, and 3.6, we have $e^{1.201r} < (2/r)t_r(6) < |H/N|/p^t < p^{2t} \leq p^{2(r-1)/(p-1)} < (p^{2/(p-1)})^r \leq 3^r$. But $e^{1.201} > 3$, a contradiction.

(ii) For the case where $r \geq 409$ and $p = 2$, if H/N is not isomorphic to $L_2(2^t)$, we have $e^{1.201r} < (2/r)t_r(6) < |H/N|/2^t < 2^{5t/3} < 2^{5r/3}$. But $e^{1.201} > 2^{5/3}$, a contradiction.

Suppose $H/N \cong L_2(2^t)$. Since $(2^{2t}-1) \mid |L_2(2^t)|$ and $2^{2t}-1$ has a prime factor q satisfying $\exp_q(2) = 2t$ (see Lemma 2 in [11]), we have $2t+1 \leq q \leq r$. Hence, $e^{1.201r} < (r/2)t_r(6) \leq 2^{2t} - 1 < 2^r$, a contradiction.

(iii) If $83 \leq r \leq 401$ and $p \geq 7$, we can deduce $e^{0.775r} < 7^{r/3}$ as above, a contradiction.

(iv) If $83 \leq r \leq 401$ and $p \leq 5$, we have $83 \mid |H/N|$ by Lemma 3.6. Similar to the argument used in the case where $23 \leq r \leq 43$, we can deduce $p^{41} - 1 \mid |H/N|$ or $p^{41} + 1 \mid |H/N|$. But $p^{41} - 1$ and $p^{41} + 1$ have a prime factor > 401 for $p \leq 5$, contrary to $r \leq 401$.

We have proved that $H/N \cong A_r$. Now set $\overline{H} := H/N \cong A_r$ and $\overline{G} := G/N$. On the other hand, we have:

$$A_r \cong \overline{H} \cong \overline{H}C_{\overline{G}}(\overline{H})/C_{\overline{G}}(\overline{H}) \leq \overline{G}/C_{\overline{G}}(\overline{H}) = N_{\overline{G}}(\overline{H})/C_{\overline{G}}(\overline{H}) \leq Aut(\overline{H}).$$

Let $K = \{x \in G \mid xN \in C_{\overline{G}}(\overline{H})\}$, then $G/K \cong \overline{G}/C_{\overline{G}}(\overline{H})$. Hence $A_r \leq G/K \leq Aut(A_r)$, and hence $G/K \cong A_r$ or $G/K \cong S_r$. If $G/K \cong A_r$, then $|K| = 2$. We have $N \leq K$, and N is a maximal solvable normal subgroup of G, then $N = K$. Hence $H/N \cong A_r = G/N$, then $|N| = 2$. So G has a normal subgroup N of order 2, generated by a central involution z. Therefore G has an element of order $2r$. Now we prove that G does not any element of order $2r$, a contradiction. At first we show that $r \parallel m_2(S_r) = m_2(G)$. We have $m_2(S_r) = \sum |cl_{S_r}(x_k)|$ such that $|x_k| = 2$. Since $2 \neq 1, r$, the cyclic structure of x_k for any k is $1^{t_1}2^{t_2}\dots l^{t_l}$, where $t_1, t_2, \dots, t_l, 1, 2, \dots, l$ are not equal to r. On the other hand, we have $|cl_{S_r}(x_k)| = r!/1^{t_1}2^{t_2}\dots l^{t_l}t_1!t_2!\dots t_l!$. Hence $m_2(S_r) = r!h$, where h is a real number. Since $m_2(S_r) \lneqq r!$, then $0 < h < 1$. Therefore $r \parallel m_2(S_r)$. We know that if P and Q are Sylow r-subgroups of G, then they are conjugate, which implies that $C_G(P)$ and $C_G(Q)$ are conjugate. Since $2r \in \pi_e(G)$, we have $m_{2r}(G) = \phi(2r)n_r(G)k = (r-1)!k$, where k is the number of cyclic subgroups of order 2 in $C_G(P_r)$. Hence $m_r(G) \mid m_{2r}(G)$. On the other hand, $2r \mid (1 + m_2(G) + m_r(G) + m_{2r}(G))$, by (2.1). Since $r \mid (1+m_r(G))$ and $r \mid m_2(G)$, then $r \mid m_{2r}(G)$. Therefore by $(r-1)! \mid m_{2r}(G)$ and $r \mid m_{2r}(G)$, we can conclude that $r! \mid m_{2r}(G)$, a contradiction. Hence G/K is not isomorphic to A_r, and hence $G/K \cong S_r$, then $|K| = 1$ and $G \cong S_r$. Thus the proof is completed. □

Corollary 3.8. *Let G be a finite group. If $|G| = |S_r|$, where r is a prime number and $|N_G(R)| = |N_{S_r}(S)|$, where $R \in Syl_r(G)$ and $S \in Syl_r(S_r)$, then $G \cong S_r$.*

Proof. It follows at once from Theorem 1. □

Corollary 3.9. *Let G be a finite group. If $|N_G(P_1)| = |N_{S_r}(P_2)|$ for every prime p, where $P_1 \in Syl_p(G)$, $P_2 \in Syl_p(S_r)$ and r is a prime number, then $G \cong S_r$.*

Proof. Since $|N_G(P_1)| = |N_{S_r}(P_2)|$ for every prime p, where $P_1 \in Syl_p(G)$, $P_2 \in Syl_p(S_r)$, we have $|P_1| = |P_2|$. Thus, $|G|_p = |S_r|_p$ for every prime p. Hence, $|G| = |S_r|$. It follows that $G \cong S_r$. □

4. Proof of the Main Theorem 2

We now prove the theorem 2 stated in the Introduction. Let G be a group such that $\mathrm{nse}(G) = \mathrm{nse}(S_r)$, where $r < 5 \times 10^8$ and $r - 2$ are prime numbers and $r \in \pi(G)$. By Lemma 2.7, we can assume that G is finite. The following lemmas reduce the problem to a study of groups with the same order with S_r.

Lemma 4.1. *If $i \in \pi_e(S_r)$, $i \neq 1$ and $i \neq r$, then $r \parallel m_i(S_r)$.*

Proof. We have $m_i(S_r) = \sum |cl_{S_r}(x_k)|$ such that $|x_k| = i$. Since $i \neq 1, r$, the cyclic structure of x_k for any k is $1^{t_1}2^{t_2}\dots l^{t_l}$, where $t_1, t_2, \dots, t_l, 1, 2, \dots, l$ are not equal to r. On the other hand, we have $|cl_{S_r}(x_k)| = r!/1^{t_1}2^{t_2}\dots l^{t_l}t_1!t_2!\dots t_l!$. Hence $m_i(S_r) = r!h$, where h is a real number. Since $m_i(S_r) \nleq r!$, then $0 < h < 1$. Therefore $r \parallel m_i(S_r)$. □

Lemma 4.2. $|P_r| = r$.

Proof. At first we prove that if $r = 5$, then $|P_5| = 5$. We know that, $\mathrm{nse}(G) = \mathrm{nse}(S_5) = \{1, 20, 24, 25, 30\}$. We show that $\pi(G) \subseteq \{2, 3, 5\}$. Since $25 \in \mathrm{nse}(G)$, it follows from (2.1) that $2 \in \pi(G)$ and $m_2 = 25$. Let $2 \neq p \in \pi(G)$. By (2.1), we have $p \in \{3, 5, 31\}$. If $p = 31$, then by (2.1), $m_{31} = 30$. On the other hand, if $62 \in \pi_e(G)$, then by (2.1), we conclude that $m_{62} = 30$ and $62|86$, which is a contradiction. Therefore $62 \notin \pi_e(G)$. So P_{31} acts fixed point freely on the set of elements of order 2, and $|P_{31}| \mid m_2$, which is a contradiction. Thus $\pi(G) \subseteq \{2, 3, 5\}$. It is easy to show that, $m_5 = 24$, by (2.1). Also if $3 \in \pi_e(G)$, then $m_3 = 20$. By (2.1), we conclude that G does not contain any element of order 15, 20 and 25. Also, we get $m_4 = 30$ and $m_8 = 24$ and G does not contain any element of order 16. Since $2, 5 \in \pi(G)$, hence we have $\pi(G) = \{2, 5\}$ or $\{2, 3, 5\}$. Suppose that $\pi(G) = \{2, 5\}$. Then $\pi_e(G) \subseteq \{1, 2, 4, 5, 8, 10\}$. Therefore $|G| = 100+20k_1+24k_2+30k_3 = 2^m \times 5^n$, where $0 \leq k_1 + k_2 + k_3 \leq 1$. Hence $5 \mid k_2$, which implies that $k_2 = 0$, and so $50 + 10k_1 + 15k_3 = 2^{m-1} \times 5^n$. Hence $2 \mid k_3$, which implies that $k_3 = 0$. It is easy to check that the only solution of the equation is $(k_1, k_2, k_3, m, n) = (0, 0, 0, 2, 2)$. Thus $|G| = 2^2 \times 5^2$. It is clear that $\pi_e(G) = \{1, 2, 4, 5, 10\}$, hence $\exp(P_2) = 4$, and P_2 is cyclic. Therefore $n_2 = m_4/\phi(4) = 30/2 = 15$, since every Sylow 2-subgroup has one element of order 2, then $m_2 \leq 15$, which is a contradiction. Hence $\pi(G) = \{2, 3, 5\}$. Since G has no element of order 15, the group P_5 acts fixed point freely on the set of elements of order 3. Therefore $|P_5|$ is a divisor of $m_3 = 20$, which implies that $|P_5| = 5$. Now suppose that $r \neq 5$, by Lemma 4.1, we have $r^2 \nmid m_i(G)$, for any $i \in \pi_e(G)$. On the other hand, if $r^3 \in \pi_e(G)$, then by (2.1)

we have $\phi(r^3) \mid m_{r^3}(G)$. Thus $r^2 \mid m_{r^3}(G)$, which is a contradiction. Therefore $r^3 \notin \pi_e(G)$. Hence $\exp(P_r) = r$ or $\exp(P_r) = r^2$. We claim that $\exp(P_r) = r$. Suppose that $\exp(P_r) = r^2$. Hence there exists an element of order r^2 in G such that $\phi(r^2) \mid m_{r^2}(G)$. Thus $r(r-1) \mid m_{r^2}(G)$. And so $m_{r^2}(G) = r(r-1)t$, where $r \nmid t$. If $|P_r| = r^2$, then P_r will be a cyclic group and we have $n_r(G) = m_{r^2}(G)/\phi(r^2) = r(r-1)t/r(r-1) = t$. Since $m_r(G) = (r-1)!$, then $(r-1)! = (r-1)n_r(G) = (r-1)t$. Therefore $t = (r-2)!$ and $m_{r^2}(G) = r(r-1)(r-2)! = r!$, which is a contradiction. If $|P_r| = r^s$, where $s \geq 3$, then by Lemma 2.6, we have $m_{r^2}(G) = r^2 l$ for some natural number l, which is a contradiction by Lemma 4.1. Thus $\exp(P_r) = r$. By Lemma 2.5, $|P_r| \mid (1 + m_r(G)) = 1 + (r-1)!$. By [12], $|P_r| = r$. □

Lemma 4.3. *$\pi(G) = \pi(S_r)$.*

Proof. By Lemma 4.2, we have $|P_r| = r$. Hence $(r-2)! = m_r(G)/\phi(r) = n_r(G) \mid |G|$. Thus $\pi((r-2)!) \subseteq \pi(G)$. Now we show that $\pi(S_r) = \pi(G)$. Let p be a prime number such that $p > r$. Suppose that $pr \in \pi_e(G)$. We have $m_{pr}(G) = \phi(pr)n_r(G)k$, where k is the number of cyclic subgroups of order p in $C_G(P_r)$. Hence $(p-1)(r-1)! \mid m_{pr}$. On the other hand, since p is prime and $p > r$, then $p - 1 > r$. Thus $(p-1)(r-1)! > r!$, then $m_{pr} > r!$, which is a contradiction. Thus $pr \notin \pi_e(G)$. Then P_p acts fixed point freely on the set of elements of order r, and so $|P_p| \mid (r-1)!$, which is a contradiction. Therefore $p \notin \pi(G)$. By the assumption $r \in \pi(G)$, hence $\pi(G) = \pi(S_r)$. □

Lemma 4.4. *G has not any element of order $2r$.*

Proof. Suppose that G has an element of order $2r$. We have

$$m_{2r}(G) = \phi(2r)n_r(G)k = (r-1)!k,$$

where k is the number of cyclic subgroups of order 2 in $C_G(P_r)$. Hence $m_r(G) \mid m_{2r}(G)$. On the other hand, $2r \mid (1 + m_2(G) + m_r(G) + m_{2r}(G))$, by (2.1). Since $r \mid (1 + m_r(G))$ and $r \mid m_2(G)$ by Lemma 4.1, $r \mid m_{2r}(G)$. Therefore by $(r-1)! \mid m_{2r}(G)$ and $r \mid m_{2r}(G)$, we can conclude that $r! \mid m_{2r}(G)$, a contradiction. □

Lemma 4.5. *G has not any element of order $3r, 5r, 7r, \ldots, pr$, where p is the prime number such that $p < r$.*

Proof. The proof of this lemma is completely similar to Lemma 4.4. □

Lemma 4.6. *If $p = r - 2$, then $|P_p| = p$ and $n_p(G) = r!/2p(p-1)$.*

Proof. Since $pr \notin \pi_e(G)$, then the group P_p acts fixed point freely on the set of elements of order r, and so $|P_p| \mid m_r(G) = (r-1)!$. Thus $|P_p| = p$. Since Sylow p-subgroups are cyclic, then $n_p(G) = m_p(G)/\phi(p) = r!/2p(p-1)$. □

Lemma 4.7. *$|G| = |S_r|$.*

Proof. We can suppose that $|S_r| = 2^{k_2}3^{k_3}5^{k_5}\cdots l^{k_l}pr$, where $k_2, k_3, k_5, \ldots, k_l$ are non-negative integers. By Lemma 4.4, the group P_2 acts fixed point freely on the set of elements of order r, and so $|P_2| \mid m_r(G) = (r-1)!$. Thus $|P_2| \mid 2^{k_2}$. Similarly by Lemma 4.5, we have $|P_3| \mid 3^{k_3}, \ldots, |P_l| \mid l^{k_l}$. Therefore $|G| \mid |S_r|$. On the other hand, we know that $(r-2)! = m_r(G)/\phi(r) = n_r(G)$ and $n_r(G) \mid |G|$ and $n_p(G) = r!/2p(p-1) \mid |G|$, then the least common multiple of $(r-2)!$ and $r!/2p(p-1)$ divide the order of G. Therefore $r!/2 \mid |G|$ and so $|G| = |A_r|$ or $|G| = |S_r|$. If $|G| = |A_r|$, by $m_r(S_r) = m_r(A_r) = (r-1)!$, then $|N_G(R)| = |N_{A_r}(S)|$, where $R \in Syl_r(G)$ and $S \in Syl_r(A_r)$, similarly to main Theorem 1, $G \cong A_r$. But we can prove that $\mathrm{nse}(G) \neq \mathrm{nse}(A_r)$. Suppose that $\mathrm{nse}(G) = \mathrm{nse}(A_r)$, since $\mathrm{nse}(G) = \mathrm{nse}(S_r)$, then $m_2(S_r) = m_2(A_r)$. On the other hand $m_2(S_r) = \sum |cl_{S_r}(x_i)|$ such that $|x_i| = 2$, since cyclic structure $1^{r-2}2$ no exists in A_r, then it is clear that $m_2(S_r) > m_2(A_r)$, a contradiction. Hence $|G| = |S_r|$. □

Now by the main Theorem 1, $G \cong S_r$, and the proof is completed.

Acknowledgment. The authors would like to thank from the referees for the valuable comments.

References

[1] Conway, J. H., Curtis, R. T., Norton, S. P., et al., Atlas of Finite Groups. *Clarendon, Oxford*, 1985.

[2] Shao, C. G., Shi, W., Jiang, Q. H., Characterization of simple K_4-groups. *Front Math, China.* **3**(2008), 355–370.

[3] Shao, C. G., Jiang, Q. H., A new characterization of Mathieu groups. *Archivum Math, (Brno) Tomus.* **46**(2010), 13–23.

[4] Shen, R., Shao, C. G., Jiang, Q. H., Shi, W., Mazuro, V., A New Characterization of A_5. *Monatsh Math.* **160** (2010), 337–341.

[5] Khatami, M., Khosravi, B., Akhlaghi, Z., A new characterization for some linear groups. *Monatsh Math.* **163**(2009), 39–50.

[6] Bi, J., Characteristic of Alternating Groups by Orders of Normalizers of Sylow Subgroups. *Algebra Colloq.* **8**(2001), 249–256.

[7] Zassenhaus, H., The theory of groups. 2nd ed, *Chelsea Publishing Company New York*, 1958.

[8] Frobenius, G., Verallgemeinerung des sylowschen satze. *Berliner sitz.* (1895), 981–993.

[9] Miller, G., Addition to a theorem due to Frobenius. *Bull. Am. Math. Soc.* **11**(1904), 6–7.

[10] Bi, J., A characterization of the symmetric groups. *Acta Math. Sinica.* **33**(1990), 70–77. (in Chinease)

[11] Bi, J., A characterization of $L_n(q)$ by the normalizers' orders of their Sylow subgroups. *Acta Math. Sinica* (New Ser). **11**(1995), 300–306.

[12] Crandal, R., Dilcher, K., Pomerance, C., A search for Wieferich and Wilson primes. *Matematics of Computation.* **66**(1997), 433–449.

Annales Mathematicae et Informaticae
***40** (2012) pp. 25–35*
http://ami.ektf.hu

On k-periodic binary recurrences

Nurettin Irmak[a], László Szalay[b]

[a]Department of Mathematics, University of Niğde
nirmak@nigde.edu.tr

[b]Institute of Mathematics, University of West Hungary
laszalay@emk.nyme.hu

Submitted November 2, 2012 — Accepted November 28, 2012

Abstract

We apply a new approach, namely the fundamental theorem of homogeneous linear recursive sequences, to k-periodic binary recurrences which allows us to determine Binet's formula of the sequence if k is given. The method is illustrated in the cases $k = 2$ and $k = 3$ for arbitrary parameters. Thus we generalize and complete the results of Edson-Yayenie, and Yayenie linked to $k = 2$ hence they gave restrictions either on the coefficients or on the initial values. At the end of the paper we solve completely the constant sequence problem of 2-periodic sequences posed by Yayenie.

Keywords: linear recurrences, k-periodic binary recurrences

MSC: 11B39, 11D61

1. Introduction

Let a, b, c, d, and q_0, q_1 denote arbitrary complex numbers, and consider the following construction of the sequence (q_n). For $n \geq 2$, the terms q_n are defined by

$$q_n = \begin{cases} aq_{n-1} + bq_{n-2}, & \text{if } n \text{ is even;} \\ cq_{n-1} + dq_{n-2}, & \text{if } n \text{ is odd.} \end{cases} \tag{1.1}$$

The sequence (q_n) is called 2-periodic binary recurrence, and it was described first by Edson and Yayenie [2]. The authors discussed the specific case $q_0 = 0$, $q_1 = 1$ and $b = d = 1$, gave the generating function and Binet-type formula of

(q_n), further they proved several identities among the terms of (q_n). In the same paper the sequence (q_n) was investigated for arbitrary initial values q_0 and q_1, but $b = d = 1$ were still assumed.

Later Yayenie [6] took one more step by determining the Binet's formula for (q_n), where b and d were arbitrary numbers, but the initial values were fixed as $q_0 = 0$ and $q_1 = 1$.

The main tool in the papers [2, 6] is to work with the generating function. In this paper we suggest a new approach, namely to apply the fundamental theorem of homogeneous liner recurrences (see Theorem 1.1). This powerful method allows us to give the Binet's formula of (q_n) for any b and d and for arbitrary initial values. Moreover, we can also handle the case when the zeros of the quadratic polynomial

$$p_2(x) = x^2 - (ac + b + d)x + bd$$

coincide. Note, that $p_2(x)$ plays an important role in the aforesaid papers, but the sequence (q_n) has not been discussed yet when $p_2(x)$ has a zero with multiplicity 2. We will see that the application of the fundamental theorem of linear recurrences is very effective and it can even be used at k-periodic sequences generally. At the end of the paper we solve an open problem concerning constant subsequences (see 2.2.2 in [6]).

The k-periodic second order linear recurrence

$$q_n = \begin{cases} a_0 q_{n-1} + b_0 q_{n-2}, & \text{if } n \equiv 0 \,(\mathrm{mod}\, k); \\ a_1 q_{n-1} + b_1 q_{n-2}, & \text{if } n \equiv 1 \,(\mathrm{mod}\, k); \\ \vdots & \vdots \\ a_{k-1} q_{n-1} + b_{k-1} q_{n-2}, & \text{if } n \equiv k-1 \,(\mathrm{mod}\, k). \end{cases} \tag{1.2}$$

was introduced by Cooper in [1], where mainly the combinatorial interpretation of the coefficients A_k and B_k appearing in the recurrence relation $q_n = A_k q_{n-k} + B_k q_{n-2k}$ was discussed. Note that Lemma 4 of the work of Shallit [4] also describes an approach to compute the coefficients for q_n. Edson, Lewis and Yayenie [3] also studied the k-periodic extension, again with $q_0 = 0$, $q_1 = 1$ and with the restrictions $b_0 = b_1 = \cdots = b_{k-1} = 1$.

At the end of the first section we recall the fundamental theorem of linear recurrences. A homogeneous linear recurrence $(G_n)_{n=0}^{\infty}$ of order k $(k \geq 1, k \in \mathbb{N})$ is defined by the recursion

$$G_n = A_1 G_{n-1} + A_2 G_{n-2} + \cdots + A_k G_{n-k} \quad (n \geq k), \tag{1.3}$$

where the initial values $G_0, \ldots, G_{k-1}$ and the coefficients $A_1, \ldots, A_k$ are complex numbers, $A_k \neq 0$ and $|G_0| + \cdots + |G_{k-1}| > 0$. The characteristic polynomial of the sequence (G_n) is the polynomial

$$g(x) = x^k - A_1 x^{k-1} - \cdots - A_k.$$

Denote by $\alpha_1, \ldots, \alpha_t$ the distinct zeros of the characteristic polynomial $g(x)$, which can there be written in the form

$$g(x) = (x - \alpha_1)^{e_1} \cdots (x - \alpha_t)^{e_t}. \tag{1.4}$$

The following result (see e.g. [5]) plays a basic role in the theory of recurrence sequences, and here in our approach.

Theorem 1.1. *Let (G_n) be a sequence satisfying the relation (1.3) with $A_k \neq 0$, and $g(x)$ its characteristic polynomial with distinct roots $\alpha_1, \ldots, \alpha_t$. Let $K = \mathbb{Q}(\alpha_1, \ldots, \alpha_t, A_1, \ldots, A_k, G_0, \ldots, G_{k-1})$ denote the extension of the field of rational numbers and let $g(x)$ be given in the form (1.4). Then there exist uniquely determined polynomials $g_i(x) \in K[x]$ of degree less than e_i $(i = 1, \ldots, t)$ such that*

$$G_n = g_1(n)\alpha_1^n + \cdots + g_t(n)\alpha_t^n \quad (n \geq 0).$$

2. *k*-periodic binary recurrences

Let $k \geq 2$ be an integer, further let q_0, q_1 and a_i, b_i, $i = 0, \ldots, k-1$ denote arbitrary complex numbers with $|q_0| + |q_1| \neq 0$ and $b_0 b_1 \cdots b_{k-1} \neq 0$.

Consider the sequence (q_n) defined by (1.2). By [1] it is known that the terms of (q_n) satisfy the recurrence relation

$$q_n = A_k q_{n-k} - (-1)^k b_0 b_1 \ldots b_{k-1} q_{n-2k} \tag{2.1}$$

of order $2k$, where the coefficient A_k is also described in [1]. Put $D = A_k^2 - 4(-1)^k b_0 b_2 \ldots b_{k-1}$, and let

$$p_k(x) = x^2 - A_k x + (-1)^k b_0 b_1 \cdots b_{k-1}$$

denote the polynomial determined by the characteristic polynomial $z^{2k} - A_k z^k + (-1)^k b_0 b_1 \cdots b_{k-1}$ of the recurrence (2.1) by the substitution $x = z^k$. The not necessarily distinct zeros of $p_k(x)$ are

$$\kappa = \frac{A_k + \sqrt{D}}{2} \quad \text{and} \quad \mu = \frac{A_k - \sqrt{D}}{2}.$$

At this point we would like to use Theorem 1.1, therefore we must distinguish two cases.

2.1. Case $D \neq 0$

If D is nonzero, then κ and μ are distinct. From Theorem 1, we deduce that there exist complex numbers κ_j and μ_j $(j = 1, \ldots, k)$ such that

$$q_n = \underbrace{\sum_{j=1}^{k} \kappa_j \varepsilon^{(j-1)n} \kappa^{n/k}}_{K_n} + \underbrace{\sum_{j=1}^{k} \mu_j \varepsilon^{(j-1)n} \kappa^{n/k}}_{M_n}, \tag{2.2}$$

where $\varepsilon = \exp(2\pi i/k)$ is a primitive root of unity of order k. If one claims to determine the coefficients κ_j and μ_j, it is sufficient to replace n by $0,\ 1,\ \ldots,\ 2k-1$ in (2.2) and, after evaluating $q_2, \ldots, q_{2k-1}$ by (1.2), to solve the system of $2k$ linear equations. Instead, we can shorten the calculations since, as we will see soon, only certain linear combinations of $\kappa_1, \ldots, \kappa_k$ and $\mu_1, \ldots, \mu_k$ are needed, respectively.

Now, by (2.2), for any non-negative integer t, we have $q_t = K_t + M_t$. Moreover,

$$q_{t+k} = \sum_{j=1}^{k} \kappa_j \varepsilon^{(j-1)(t+k)} \kappa^{(t+k)/k} + \sum_{j=1}^{k} \mu_j \varepsilon^{(j-1)(t+k)} \kappa^{(t+k)/k} = \kappa K_t + \mu M_t. \quad (2.3)$$

Since the determinant $\mu - \kappa$ of the system of two linear equations

$$\begin{cases} K_t + M_t = q_t \\ \kappa K_t + \mu M_t = q_{t+k} \end{cases} \quad (2.4)$$

is non-zero, therefore (2.4) possesses the unique solution

$$K_t = \frac{q_{t+k} - \mu q_t}{\kappa - \mu}, \quad M_t = -\frac{q_{t+k} - \kappa q_t}{\kappa - \mu}.$$

To give the explicit formula for the term of the sequence (q_n), we use the technique described in (2.3) for $n = sk + t$ and t with $0 \le t < k$. It is easy to see that $q_n = q_{sk+t} = \kappa^s K_t + \mu^s M_t$. Hence we proved the following theorem.

Theorem 2.1. *In the case $D \neq 0$, the n^{th} term of the sequence (q_n) satisfies*

$$q_n = \frac{q_{k+(n \bmod k)} - \mu q_{n \bmod k}}{\kappa - \mu} \kappa^{\lfloor n/k \rfloor} - \frac{q_{k+(n \bmod k)} - \kappa q_{n \bmod k}}{\kappa - \mu} \mu^{\lfloor n/k \rfloor}.$$

2.2. Case $D = 0$

If D is zero, then κ and μ coincide with $A_k/2$. By Theorem 1, there exist complex numbers u_j and v_j, $j = 1, \ldots, k$ such that

$$q_n = \sum_{j=1}^{k} (u_j n + v_j) \varepsilon^{(j-1)n} \kappa^{n/k} = nU_n + V_n, \quad (2.5)$$

where

$$U_n = \sum_{j=1}^{k} u_j \varepsilon^{(j-1)n} \kappa^{n/k}, \quad V_n = \sum_{j=1}^{k} v_j \varepsilon^{(j-1)n} \kappa^{n/k}. \quad (2.6)$$

Then $q_t = tU_t + V_t$, together with (2.5) and (2.6) provides $q_{t+k} = \kappa((t+k)U_t + V_t)$. The unique solution of the system

$$\begin{cases} tU_t + V_t = q_t \\ \kappa(t+k)U_t + \kappa V_t = q_{t+k} \end{cases}$$

is

$$U_t = \frac{q_{t+k} - \kappa q_t}{\kappa k}, \quad V_t = -\frac{tq_{t+k} - (t+k)\kappa q_t}{\kappa k}.$$

Consequently, if $n = sk + t$ with $0 \le t < k$ then, clearly, $q_n = \kappa^s(U_t n + V_t)$, and by the notation

$$\omega = q_{t+k} - \kappa q_t, \quad \nu = tq_{t+k} - (t+k)\kappa q_t,$$

the following theorem holds.

Theorem 2.2. *If $D = 0$ then*

$$q_n = \frac{1}{k}\left(\omega n + \nu\right)\kappa^{\lfloor n/k \rfloor - 1},$$

where $\omega = q_{k+(n \bmod k)} - \kappa q_{n \bmod k}$ and $\nu = -(n \bmod k)\, q_{k+(n \bmod k)} + (k + (n \bmod k))\kappa q_{n \bmod k}$.

Note, that the application of Theorems 2.1 and 2.2 results a more precise formula for the term q_n if k is fixed. In the next two sections, we go into details in the cases $k = 2$ and $k = 3$. We derive Theorem 5 in [2] as a corollary of Theorem 2.1 with $k = 2$.

3. The 2-periodic binary recurrences

Suppose that $bd \neq 0$ and $|q_0| + |q_1| \neq 0$ hold in (1.1). It is known, that the terms of the recurrence (q_n) satisfy the recurrence relation

$$q_n = (ac + b + d)q_{n-2} - bdq_{n-4}, \quad n \ge 4$$

of order four, where the initial values are, obviously, q_0, q_1, $q_2 = aq_1 + bq_0$ and $q_3 = (ac + d)q_1 + bcq_0$. Put $D = (ac + b + d)^2 - 4bd$. Thus the zeros of the polynomial $p_2(x) = x^2 - (ac + b + d)x + bd$ are

$$\kappa = \frac{ac + b + d + \sqrt{D}}{2} \qquad \text{and} \qquad \mu = \frac{ac + b + d - \sqrt{D}}{2}.$$

3.1. Case $D \neq 0$

First assume that n is even, i.d., $t = (n \bmod 2) = 0$ holds in Theorem 2.1. Thus we obtain

$$q_n = \frac{q_2 - \mu q_0}{\kappa - \mu}\,\kappa^{\lfloor n/2 \rfloor} - \frac{q_2 - \kappa q_0}{\kappa - \mu}\,\mu^{\lfloor n/2 \rfloor}.$$

Clearly, $q_2 - \mu q_0 = aq_1 + (b - \mu)q_0$, further $q_2 - \kappa q_0 = aq_1 + (b - \kappa)q_0$.

Suppose now, that n is odd, i.d., $t = 1$. Now Theorem 2.1 results

$$q_n = \frac{q_3 - \mu q_1}{\kappa - \mu}\,\kappa^{\lfloor n/2 \rfloor} - \frac{q_3 - \kappa q_1}{\kappa - \mu}\,\mu^{\lfloor n/2 \rfloor}.$$

Obviously, $q_3 - \mu q_1 = (ac + d - \mu)q_1 + (bc)q_0 = (\kappa - b)q_1 + (bc)q_0$, similarly $q_3 - \kappa q_1 = (\mu - b)q_1 + (bc)q_0$.

To join the even and odd cases together, we introduce

$$e_\kappa = a^{1-\xi(n)}(\kappa - b)^{\xi(n)}q_1 + (b - \mu)^{1-\xi(n)}(bc)^{\xi(n)}q_0$$

and

$$e_\mu = a^{1-\xi(n)}(\mu - b)^{\xi(n)}q_1 + (b - \kappa)^{1-\xi(n)}(bc)^{\xi(n)}q_0,$$

where $\xi(n) = (n \bmod 2)$ is the parity function. Thus

$$q_n = \frac{e_\kappa\, \kappa^{\lfloor n/2 \rfloor} - e_\mu\, \mu^{\lfloor n/2 \rfloor}}{\kappa - \mu}. \tag{3.1}$$

Observe that (3.1) returns with the explicit formula given in Theorem 5 of [2] if $b = d = 1$ and $q_0 = 0$, $q_1 = 1$. Indeed, now $e_\kappa = a^{1-\xi(n)}(\kappa - 1)^{\xi(n)}$, $e_\mu = a^{1-\xi(n)}(\mu - 1)^{\xi(n)}$, which together with $ac\kappa = (\kappa - 1)^2$ and $ac\mu = (\mu - 1)^2$ provide

$$q_n = \frac{a^{1-\xi(n)}}{(ac)^{\lfloor n/2 \rfloor}} \frac{(\kappa - 1)^n - (\mu - 1)^n}{(\kappa - 1) - (\mu - 1)}. \tag{3.2}$$

Clearly, by $\alpha = \kappa - 1$ and $\beta = \mu - 1$, (3.2) coincides with the statement of Theorem 5 in [2].

3.2. Case $D = 0$

Note, that neither [2] nor [6] worked this subcase out. Observe, that $D = 0$ is possible, for example, let $b = rs^2$, $d = rt^2$, further $a = r$ and $c = 4st - s^2 - t^2$. Clearly, $\kappa = \mu = (ac + b + d)/2$.

Assume first that n is even, or equivalently $t = 0$. Then $\omega = q_2 - \kappa q_0 = aq_1 + (b - \kappa)q_0$, while $\nu = 2\kappa q_0$.

Supposing $t = 1$, it gives $\omega = q_3 - \kappa q_1 = (ac + d - \kappa)q_1 + (bc)q_0 = (\kappa - b)q_1 + (bc)q_0$ and $\nu = -(q_3 - 3\kappa q_1) = (\kappa + b)q_1 - (bc)q_0$.

Henceforward,

$$q_n = \frac{1}{2}(\omega n + \nu)\kappa^{\lfloor n/2 \rfloor - 1}$$

describes the general case, where $\omega = a^{1-\xi(n)}(\kappa - b)^{\xi(n)}q_1 + (b - \kappa)^{1-\xi(n)}(bc)^{\xi(n)}q_0$ and $\nu = \xi(n)(\kappa + b)q_1 + (-1)^{\xi(n)}(2\kappa)^{1-\xi(n)}(bc)^{\xi(n)}q_0$.

4. The 3-periodic binary recurrences

This section follows the structure of the previous one. Let a, b, c, d, e, f and q_0, q_1 are arbitrary complex numbers with $bdf \neq 0$ and $|q_0| + |q_1| \neq 0$. For $n \geq 2$, the terms of the sequence (q_n) are defined by

$$q_n = \begin{cases} aq_{n-1} + bq_{n-2}, & \text{if } n \equiv 0 \,(\mathrm{mod}\, 3); \\ cq_{n-1} + dq_{n-2}, & \text{if } n \equiv 1 \,(\mathrm{mod}\, 3); \\ eq_{n-1} + fq_{n-2}, & \text{if } n \equiv 2 \,(\mathrm{mod}\, 3). \end{cases}$$

It is known, that recurrence (q_n) satisfies the recurrence relation

$$q_n = (ace + bc + de + af)\, q_{n-3} + bdf q_{n-6}$$

of order six, where the initial values are

$$\begin{aligned} q_0, q_1, q_2 &= eq_1 + fq_0, \\ q_3 &= (ae + b)\, q_1 + af q_0, \\ q_4 &= (ace + bc + de)\, q_1 + (acf + df)\, q_0, \\ q_5 &= \left(ace^2 + bce + de^2 + aef + bf\right) q_1 + \left(acef + def + af^2\right) q_0. \end{aligned}$$

Put $D = (ace + bc + de + af)^2 + 4bdf$. Thus, the roots of the polynomial

$$p_3(x) = x^2 - (ace + bc + de + af)\, x - bdf$$

are

$$\kappa = \frac{(ace + bc + de + af) + \sqrt{D}}{2} \quad \text{and} \quad \mu = \frac{(ace + bc + de + af) - \sqrt{D}}{2}.$$

In the sequel, we need the sequence (a_n) defined by $a_n = 1$ if 3 divides n, and $a_n = 0$ otherwise.

4.1. Case $D \neq 0$

The consequence of Theorem 2.1 is the nice formula

$$q_n = \frac{e_\kappa \kappa^{\lfloor n/3 \rfloor} - e_\mu \mu^{\lfloor n/3 \rfloor}}{\kappa - \mu},$$

where

$$\begin{aligned} e_\kappa = {} & (ae + b)^{a_n} (\kappa - af)^{a_{n+2}} (e\kappa + fb)^{a_{n+1}} q_1 \\ & + (af - \mu)^{a_n} (f(ac + d))^{a_{n+2}} (f(\kappa - bc))^{a_{n+1}} q_0, \end{aligned}$$

and

$$\begin{aligned} e_\mu = {} & (ae + b)^{a_n} (\mu - af)^{a_{n+2}} (e\mu + fb)^{a_{n+1}} q_1 \\ & + (af - \kappa)^{a_n} (f(ac + d))^{a_{n+2}} (f(\mu - bc))^{a_{n+1}} q_0. \end{aligned}$$

Indeed, for $t = 0, 1, 2$

$$q_{t+3} - \mu q_t = \begin{cases} (ae + b)q_1 + (af - \mu)q_0, & \text{if } t = 0; \\ (\kappa - af)q_1 + (ac + d)fq_0, & \text{if } t = 1; \\ (e\kappa + fb)q_1 + (\kappa - bc)fq_0, & \text{if } t = 2, \end{cases} \tag{4.1}$$

and $q_{t+3} - \kappa q_t$ can similarly be obtained from (4.1) by switching κ and μ.

4.2. Case $D = 0$

When $t = 0$ we obtain $\omega = (ae + b)q_1 + (af - \kappa)q_0$, $\nu = 3\kappa q_0$. Secondly, $t = 1$ yields $\omega = (\kappa - af)q_1 + (ac + d)fq_0$ and $\nu = (2\kappa + af)q_1 - (ac + d)fq_0$. Finally, $\omega = (\kappa e + bf)q_1 + (\kappa - bc)fq_0$ and $\nu = (\kappa e - 2bf)q_1 + (\kappa + 2bc)fq_0$ when $t = 2$.

So, we obtain

$$q_n = \frac{1}{3}\left(\omega n + \nu\right)\kappa^{\lfloor n/3 \rfloor - 1},$$

where

$$\begin{aligned} w = {} & (ae + b)^{a_n}\left(\kappa - af\right)^{a_{n+2}}\left(\kappa e - bf\right)^{a_{n+1}} q_1 \\ & + \left(af - \kappa\right)^{a_n}\left((ac + d)\, f\right)^{a_{n+1}}\left((\kappa - bc)f\right)^{a_{n+2}} q_0 \end{aligned}$$

and

$$\begin{aligned} \nu = {} & (1 - a_n)\left(2\kappa + af\right)^{a_{n+2}}\left(\kappa e - 2bf\right)^{a_{n+1}} q_1 \\ & + \left(3\kappa\right)^{a_n}\left(-(ac + d)f\right)^{a_{n+2}}\left((\kappa + 2bc)f\right)^{a_{n+1}} q_0. \end{aligned}$$

5. Constant subsequences in 2-periodic binary recurrences

In the last section we solve the problem posed in 2.2.2 of [6]. There, after pointing on few examples, the author claim a general sufficiency condition for the sequence (1.1) to be constant from a term q_ν (actually, $\nu = 1$ was asked in [6]). The forthcoming theorem describes the complete answer.

Theorem 5.1. *The sequence (q_n) takes the constant value $q \in \mathbb{C}$ from the ν^{th} terms ($\nu \geq 0$) if and only if one of the following cases holds.*

1. *$q_0 = q_1 = 0$, further $a,\ b,\ c,\ d$ are arbitrary, ($\nu = 0$, $q = 0$),*
2. *$q_0 = q_1 = q \neq 0$, $a + b = 1$, $c + d = 1$, ($\nu = 0$, $q \neq 0$),*
3. *$q_0 \neq 0$ is arbitrary, $q_1 = 0$, $b = 0$, moreover $a,\ c,\ d$ are arbitrary, ($\nu = 1$, $q = 0$),*
4. *$q_0 \neq q$ is arbitrary and $q_1 = q$ with $q \neq 0$, and $a = 1$, $b = 0$, $c + d = 1$, ($\nu = 1$, $q \neq 0$),*
5. *q_0 and $q_1 \neq 0$ are arbitrary, $b,\ c$ are arbitrary, $a = -bq_0/q_1$, $d = 0$, ($\nu = 2$, $q = 0$),*
6. *q_0 and $q_1 \neq q$ are arbitrary with $q_1 \neq q_0$ and $q = aq_1 + bq_0$, where $a + b = 1$, $a \neq 1$, $c = 1$, $d = 0$, ($\nu = 2$, $q \neq 0$),*
7. *q_0 and $q_1 \neq 0$ are arbitrary, $a \neq 0$ and c are arbitrary, $b = 0$, $d = -ac$, ($\nu = 3$, $q = 0$),*

8. *q_0 and $q_1 \neq cq_0$ are arbitrary, where $a \neq 0$ and $c \neq 0$ are arbitrary, $b = -ac$, $d = 0$, ($\nu = 4$, $q = 0$).*

Proof. Obviously, each of the conditions appearing in Theorem 5.1 is sufficient. We are going to show that one of them is necessary. Suppose that the sequence (q_n) takes the constant value $q \in \mathbb{C}$ from the ν^{th} terms.

I. First assume that $\nu \geq 5$ is an integer. We introduce the notation $(u, v) = (a, b)$ and $(\breve{u}, \breve{v}) = (c, d)$ if ν is odd, while $(u, v) = (c, d)$ and $(\breve{u}, \breve{v}) = (a, b)$ if ν is even. Then the equations

$$\begin{aligned} q_{\nu-3} &= uq_{\nu-4} + vq_{\nu-5} & \qquad q_{\nu-2} &= \breve{u}q_{\nu-3} + \breve{v}q_{\nu-4} \\ q_{\nu-1} &= uq_{\nu-2} + vq_{\nu-3} & \qquad q &= \breve{u}q_{\nu-1} + \breve{v}q_{\nu-2} \\ q &= uq + vq_{\nu-1} & \qquad q &= \breve{u}q + \breve{v}q \\ q &= uq + vq \end{aligned}$$

hold, where $q \neq q_{\nu-1}$. The last two equations in the left column imply $v(q_{\nu-1} - q) = 0$. Therefore $v = 0$ follows, and it simplifies the whole left column.

If $q \neq 0$ then $u = 1$ and $\breve{u} + \breve{v} = 1$ fulfill. Hence $q_{\nu-1} = q_{\nu-2}$, consequently $q = \breve{u}q_{\nu-1} + \breve{v}q_{\nu-2}$ leads to $q = q_{\nu-1}$ and we arrived at a contradiction.

Consider now the case $q = 0$. Thus $q_{\nu-1} \neq 0$, and then we have the system

$$\begin{aligned} q_{\nu-3} &= uq_{\nu-4} & \qquad q_{\nu-2} &= \breve{u}q_{\nu-3} + \breve{v}q_{\nu-4} \\ q_{\nu-1} &= uq_{\nu-2} & \qquad 0 &= \breve{u}q_{\nu-1} + \breve{v}q_{\nu-2} \end{aligned}$$

to examine. Clearly, $uq_{\nu-2} \neq 0$. The equalities in the second row provide $0 = u\breve{u}q_{\nu-2} + \breve{v}q_{\nu-2}$, subsequently $(u\breve{u} + \breve{v})q_{\nu-2} = 0$, and then $u\breve{u} + \breve{v} = 0$. Insert it to $q_{\nu-2} = u\breve{u}q_{\nu-4} + \breve{v}q_{\nu-4}$ (coming from the first row), and we obtain $q_{\nu-2} = 0$, which is impossible.

Hence, we have shown that if the constant subsequence of (q_n) starts at the term q_ν, then necessarily $\nu \leq 4$.

II. In the second place we assume that $\nu \leq 4$ and distinguish five cases. Note, that for the subscript $k \geq \nu$ the equalities $q_{k+2} = aq_{k+1} + bq_k$, $q_{k+2} = cq_{k+1} + dq_k$ simplify to

$$q = aq + bq, \quad q = cq + dq, \tag{5.1}$$

respectively.

$\nu = 0$**.** If $q = 0$ then $q_0 = q_1 = 0$ and, trivially, all the coefficients a, b, c and d are arbitrary. If $q \neq 0$ then $q_0 = q_1 = q$ and (5.1) must hold. Consequently, $a + b = 1$ and $c + d = 1$ follow.

$\nu = 1$**.** Here $q_0 \neq q$. Further, $q = aq + bq_0$, together with the first equality of (5.1) provides $b(q_0 - q) = 0$. Thus $b = 0$.

Clearly, $q = 0$ satisfies both (5.1) and $q = aq + bq_0$ without further restrictions on a, b and c.

If q is non-zero, then (5.1) and $b = 0$ imply $a = 1$ and $c + d = 1$.

$\nu = 2$. Besides (5.1), we also have

$$q = aq_1 + bq_0, \quad q = cq + dq_1 \tag{5.2}$$

with $q_1 \neq q$. The last equality and the second property of (5.1) give $d = 0$ via $d(q_1 - q) = 0$.

Assume first $q = 0$. Then, except $0 = aq_1 + bq_0$, all the equalities in (5.1) and (5.2) are fulfilled. Since $q_1 \neq 0$, we can write $a = -bq_0/q_1$. Obviously b and c are arbitrary.

If $q \neq 0$ then $c = 1$ and $a + b = 1$ follow. The value of the constant q is $aq_1 + bq_0$. Observe, that $a \neq 1$ otherwise $b = 0$, and then $q_1 = q$ would come.

$\nu = 3$. Now $q_2 \neq q$. The conditions $q_2 = aq_1 + bq_0$, $q = cq_2 + dq_1$, $q = aq + bq_2$ and (5.1) are valid. Thus $b(q_2 - q)$ vanish, i.e. $b = 0$. Hence we obtain the system

$$\begin{array}{rcl} q_2 & = & aq_1 \\ q & = & aq \end{array} \qquad\qquad \begin{array}{rcl} q & = & cq_2 + dq_1 \\ q & = & cq + dq \end{array}$$

Suppose first that $q = 0$. Then $q_2 = aq_1$ and $0 = cq_2 + dq_1$ provide $0 = (ac + d)q_1$. Since $q_1 = 0$ would give $q_2 = 0$ therefore $ac + d$ must be zero, so $d = -ac$. Also $a \neq 0$ holds, otherwise $q_2 = 0$ leads to a contradiction. Clearly, c is arbitrary.

Assume now that q is non-zero. Thus, from the last system above, we conclude $a = 1$, $c + d = 1$ and $q_2 = q_1$. Hence, the remaining equation $q = cq_2 + dq_1$ becomes $q = cq_2 + (1 - c)q_2$, and we arrived at a contradiction by $q \neq q_2$. Subsequently, $q \neq 0$ does not provide a constant sequence from the third term.

$\nu = 4$. The technique we apply resembles us to the previous cases. Here $q_3 \neq q$. We have $q_2 = aq_1 + bq_0$, $q_3 = cq_2 + dq_1$, $q = aq_3 + bq_2$, $q = cq + dq_3$ and (5.1). Similarly, $d(q_3 - q)$ implies $d = 0$. Thus

$$\begin{array}{l} q_2 = aq_1 + bq_0 \\ q = aq_3 + bq_2 \\ q = aq + bq \end{array} \qquad\qquad \begin{array}{l} q_3 = cq_2 \\ q = cq \end{array}$$

If $q = 0$ then $q_3 = cq_2 \neq 0$, further $0 = aq_3 + bq_2$ and $q_3 = cq_2$ yield $ac + b = 0$. Clearly, $c \neq 0$. Moreover $a \neq 0$ holds, otherwise $b = 0$ and $q_2 = 0$ and $q_3 = 0$ follow. Finally, $q_1 \neq cq_0$ since $q_2 \neq 0$.

The assertion $q \neq 0$, similarly to the case $\nu = 3$, leads to a contradiction. □

References

[1] Cooper, C., An identity for period k second order linear recurrence systems, *Congr. Numer.*, 200 (2010), 95–106.

[2] Edson, M., Yayenie, O., A new generalization of Fibonacci sequence and extended Binet's formula, *Integers*, 9 (2009), 639–654.

[3] Edson, M., Lewis, S., Yayenie, O., The k-periodic Fibonacci sequence and an extended Binet's formula, *Integers*, 11 (2011), Paper #A32.

[4] Shallit, J., Numeration systems, linear recurrences, and regular sets, *Inform. and Comput.* 113 (1994), 331–347.

[5] Shorey, T. N., Tijdeman, R., *Exponential Diophantine Equations*, Cambridge University Press, 1986.

[6] Yayenie, O., A note on generalized Fibonacci sequences, *Appl. Math. Comp.* 217 (2011), 5603–5611.

Annales Mathematicae et Informaticae
***40** (2012) pp. 37–46*
http://ami.ektf.hu

Control point based representation of inellipses of triangles*

Imre Juhász

Department of Descriptive Geometry, University of Miskolc, Hungary
agtji@uni-miskolc.hu

Submitted April 12, 2012 — Accepted May 20, 2012

Abstract

We provide a control point based parametric description of inellipses of triangles, where the control points are the vertices of the triangle themselves. We also show, how to convert remarkable inellipses defined by their Brianchon point to control point based description.

Keywords: inellipse, cyclic basis, rational trigonometric curve, Brianchon point

MSC: 65D17, 68U07

1. Introduction

It is well known from elementary projective geometry that there is a two-parameter family of ellipses that are within a given non-degenerate triangle and touch its three sides. Such ellipses can easily be constructed in the traditional way (by means of ruler and compasses), or their implicit equation can be determined. We provide a method using which one can determine the parametric form of these ellipses in a fairly simple way.

Nowadays, in Computer Aided Geometric Design (CAGD) curves are represented mainly in the form

$$\begin{cases} \mathbf{g}(u) = \sum_{j=0}^{n} F_j(u)\,\mathbf{d}_j \\ F_j : [a,b] \to \mathbb{R},\ u \in [a,b] \subset \mathbb{R},\ \mathbf{d}_j \in \mathbb{R}^\delta,\ \delta \geq 2 \end{cases}$$

*This research was carried out as a part of the TAMOP-4.2.1.B-10/2/KONV-2010-0001 project with support by the European Union, co-financed by the European Social Fund.

where $\mathbf{d}_j$ are called control points and $F_j(u)$ are blending functions. (The most well-known blending functions are Bernstein polynomials and normalized B-spline basis functions, cf. [4].)

2. Cyclic curves and their rational extension

In [7] a new set of blending functions - called cyclic basis - have been introduced that are suitable for the description of closed trigonometric curves (the coordinate functions of which are trigonometric polynomials). The cyclic basis of the vector space

$$\mathcal{V}_n = \langle 1, \cos(u), \sin(u), \ldots, \cos(nu), \sin(nu) \rangle$$

of trigonometric polynomials of degree at most $n \geq 1$ is

$$\left\{ C_{i,n}(u) = \frac{c_n}{2^n} \left(1 + \cos\left(u + i\frac{2\pi}{2n+1} \right) \right)^n : u \in [-\pi, \pi] \right\}_{i=0}^{2n}, \tag{2.1}$$

where constant

$$c_n = \frac{2^{2n}}{(2n+1)\binom{2n}{n}}$$

fulfills the recursion

$$\begin{cases} c_1 = \frac{2}{3}, \\ c_n = \frac{2n}{2n+1} c_{n-1}, n \geq 2. \end{cases}$$

Observe, that basis (2.1) consists of 2π-periodic functions, thus we can study the properties of this basis and the corresponding curve on any interval of length 2π.

By means of these basis functions we can specify cyclic curves in the following way.

Definition 2.1. The curve

$$\mathbf{a}(u) = \sum_{i=0}^{2n} C_{i,n}(u)\, \mathbf{d}_i, \ u \in [-\pi, \pi], \tag{2.2}$$

is called cyclic curve of degree $n \geq 1$ that is uniquely determined by its control points $\mathbf{d}_i \in \mathbb{R}^\delta$, $(\delta \geq 2)$ and basis functions (2.1).

Cyclic curves have the following advantageous properties:

- singularity free parametrization (the curve is of C^∞ continuity at all regular points and at singular points non-vanishing left and right derivatives exist);
- convex hull property;
- cyclic symmetry (the shape of the curve does not change when its control points are cyclically permuted);

- closure for the affine transformation of their control points;
- pseudo local controllability;
- variation diminishing.

A simple exact formula has also been provided in [8] for the conversion from the traditional trigonometric representation to the control points based cyclic one. This facilitates the exact control point based description of several remarkable closed curves, such as epi- and hypocycloids, Lissajous curves, torus knots and foliums.

Associating weights with control points of curve (2.2) we can describe closed rational trigonometric polynomial curves in the form

$$\begin{cases} \mathbf{g}(u) = \sum_{i=0}^{2n} \mathbf{d}_i R_{i,n}(u) \\ R_{i,n}(u) = \dfrac{w_i C_{i,n}(u)}{\sum_{j=0}^{2n} w_j C_{j,n}(u)} \end{cases}, u \in [-\pi, \pi] \tag{2.3}$$

where $0 \leq w_i \in \mathbb{R}, \left(\sum_{i=0}^{2n} w_i \neq 0\right)$ are the associated weights (cf. [5]). When all weights are equal, we obtain the cyclic curve (2.2) as a special case.

Curve (2.3) can also be considered as the central projection of the curve

$$\mathbf{g}^w(u) = \sum_{i=0}^{2n} C_{i,n}(u) \begin{bmatrix} w_i \mathbf{d}_i \\ w_i \end{bmatrix}, u \in [-\pi, \pi]$$

in the $\delta + 1$ dimensional space from the origin on to the δ dimensional hyperplane $w = 1$ (assuming that the last coordinate of space $\mathbb{R}^{\delta+1}$ is denoted by w). Curve $\mathbf{g}^w$ is called the pre-image of curve $\mathbf{g}$. This central projection concept facilitates to study the properties of curve $\mathbf{g}$. Curve $\mathbf{g}$ inherits all properties of $\mathbf{g}^w$ that are invariant under central projection, such as continuity, incidence, colinearity and variation diminishing. Curve (2.3) is closed for the projective transformation of its control points, i.e. the curve determined by the transformed control points coincides with the transformed curve. The transformation has to be performed in the pre-image space, therefore not only control points but weights will also be altered.

3. Inellipses of a triangle

Inellipses of a non-degenerate triangle can be constructed by applying the theorem of Brianchon (cf. [2]). The implicit representation of them can also be determined cf. [9]. We provide a control point based parametric representation of inellipses, by means of rational trigonometric curves (2.3).

In [7] the following theorem has been proved.

Theorem 3.1. *If $n = 1$ the curve (2.2) is the ellipse that touches the sides of the control triangle at its midpoints, and the centre of the ellipse is the centroid of the triangle, i.e. the ellipse is the Steiner inellipse of the control triangle.*

In order to describe all inellipses of a triangle, we consider the case $n = 1$ of the rational extension of cyclic curves. Let us denote the position vectors of the vertices of the given triangle by $\mathbf{c}_0, \mathbf{c}_1$ and $\mathbf{c}_2$. The rational trigonometric curve of degree one, determined by these points (as control points) is

$$\begin{cases} \mathbf{g}(u) = \sum_{i=0}^{2} R_{i,1}(u)\,\mathbf{c}_i, \;\; u \in [-\pi, \pi], w_i > 0 \\ R_{i,1}(u) = \dfrac{w_i C_{i,n}(u)}{\sum_{j=0}^{2} w_j C_{j,n}(u)} \\ C_{i,1}(u) = \dfrac{1}{3}\left(1 + \cos\left(u + \dfrac{2\pi i}{3}\right)\right) \end{cases} \tag{3.1}$$

Weights are determined up to a non-zero scaling factor, i.e. weights w_i and λw_i, $0 < \lambda \in \mathbb{R}$ specify the same curve, therefore we can assume without the loss of generality that $w_0 = 1$. Curve (3.1) is also an ellipse, since it is a central projection of an ellipse that does not intersect the vanishing plane if the weights are non-negative (cf. Fig. 1).

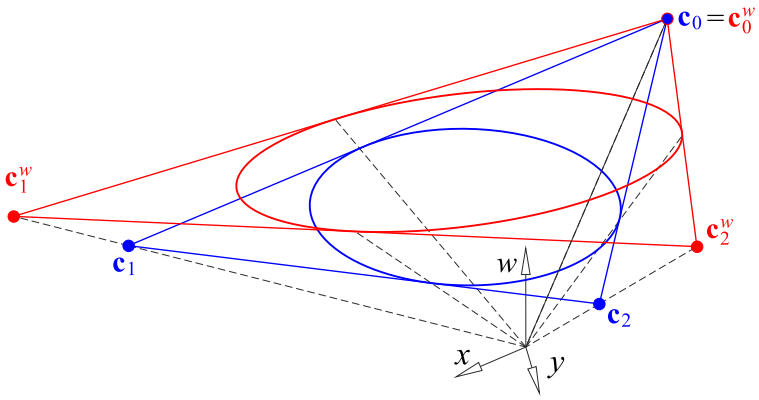


Figure 1: The inscribed ellipse (blue) and its pre-image (red).

Substituting $\pi, \pi/3$ and $-\pi/3$ into $R_{i,1}(u)$ we obtain the values

	$R_{0,1}(u)$	$R_{1,1}(u)$	$R_{2,1}(u)$
π	0	$\frac{w_1}{w_1+w_2}$	$1 - \frac{w_1}{w_1+w_2}$
$\pi/3$	$\frac{1}{1+w_2}$	0	$1 - \frac{1}{1+w_2}$
$-\pi/3$	$\frac{1}{1+w_1}$	$1 - \frac{1}{1+w_1}$	0

which means that the ellipse touches the sides of the control triangle at its

points

$$\begin{aligned}\mathbf{m}_2 &= \tfrac{1}{1+w_1}\mathbf{c}_0 + \left(1 - \tfrac{1}{1+w_1}\right)\mathbf{c}_1,\\ \mathbf{m}_0 &= \tfrac{w_1}{w_1+w_2}\mathbf{c}_1 + \left(1 - \tfrac{w_1}{w_1+w_2}\right)\mathbf{c}_2,\\ \mathbf{m}_1 &= \tfrac{1}{1+w_2}\mathbf{c}_0 + \left(1 - \tfrac{1}{1+w_2}\right)\mathbf{c}_2.\end{aligned} \tag{3.2}$$

Thus, the curve

$$\mathbf{g}(u) = \frac{\mathbf{c}_0 C_{0,1}(u) + w_1\mathbf{c}_1 C_{1,1}(u) + w_2\mathbf{c}_2 C_{2,1}(u)}{C_{0,1}(u) + w_1 C_{1,1}(u) + w_2 C_{2,1}(u)} \tag{3.3}$$

is the parametric representation of the two-parameter ($w_1, w_2 \in \mathbb{R}$) family of inellipses of the triangle with vertices $\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2$, and the points of contact are determined by equalities (3.2).

Two points of contact can arbitrarily be chosen on the sides, from which weights can be determined that uniquely specifies the corresponding inellipse. Let us assume that points of contact $\mathbf{m}_0$ and $\mathbf{m}_1$ are specified on the sides $\mathbf{c}_1\mathbf{c}_2$ and $\mathbf{c}_0\mathbf{c}_2$, respectively. Point $\mathbf{m}_0$ can be written as a barycentric combination of points $\mathbf{c}_1$ and $\mathbf{c}_2$ in the form

$$\mathbf{m}_0 = \alpha\mathbf{c}_1 + (1-\alpha)\,\mathbf{c}_2,\ \alpha \in (0,1)$$

analogously

$$\mathbf{m}_1 = \beta\mathbf{c}_0 + (1-\beta)\,\mathbf{c}_2,\ \beta \in (0,1)$$

Since, the barycentric combination of points of straight line segments with respect to the endpoints is unique, we have the equalities

$$\alpha = \frac{w_1}{w_1 + w_2},\ \beta = \frac{1}{1+w_2}$$

from which we obtain the weights

$$w_2 = \frac{1-\beta}{\beta},\ w_1 = \frac{\alpha(1-\beta)}{\beta(1-\alpha)}$$

that are needed for the representation (3.3).

There are several remarkable inellipses of triangles, a collection can be found at [9]. These are often specified by the trilinear coordinates of their Brianchon point. The Brianchon point of an inellipse of a triangle is the common point of those lines that join point of contacts with the opposite vertices of the triangle, cf. Fig. 2.

Trilinear coordinates (α, β, γ) of a point $\mathbf{p}$ with respect to a reference triangle are an ordered triplet of numbers, each of which is proportional to the directed distance from $\mathbf{p}$ to one of the sides (cf. [10]). The relation between the directed distances and the trilinear coordinates is

$$\begin{aligned}a_1 = k\alpha,\ b_1 = k\beta,\ c_1 = k\gamma\\ k = \frac{2\Delta}{a\alpha + b\beta + c\gamma},\end{aligned} \tag{3.4}$$

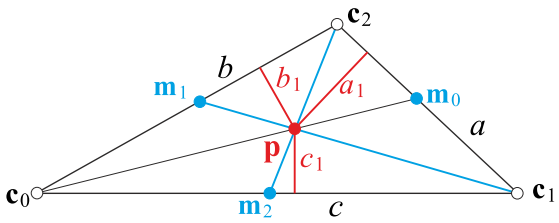


Figure 2: Trilinear coordinates.

where a, b, c denotes the length of the sides, a_1, b_1, c_1 are the corresponding directed distances (cf. Fig. 2) and Δ is the area of the triangle. Distances a_1, b_1, c_1 are also called exact trilinear coordinates.

In what follows, we show how to compute the weights of an inellipse from the trilinear coordinates of its Brianchon point. Assuming that we know the trilinear coordinates (α, β, γ) of the Brianchon point $\mathbf{p}$ of the inellipse, we can calculate its exact trilinear coordinates (a_1, b_1, c_1) by means of (3.4) and the barycentric coordinates (p_0, p_1, p_2) of $\mathbf{p}$ with respect to the vertices $\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2$ in the form

$$p_0 = \frac{aa_1}{2\Delta},\ p_1 = \frac{bb_1}{2\Delta},\ p_2 = \frac{cc_1}{2\Delta},$$

i.e.

$$\mathbf{p} = p_0\mathbf{c}_0 + p_1\mathbf{c}_1 + p_2\mathbf{c}_2.$$

Since points of contact of the inellipse and the triangle are those points where the joining line of the Brianchon point and the vertices meet the opposite sides of the triangle (cf. Fig. 2), point of contact $\mathbf{m}_2$ can be written in the form

$$\mathbf{m}_2 = \lambda\mathbf{c}_2 + (1-\lambda)\,\mathbf{p},\ \lambda \in \mathbb{R},$$

i.e.

$$\begin{aligned}\mathbf{m}_2 &= \lambda\mathbf{c}_2 + (1-\lambda)\,(p_0\mathbf{c}_0 + p_1\mathbf{c}_1 + p_2\mathbf{c}_2)\\ &= (1-\lambda)\,p_0\mathbf{c}_0 + (1-\lambda)\,p_1\mathbf{c}_1 + (\lambda + (1-\lambda)\,p_2)\,\mathbf{c}_2.\end{aligned}$$

Point $\mathbf{m}_2$ is on the side $\mathbf{c}_0\mathbf{c}_1$, thus

$$\lambda + (1-\lambda)\,p_2 = 0$$

from which

$$\lambda = \frac{-p_2}{1-p_2},$$

consequently

$$\mathbf{m}_2 = \frac{p_0}{1-p_2}\mathbf{c}_0 + \frac{p_1}{1-p_2}\mathbf{c}_1$$

$$=\frac{p_0}{1-p_2}\mathbf{c}_0+\left(1-\frac{p_0}{1-p_2}\right)\mathbf{c}_1.$$

Applying considerations above, points of contact $\mathbf{m}_0$ and $\mathbf{m}_1$ are

$$\mathbf{m}_0=\frac{p_1}{1-p_0}\mathbf{c}_1+\left(1-\frac{p_1}{1-p_0}\right)\mathbf{c}_2,$$
$$\mathbf{m}_1=\frac{p_0}{1-p_1}\mathbf{c}_0+\left(1-\frac{p_0}{1-p_1}\right)\mathbf{c}_2.$$

The corresponding weights are

$$w_1=\frac{\frac{p_1}{1-p_0}\left(1-\frac{p_0}{1-p_b}\right)}{\frac{p_0}{1-p_1}\left(1-\frac{p_1}{1-p_0}\right)}=\frac{p_1}{p_0}$$

and

$$w_2=\frac{1-\frac{p_0}{1-p_1}}{\frac{p_0}{1-p_1}}=\frac{p_2}{p_0},$$

i.e. the weights of the rational representation can easily be obtained from the barycentric coordinates of the Brianchon point. In the next subsection we specify the weights of some remarkable inellipses. We will assume that $w_0=1$.

3.1. Incircle

Incircle can be considered as a special inellipse (of equal axes), the barycentric coordinates of its Brianchon point are

$$p_0=\frac{-a^2+b^2+c^2-2bc}{a^2+b^2+c^2-2\left(ab+ac+bc\right)},$$
$$p_1=\frac{a^2-b^2+c^2-2ac}{a^2+b^2+c^2-2\left(ab+ac+bc\right)},$$
$$p_2=\frac{a^2+b^2-c^2-2ab}{a^2+b^2+c^2-2\left(ab+ac+bc\right)},$$

therefore the missing weights are

$$w_1=\frac{b-a+c}{a-b+c},$$
$$w_2=\frac{b-a+c}{a-b+c}.$$

3.2. Brocard inellipse

The Brocard inellipse of a triangle touches the sides at the intersections of the sides with the symmedians, cf. [1]. A symmedian of a triangle is a line obtained by reflecting a median with respect to the corresponding angular bisector, cf. Fig.3

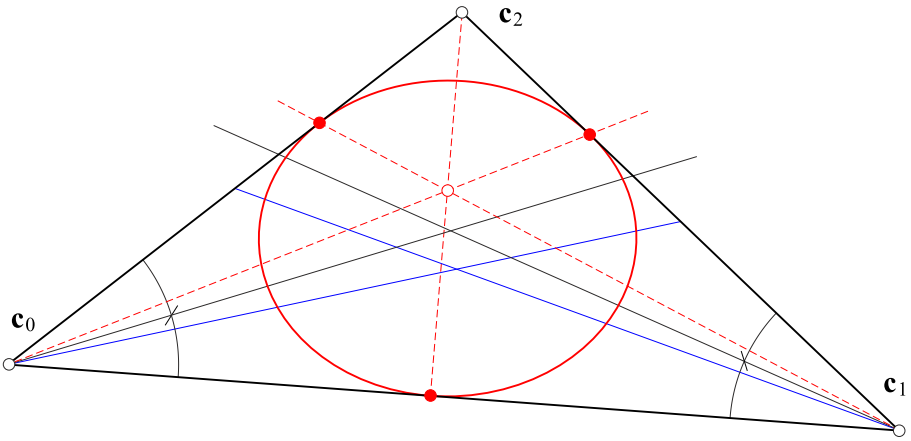


Figure 3: Brocard inellipse

Therefore, the Brianchon point of the Brocard inellipse is the symmedian point of the triangle, and its trilinear coordinates are

$$\alpha = a,\ \beta = b,\ \gamma = c,$$

from which we can compute its exact trilinear coordinates

$$\begin{aligned} a_1 &= \frac{2\Delta a}{a^2+b^2+c^2}, \\ b_1 &= \frac{2\Delta b}{a^2+b^2+c^2}, \\ c_1 &= \frac{2\Delta c}{a^2+b^2+c^2} \end{aligned}$$

and its barycentric coordinates

$$\begin{aligned} p_0 &= \frac{a^2}{a^2+b^2+c^2}, \\ p_1 &= \frac{b^2}{a^2+b^2+c^2}, \\ p_2 &= \frac{c^2}{a^2+b^2+c^2}. \end{aligned}$$

These determine the weights

$$w_1 = \frac{b^2}{a^2} \text{ and } w_2 = \frac{c^2}{a^2}.$$

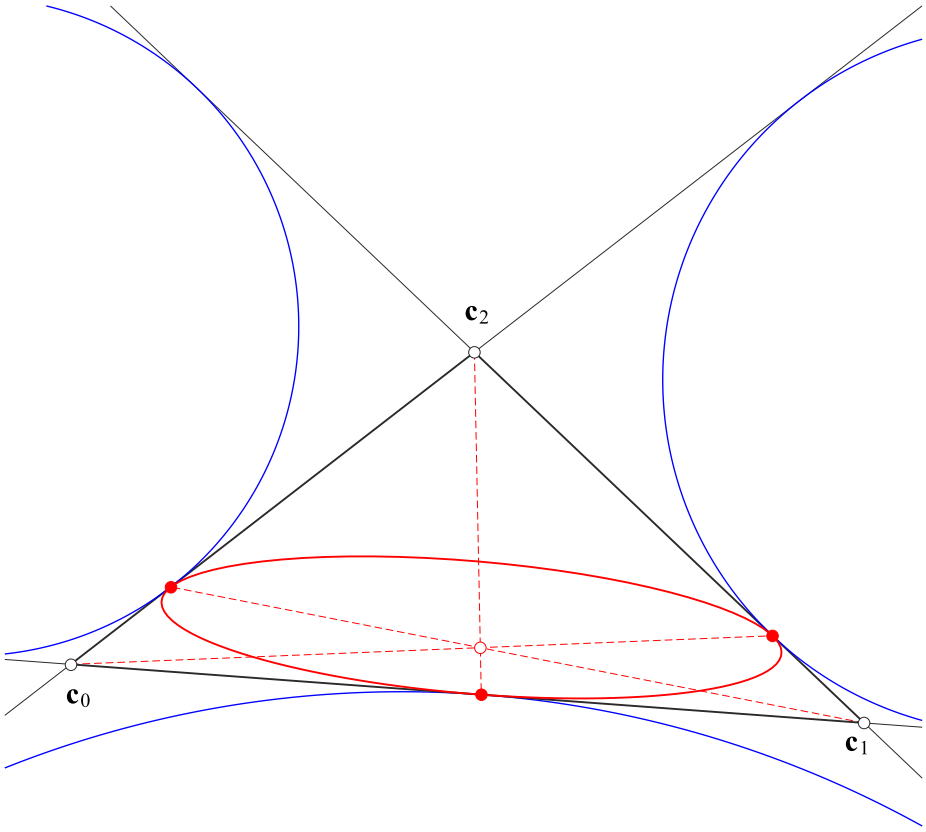


Figure 4: Mandart inellipse

3.3. Mandart inellipse

The Mandart inellipse (cf. [6], [3]) touches the three sides of the triangle at those points where the three external touching (extouch) circles do, cf. Fig. 4. The trilinear coordinates of the corresponding Brianchon point are

$$\alpha = \frac{-a+b+c}{a},\ \beta = \frac{a-b+c}{b},\ \gamma = \frac{a+b-c}{c}.$$

Its exact trilinear coordinates are

$$a_1 = \frac{2\Delta\,(-a+b+c)}{a\,(a+b+c)},$$
$$b_1 = \frac{2\Delta\,(a-b+c)}{b\,(a+b+c)},$$
$$c_1 = \frac{2\Delta\,(a+b-c)}{c\,(a+b+c)}$$

its barycentric coordinates are

$$p_0 = \frac{-a+b+c}{a+b+c},$$
$$p_1 = \frac{a-b+c}{a+b+c},$$
$$p_2 = \frac{a+b-c}{a+b+c}$$

and the pending weights are

$$w_1 = \frac{a-b+c}{-a+b+c} \text{ and } w_2 = \frac{a+b-c}{-a+b+c}.$$

References

[1] Cavallaro, V., On Brocard's Ellipse, *National Mathematics Magazine*, 14(8) (1940), 445–448.

[2] Coxeter, H. S. M., *Projective Geometry (2nd ed.)*, Springer-Verlag, 1987.

[3] Gibert, B., Generalized Mandart Conics, *Forum Geometricorum*, 4 (2004), 177–198.

[4] Hoschek, J., Lasser, D., *Fundamentals of Computer Aided Geometric Design*, AK Peters, Wellesley, 1993.

[5] Juhász, I., Róth, Á., Closed rational trigonometric curves and surfaces, *Journal of Computational and Applied Mathematics*, 234(8) (2010) 2390–2404 doi: 10.1016/j.cam.2010.03.009

[6] Mandart, H., Sur une ellipse associée au triangle, *Mathesis*, 1894., 241–245.

[7] Róth, Á., Juhász, I., Schicho, J., Hoffmann, M., A cyclic basis for closed curve and surface modeling, *Computer Aided Geometric Design*, 26(5) (2009) 528–546, doi:10.1016/j.cagd.2009.02.002

[8] Róth, Á., Juhász, I., Control point based exact description of a class of closed curves and surfaces, *Computer Aided Geometric Design*, 27(2) (2010) 179–201, doi: 10.1016/j.cagd.2009

[9] Weisstein, E. W., Inellipse, *From MathWorld–A Wolfram Web Resource*, http://mathworld.wolfram.com/Inellipse.html

[10] Weisstein, E. W., Trilinear Coordinates, *From MathWorld–A Wolfram Web Resource*, http://mathworld.wolfram.com/TrilinearCoordinates.html

Annales Mathematicae et Informaticae
***40** (2012) pp. 47–55*
http://ami.ektf.hu

About the equation $B_m^{(a,b)} = f(x)^*$

Kálmán Liptai[a], Péter Olajos[b]

[a]Eszterházy Károly College,
Institute of Mathematics and Informatics,
H–3300 Eger, Eszterházy tér 1, Hungary
liptaik@ektf.hu

[b]University of Miskolc
Department of Applied Mathematics
H–3515 Miskolc-Egyetemváros, Hungary
matolaj@uni-miskolc.hu

Submitted April 9, 2012 — Accepted September 24, 2012

Abstract

Let a, b be nonnegative coprime integers. We call an integer $an + b \in \mathbb{N}$ (denoted by $B_m^{(a,b)}$) an (a, b)-type balancing number if

$$(a+b) + (2a+b) + \cdots + (a(n-1)+b) = (a(n+1)+b) + \cdots + (a(n+r)+b)$$

for some $r \in \mathbb{N}$.

In this paper we consider and give numerical results for the equation $B_m^{(a,b)} = f(x)$ where $B_m^{(a,b)}$ is an (a, b)-type balancing number and $f(x)$ is a polynomial belonging to combinatorial numbers (that is binomial coefficients, power sums and products of consecutive integers).

Moreover we investigate the equation when an (a, b)-type balancing number with different parameters are equal to a Fibonacci or a Lucas number. In this case we use a parallel program to find the solutions of simultaneous Pell equations.

Keywords: balancing numbers, elliptic curves, Magma, combinatorial numbers, parallel algorithm

MSC: 11D25, 11D41, 11D45

*Supported in part by Grant T-48945, T-48791 from the Hungarian National Foundation for Scientific Research and TéT SK-8/2008. This research was also carried out as part of the TAMOP-4.2.1.B-10/2/KONV-2010-0001 project with support by the European Union, co-financed by the European Social Fund.

1. Introduction

A positive integer n is called a balancing number (see [2] and [4]) if

$$1 + 2 + \cdots + (n-1) = (n+1) + (n+2) + \cdots + (n+r)$$

for some $r \in \mathbb{Z}^+$. Here r is called the balancer corresponding to the balancing number n. Denote by B_m the mth term of the sequence of balancing numbers. For example 6 and 35 are balancing numbers with balancers 2 and 14, respectively.

K. Liptai [5, 6] proved that there is no Fibonacci and Lucas balancing numbers. In these proofs the same method were used which is based on the result of Baker and Davenport (see [1]). Using an other way from L. Szalay [12] got the same result. This method used a program by Magma [9], but later G. Szekrényesi [13] made a parallel program which was faster than earlier one and arbitrarily large coefficients were used. This program used the fast algorithm for finding solutions of "small solutions" of Thue equations or inequalities. In this case we know about the integer solutions (x, y) that $|y| < 10^{500}$. Using this program we investigated the problem of existence of Fibonacci or Lucas numbers among balancing numbers (for details see [13]).

To prove one of our main results we need the following lemma of P. E. Ferguson (see [3]).

Lemma 1.1. *The only solutions of the equation*

$$x^2 - 5y^2 = \pm 4 \tag{1.1}$$

are $x = \pm L_n$, $y = \pm F_n$ $(n = 0, 1, 2, \ldots)$, where L_n and F_n are the n^{th} terms of the Lucas and Fibonacci sequences, respectively.

Later K. Liptai, F. Luca, Á. Pintér and L. Szalay [7] generalized the balancing numbers which are called (k, l)-power numerical center.

Let y, k, l be fixed positive integers with $y \geq 4$. A positive integer x $(x \leq y - 2)$ is called a (k, l)-power numerical center for y if

$$1^k + \cdots + (x-1)^k = (x+1)^l + \cdots + (y-1)^l.$$

They [7] proved several effective and ineffective finiteness statements for (k, l)-power numerical center using Baker-type diophantine results and Bilu-Tichy theorem. There is another generalization of balancing numbers (see [8]).

Let $a > 0$ and $b \geq 0$ be coprime integers. We call an integer $an + b \in \mathbb{N}$ an (a, b)-type balancing number if

$$(a+b) + (2a+b) + \cdots + (a(n-1)+b) = (a(n+1)+b) + \cdots + (a(n+r)+b)$$

for some $n, r \in \mathbb{N}$. Here r is called the balancer corresponding to the balancing number $an + b$ denoted by $B_m^{(a,b)}$.

T. Kovács, K. Liptai and P. Olajos [8] got a simple proposition for (a, b)-type balancing numbers.

Lemma 1.2. *If* $B_m^{(a,b)}$ *is an* (a,b)*-type balancing number then the following equation*

$$z^2 - 8\left(B_m^{(a,b)}\right)^2 = a^2 - 4ab - 4b^2 \tag{1.2}$$

is valid for some $z \in \mathbb{Z}$.

In the case when $a = 2$ and $b = 1$ P. Olajos [10] proved that

$$B_{m+2}^{(2,1)} = 6 \cdot B_{m+1}^{(2,1)} - 1 \cdot B_m, \ (m \geq 1), \text{ where } B_1^{(2,1)} = 17, \ B_2^{(2,1)} = 99.$$

He also considered Fibonacci and Lucas numbers among $(2,1)$-type balancing numbers.

Let us consider the equation

$$B_m^{(a,b)} = f(x) \tag{1.3}$$

where $f(x)$ is a polynomial with integer coefficients.

They [8] proved finiteness results for equation (1.3) in the cases when $f(x)$ is a monic polynomial or perfect power. The authors proved another finiteness result also when $f(x)$ is equal to a combinatorial number.

For all $k, x \in \mathbb{N}$ let

$$\begin{aligned} S_k(x) &= 1^k + 2^k + \cdots + (x-1)^k, \\ T_k(x) &= -1^k + 2^k - \cdots + (-1)^{x-1}(x-1)^k, \\ \Pi_k(x) &= x(x+1)\ldots(x+k-1). \end{aligned}$$

Lemma 1.3. *Let* $k \geq 2$ *and* $f(x)$ *be one of the polynomials* $\binom{x}{k}$, $\Pi_k(x)$, $S_{k-1}(x)$, $T_k(x)$. *Then the solutions of equation* (1.3) *satisfy* $\max(m, |x|) < c_1(a,b,k)$, *where* $c_1(a,b,k)$ *is an effectively computable constant depending only on* a, b *and* k.

In this paper they also considered all solutions $(x,y) \in \mathbb{Z}^2$ of equation

$$B_m^{(a,b)} = f(x)$$

when $a^2 - 4ab - 4b^2 = 1$ and $f(x) \in \{\binom{x}{2}, \binom{x}{3}, \binom{x}{4}\}$, $\Pi_2(x)$, $\Pi_3(x)$, $\Pi_4(x)$, $S_1(x)$, $S_2(x)$, $S_3(x)$, $S_5(x)$. For more details see [8].

Later Sz. Tengely [14] proved that the equation

$$B_m = x(x+1)(x+2)(x+3)(x+4)$$

has no solution. The author combined Baker's method and the so-called Mordell-Weil sieve to obtain all solutions.

The authors [8] partially solved equation (1.3), because they considered only the cases when $a^2 - 4ab - 4b^2 = 1$. In the following chapter we discuss this problem with certain conditions and not only for the cases above.

2. Numerical results

2.1. Results by MAGMA

By Lemma 1.3 we know that there are only finite number of solutions of equation (1.3). In the cases when $a \in [1,9]$, $b \in [0,7]$, $a \geq b$ and $\gcd(a,b) = 1$ we get the following result:

Theorem 2.1. *Let $2 \leq k \leq 4$ and $f(x)$ be one of the polynomials $\binom{x}{k}$, $\Pi_k(x)$, $S_{k-1}(x)$ and $a \in [1,9]$, $b \in [0,7]$ where $a \geq b$ and $\gcd(a,b) = 1$. Then the solutions $(B_m^{(a,b)}, x)$ of equation (1.3) are in the following table:*

a	b	$B_m^{(a,b)}$	$f(x)$	x	k
1	0	1	$\binom{x}{k}$	2	2
1	0	1	$\binom{x}{k}$	3	3
1	0	1	$\binom{x}{k}$	4	4
1	0	6	$\binom{x}{k}$	4	2
1	0	35	$\binom{x}{k}$	7	3
1	0	35	$\binom{x}{k}$	7	4
1	1	4	$\binom{x}{k}$	4	3
1	0	1	$S_{k-1}(x)$	2	2
1	0	6	$S_{k-1}(x)$	4	2
1	0	1	$S_{k-1}(x)$	2	3
1	0	204	$S_{k-1}(x)$	9	3
1	0	1	$S_{k-1}(x)$	2	4
1	0	6	$\Pi_k(x)$	2	2
7	5	600	$\Pi_k(x)$	24	2
1	0	6	$\Pi_k(x)$	1	3

Remark 2.2. We mention that in the case $k = 1$ we get infinitely many solutions for equation (1.3) since in this case the equation is a Pell-equation.

2.2. Results by a parallel program

In this subsection we consider the cases when $B_m^{(a,b)} = F_l$ or $B_m^{(a,b)} = L_p$ where F_l and L_p are Fibonacci and Lucas numbers, respectively. Let us consider the equation (1.1) and (1.2). In the first case above we get the following simultaneous Pell equations:

$$5x^2 - y^2 = \pm 4, \tag{2.1}$$

$$8x^2 - z^2 = -1(a^2 - 4ab - 4b^2), \tag{2.2}$$

where $x = B_m^{(a,b)} = F_l$. In the second case we have to solve the following:

$$x^2 - 5y^2 = \pm 4, \tag{2.3}$$

$$8x^2 - z^2 = -1(a^2 - 4ab - 4b^2), \tag{2.4}$$

where $x = B_m^{(a,b)} = L_p$. We use the parallel program from G. Szekrényesi to get the solutions of the equation systems above. So our numerical results is the following theorem.

Theorem 2.3. *If* $a \in [1, 9]$, $b \in [0, 7]$, $a \geq b$ *and* $\gcd(a, b) = 1$ *we get the following "small solutions" of equations* $B_m^{(a,b)} = F_l$ *or* $B_m^{(a,b)} = L_p$ *detailed in the next tables (that is there is an upper bound for integer unknowns in Thue inequalities which is equal to* 10^{500}*):*

a	b	m	r	$B_m^{(a,b)} = F_l$	l
1	*0*	*1*	*0*	*1*	*1 or 2*
7	*1*	*228*	*94*	*1597*	*17*

a	b	m	r	$B_m^{(a,b)} = L_p$	p
1	*0*	*1*	*0*	*1*	*1*
1	*1*	*3*	*1*	*4*	*1*
1	*1*	*10*	*4*	*11*	*6*

3. Proofs

3.1. Proof of Theorem 2.1

Consider the equation (1.3) when $f(x)$ one of polynomials $\binom{x}{2}$, $\binom{x}{3}$ and $\binom{x}{4}$. Using the transformations $X = 2x - 1$, $X = (x - 1)^2$, $X = x^2 - 3x + 1$ respectively to the polynomials above then we get the following by Lemma 1.2:

$$(2^2 z)^2 = 2X^4 - 4X^2 + 2 - 16C(a, b),$$
$$(6z)^2 = X^3 - 4X^2 + 4X - 36C(a, b),$$
$$\left(2^2 3z\right)^2 = 2X^4 - 4X^2 + 2 - 144C(a, b),$$

where $C(a, b)$ denotes the quantity $-(a^2 - 4ab - 4b^2)$.

These types of equations are solvable by MAGMA (`IntegralQuarticPoints` and `IntegralPoints`), so after testing them we get the solutions above.

Let us consider two example of using MAGMA commands. In the first example set the parameters as the following: $k = 2$, $a = 2$, $b = 1$. In this case we have to use the transformations above that is we have to solve the equation

$$(2^2 z)^2 = 2X^4 - 4X^2 - 126.$$

The suitable command is `IntegralQuarticPoints([2,0,-4,0,-126])`. We get the solutions `(5,32),(3,0)` for $(X, 2^2 z)$. Using these results we know that no solutions for $B_m^{(2,1)}$, because 3 and 1 are not $(2, 1)$-type balancing numbers. We used the property that $B_m^{(2,1)} \geq 17$.

Let us consider the second one. In this case let parameters $k = 3$, $a = 2$, $b = 1$. Our equation is the following:

$$(6z)^2 = X^3 - 4X^2 + 4X - 288.$$

Using the commands `IntegralPoints(EllipticCurve([0,-4,0,4,-288]))` we get the solution $(X, 6z) = (8, 0)$, that is there is no $(2, 1)$-type balancing number with the main property above.

Now let $f(x)$ be equal to $S_{k-1}(x)$. If $k = 2$ then $S_1(x) = \binom{x}{2}$ that is we get the solutions.

Using the transformations $X = 2(2x-1)^2$, $X = \binom{x}{2}$ respectively to the equation (1.2) when $f(x) = S_2(x)$ and $f(x) = S_3(x)$ we get

$$\left(2^3 3z\right)^2 = X^3 - 4X^2 + 4X - 576C(a, b),$$
$$z^2 = 8X^4 - C(a, b).$$

By MAGMA we get the solutions by the commands `IntegralQuarticPoints` and `IntegralPoints` above when $f(x) = S_{k-1}(x)$.

At last let $f(x) = \Pi_2(x)$, $\Pi_3(x)$, $\Pi_4(x)$ and by using the transformations $X = 2x + 1$, $X = 2(x + 1)^2$ and $x^2 + 3x + 1$ we get the following from (1.2)

$$(2z)^2 = 2X^4 - 4X^2 + 2 - 4C(a, b),$$
$$z^2 = X^3 - 4X^2 + 4X - C(a, b),$$
$$z^2 = 8X^4 - 16X^2 + 8 - C(a, b).$$

By MAGMA we get the solutions above. We have to mention that in the case $a = 2$, $b = 1$ of $\Pi_4(x)$ we get a singular equation, because $C(2, 1) = 8$ and the curve $z^2 = 8X^4 - 16X^2$ is singular. Theres is no problem, because $\Pi_4(x)$ is even, but all $B_m^{(2,1)}$ are odd that is there is no solution of the equation (1.3).

3.2. Proof of Theorem 2.3

Let us consider first the case when $B_m^{(a,b)} = F_l$. We have to solve the simultaneous Pell equations by the parallel program (G. Szekrényesi [13]) or by MAGMA (L. Szalay [12]). We used the parallel one to determine the "small" (less then 10^{500}) solutions of system of the equations (2.1) and (2.2). It means that this program besides others could not find all solutions.

Generally the parallel program have been faster than others (e.g Maple, Magma or Kant). It uses the fast algorithm for finding the "small" integer solutions of Thue inequalities in parallel way by the method from Pethő and Schulenberg [11]. The program also containes a solver for simultaneous Pell equations, which is based on the algorithm of L. Szalay [12]. The program could use arbitrarily large coefficients which is inpossible in others.

The results detailed in the next table. I have to mention that the sign + or − in the table below denotes the correct sign of the right hand side of the first equation of our Pell system. Denote the expression $-(a^2 - 4ab - 4b^2)$ by $C(a, b)$ again.

a	b	$C(a,b)$	$+(x,y,z)$	$-(x,y,z)$
1	0	-1	(1,1,3)	(0,2,1); (1,3,3)
1	1	7	(2,4,5); (1,1,1)	(1,3,1)
2	1	8	(1,1,0)	(1,3,0); (3,7,8)
3	1	7	(2,4,5); (1,1,1)	(1,3,1)
3	2	31	(2,4,1); (233,521,659); (5,11,13)	–
4	1	4	(1,1,2); (5,11,14)	(1,3,2)
4	3	68	–	(3,7,2)
5	1	-1	(1,1,3)	(0,2,1); (1,3,3)
5	2	31	(2,4,1); (233,521,659); (5,11,13)	–
5	3	71	–	(8,18,21); (3,7,1)
5	4	119	(5,11,9)	–
6	1	-8	(1,1,4)	(1,3,4)
6	5	184	(5,11,4)	–
7	1	-17	(2,4,7);(13,29,37);(1,1,5);(1597,3571,4517)	(8,18,23); (1,3,5)
7	2	23	(2,4,3)	(3,7,7)
7	3	71	–	(8,18,21); (3,7,1)
7	4	127	(13,29,35)	–
7	5	191	(5,11,3)	–
7	6	263	(13,29,33)	–
8	1	-28	(1,1,6)	(3,7,10); (1,3,6)
8	3	68	–	(3,7,2)
8	5	196	(13,29,34); (5,11,2)	–
8	7	356	–	–
9	1	-41	(1,1,7)	(1,3,7)
9	2	7	(2,4,5); (1,1,1)	(1,3,1)
9	4	127	(13,29,35)	–
9	5	199	(5,11,1)	–
9	7	367	(89,199,251)	–

Using tha data from this table, we can get the solutions of Theorem 2.3 for Fibonacci balancing numbers.

Consider now the cases of Lucas numbers that is the equations (2.3) and (2.4). We get the following solutions detailed in the table below:

a	b	$C(a,b)$	$+(x,y,z)$	$-(x,y,z)$
1	0	-1	–	(1,1,3)
1	1	7	(2,0,5)	(4,2,11); (11,5,31); (1,1,1)
2	1	8	(3,1,8)	(1,1,0)
3	1	7	(2,0,5)	(4,2,11); (11,5,31); (1,1,1)
3	2	31	(2,0,1); (7,3,19)	–
4	1	4	–	(29,13,82); (1,1,2)
4	3	68	(3,1,2); (7,3,18)	(11,5,30)
5	1	-1	–	(1,1,3)
5	2	31	(2,0,1); (7,3,19)	–
5	3	71	(3,1,1)	–
5	4	119	–	(4,2,3)
6	1	-8	(7,3,20)	(1,1,4)
6	5	184	–	(11,5,28)
7	1	-17	(2,0,7); (47,21,133)	(76,34,215); (1,1,5)
7	2	23	(3,1,7); (2,0,3)	–
7	3	71	(3,1,1)	–
7	4	127	–	(4,2,1); (11,5,29)
7	5	191	(18,8,49)	–
7	6	263	–	–
8	1	-28	(3,1,10)	(1,1,6)
8	3	68	(3,1,2); (7,3,18)	(11,5,30)
8	5	196	(7,3,14)	–
8	7	356	(7,3,6)	–
9	1	-41	–	(4,2,13); (1,1,7)
9	2	7	(2,0,5)	(4,2,11); (11,5,31); (1,1,1)
9	4	127	–	(4,2,1); (11,5,29)
9	5	199	–	–
9	7	367	(7,3,5)	–

References

[1] Baker, A., Davenport, H., The equations $3x^2 - 2 = y^2$ and $8x^2 - 7 = z^2$, *Quart. J. Math. Oxford*, (2) 20 (1969), 129–137.

[2] Behera, A., Panda, G. K., On the square roots of triangular numbers, *Fibonacci Quarterly*, 37 No. 2 (1999), 98–105.

[3] Ferguson, D. E., Letter to the editor, *Fibonacci Quarterly*, 8 (1970), 88–89.

[4] Finkelstein, R. P., The House Problem, *American Math. Monthly*, 72 (1965), 1082–1088.

[5] Liptai, K., Fibonacci balancing numbers, *Fibonacci Quarterly*, 42 (2004), 330–340.

[6] Liptai, K., Lucas balancing numbers, *Acta Math. Univ. Ostrav.*, 14 No. 1 (2006), 43–47.

[7] Liptai, K., Luca, F., Pintér, Á., Szalay, L., Generalized balancing numbers, *Indag. Math. N. S.*, 20 (2009), 87–100.

[8] Kovács, T., Liptai, K., Olajos, P., On (a; b)-balancing numbers, *Publ. Math. Debrecen*, 77/3-4 (2010), 485–498.

[9] MAGMA Computational Algebra System, Computational Algebra Group School of Mathematics and Statistics, *University of Sydney*, NSW 2006, Australia. `http://magma.maths.usyd.edu.au/magma/`

[10] Olajos, P., A (2,1) típusú balansz számokról, *GÉP folyóirat*, LXIII. évfolyam, 4 (2012), 59–62.

[11] Pethő, A., Schulenberg, R., Effektives lösen von Thue gleichungen, *Publ. Math. Debrecen*, 34 (1987), 189–196.

[12] Szalay, L., On the resolution of simultaneous Pell equations, *Annales Mathematicae et Informaticae*, 34 (2007), 77–87.

[13] Szekrényesi, G., Parallel algorithm for determining the "small solutions" of Thue equations, *Annales Mathematicae et Informaticae*, (submitted).

[14] Tengely, Sz., Balancing numbers which are products of consecutive integers, *Monatshefte für Mathematik*, (under submission).

Annales Mathematicae et Informaticae
***40** (2012) pp. 57–76*
http://ami.ektf.hu

Polynomials whose coefficients are *k*-Fibonacci numbers

Toufik Mansour, Mark Shattuck

Department of Mathematics, University of Haifa, 31905 Haifa, Israel
tmansour@univ.haifa.ac.il, maarkons@excite.com

Submitted June 1, 2012 — Accepted October 13, 2012

Abstract

Let $\{a_n\}_{n\geq 0}$ denote the linear recursive sequence of order k ($k \geq 2$) defined by the initial values $a_0 = a_1 = \cdots = a_{k-2} = 0$ and $a_{k-1} = 1$ and the recursion $a_n = a_{n-1} + a_{n-2} + \cdots + a_{n-k}$ if $n \geq k$. The a_n are often called *k-Fibonacci* numbers and reduce to the usual *Fibonacci* numbers when $k = 2$. Let $P_{n,k}(x) = a_{k-1}x^n + a_k x^{n-1} + \cdots + a_{n+k-2}x + a_{n+k-1}$, which we will refer to as a *k-Fibonacci coefficient* polynomial. In this paper, we show for all k that the polynomial $P_{n,k}(x)$ has no real zeros if n is even and exactly one real zero if n is odd. This generalizes the known result for the $k = 2$ and $k = 3$ cases corresponding to Fibonacci and Tribonacci coefficient polynomials, respectively. It also improves upon a previous upper bound of approximately k for the number of real zeros of $P_{n,k}(x)$. Finally, we show for all k that the sequence of real zeros of the polynomials $P_{n,k}(x)$ when n is odd converges to the opposite of the positive zero of the characteristic polynomial associated with the sequence a_n. This generalizes a previous result for the case $k = 2$.

Keywords: k-Fibonacci sequence, zeros of polynomials, linear recurrences

MSC: 11C08, 13B25, 11B39, 05A20

1. Introduction

Let the recursive sequence $\{a_n\}_{n\geq 0}$ of order k ($k \geq 2$) be defined by the initial values $a_0 = a_1 = \cdots = a_{k-2} = 0$ and $a_{k-1} = 1$ and the linear recursion

$$a_n = a_{n-1} + a_{n-2} + \cdots + a_{n-k}, \qquad n \geq k. \tag{1.1}$$

The numbers a_n are sometimes referred to as *k-Fibonacci* numbers (or *generalized Fibonacci* numbers) and reduce to the usual *Fibonacci* numbers F_n when $k = 2$ and to the *Tribonacci* numbers T_n when $k = 3$. (See, e.g., A000045 and A000073 in [11].) The sequence a_n was first considered by Knuth [3] and has been a topic of study in enumerative combinatorics. See, for example, [1, Chapter 3] or [9] for interpretations of a_n in terms of linear tilings or k-filtering linear partitions, respectively, and see [10] for a q-generalization of a_n.

Garth, Mills, and Mitchell [2] introduced the definition of the Fibonacci coefficient polynomials $p_n(x) = F_1x^n + F_2x^{n-1} + \cdots + F_nx + F_{n+1}$ and–among other things–determined the number of real zeros of $p_n(x)$. In particular, they showed that $p_n(x)$ has no real zeros if n is even and exactly one real zero if n is odd. Later, this result was extended by Mátyás [5, 6] to more general second order recurrences. The same result also holds for the Tribonacci coefficient polynomials $q_n(x) = T_2x^n + T_3x^{n-1} + \cdots + T_{n+1}x + T_{n+2}$, which was shown by Mátyás and Szalay [8].

If $k \geq 2$ and $n \geq 1$, then define the polynomial $P_{n,k}(x)$ by

$$P_{n,k}(x) = a_{k-1}x^n + a_kx^{n-1} + \cdots + a_{n+k-2}x + a_{n+k-1}. \tag{1.2}$$

We will refer to $P_{n,k}(x)$ as a *k-Fibonacci coefficient* polynomial. Note that when $k = 2$ and $k = 3$, the $P_{n,k}(x)$ reduce to the Fibonacci and Tribonacci coefficient polynomials $p_n(x)$ and $q_n(x)$ mentioned above. In [7], the following result was obtained concerning the number of real zeros of $P_{n,k}(x)$ as a corollary to a more general result involving sequences defined by linear recurrences with non-negative integral weights.

Theorem 1.1. *Let h denote the number of real zeros of the polynomial $P_{n,k}(x)$ defined by* (1.2) *above. Then we have*
(i) $h = k - 2 - 2j$ for some $j = 0, 1, \ldots, (k-2)/2$, if k and n are even,
(ii) $h = k - 1 - 2j$ for some $j = 0, 1, \ldots, (k-2)/2$, if k is even and n is odd,
(iii) $h = k - 1 - 2j$ for some $j = 0, 1, \ldots, (k-1)/2$, if k is odd and n is even,
(iv) $h = k - 2j$ for some $j = 0, 1, \ldots, (k-1)/2$, if k and n are odd.

For example, Theorem 1.1 states when $k = 3$ that the number of real zeros of the polynomial $P_{n,3}(x)$ is either 0 or 2 if n is even or 1 or 3 if n is odd. As already mentioned, it was shown in [8] that $P_{n,3}(x)$ possesses no real zeros when n is even and exactly one real zero when n is odd.

In this paper, we show that the polynomial $P_{n,k}(x)$ possesses the smallest possible number of real zeros in every case and prove the following result.

Theorem 1.2. *Let $k \geq 2$ be a positive integer and $P_{n,k}(x)$ be defined by* (1.2) *above. Then we have the following:*
(i) If n is even, then $P_{n,k}(x)$ has no real zeros.
(ii) If n is odd, then $P_{n,k}(x)$ has exactly one real zero.

We prove Theorem 1.2 as a series of lemmas in the third and fourth sections below, and have considered separately the cases for even and odd k. Combining

Theorems 3.5 and 4.5 below gives Theorem 1.2. The crucial steps in our proofs of Theorems 3.5 and 4.5 are Lemmas 3.2 and 4.2, respectively, where we make a comparison of consecutive derivatives of a polynomial evaluated at the point $x = 1$. This allows us to show that there is exactly one zero when $x \leq -1$ in the case when n is odd. We remark that our proof, when specialized to the cases $k = 2$ and $k = 3$, provides an alternative proof to the ones given in [2] and [8], respectively, in these cases. In the final section, we show for all k that the sequence of real zeros of the polynomials $P_{n,k}(x)$ for n odd converges to $-\lambda$, where λ is the positive zero of the characteristic polynomial associated with the sequence a_n (see Theorem 5.5 below). This generalizes the result for the $k = 2$ case, which was shown in [2].

2. Preliminaries

We seek to determine the number of real zeros of the polynomial $P_{n,k}(x)$. By the following lemma, we may restrict our attention to the case when $x \leq -1$.

Lemma 2.1. *If $k \geq 2$ and $n \geq 1$, then the polynomial $P_{n,k}(x)$ has no zeros on the interval $(-1, \infty)$.*

Proof. Clearly, the equation $P_{n,k}(x) = 0$ has no roots if $x \geq 0$ since it has positive coefficients. Suppose $-1 < x < 0$. If n is odd, then

$$a_{k+2j-1}x^{n-2j} + a_{k+2j}x^{n-2j-1} > 0, \qquad 0 \leq j \leq (n-1)/2,$$

since $x^{n-2j-1} > -x^{n-2j} > 0$ if $-1 < x < 0$ and $a_{k+2j} \geq a_{k+2j-1} > 0$. This implies

$$P_{n,k}(x) = \sum_{j=0}^{\frac{n-1}{2}} (a_{k+2j-1}x^{n-2j} + a_{k+2j}x^{n-2j-1}) > 0.$$

Similarly, if n is even, then

$$P_{n,k}(x) = a_{k-1}x^n + \sum_{j=0}^{\frac{n-2}{2}} (a_{k+2j}x^{n-2j-1} + a_{k+2j+1}x^{n-2j-2}) > 0.$$

□

So we seek the zeros of $P_{n,k}(x)$ where $x \leq -1$, equivalently, the zeros of $P_{n,k}(-x)$ where $x \geq 1$. For this, it is more convenient to consider the zeros of $g_{n,k}(x)$ given by

$$g_{n,k}(x) := c_k(-x)P_{n,k}(-x), \tag{2.1}$$

see [7], where

$$c_k(x) := x^k - x^{k-1} - x^{k-2} - \cdots - x - 1 \tag{2.2}$$

denotes the *characteristic polynomial* associated with the sequence a_n.

By [7, Lemma 2.1], we have

$$\begin{aligned} g_{n,k}(x) &= (-x)^{n+k} - a_{n+k}(-x)^{k-1} - (a_{n+1} + a_{n+2} + \cdots + a_{n+k-1})(-x)^{k-2} \\ &\quad - \cdots - (a_{n+k-2} + a_{n+k-1})(-x) - a_{n+k-1} \\ &= (-x)^{n+k} - a_{n+k}(-x)^{k-1} - \sum_{r=1}^{k-1} \left(\sum_{j=r}^{k-1} a_{n+j} \right) (-x)^{k-r-1}. \end{aligned} \tag{2.3}$$

We now wish to study the zeros of $g_{n,k}(x)$, where $x \geq 1$. In the subsequent two sections, we undertake such a study, considering separately the even and odd cases for k.

3. The case k even

Throughout this section, k will denote a positive even integer. We consider the zeros of the polynomial $g_{n,k}(x)$ where $x \geq 1$, and for this, it is more convenient to consider the zeros of the polynomial

$$f_{n,k}(x) := (1+x) g_{n,k}(x), \tag{3.1}$$

where $x \geq 1$.

First suppose n is odd. Note that when k is even and n is odd, we have

$$\begin{aligned} f_{n,k}(x) &= -x^{n+k}(1+x) + a_{n+k}x^k + a_n x^{k-1} - a_{n+1}x^{k-2} + a_{n+2}x^{k-3} \\ &\quad - \cdots + a_{n+k-2}x - a_{n+k-1} \\ &= -x^{n+k}(1+x) + a_{n+k}x^k + \sum_{r=0}^{k-1} (-1)^r a_{n+r} x^{k-r-1}, \end{aligned} \tag{3.2}$$

by (2.3) and the recurrence for a_n. In the lemmas below, we ascertain the number of the zeros of the polynomial $f_{n,k}(x)$ when $x \geq 1$. We will need the following combinatorial inequality.

Lemma 3.1. *If $k \geq 4$ is even and $n \geq 1$, then*

$$a_{n+k+1} \geq \sum_{r=0}^{\frac{k}{2}-1} 2^{\frac{k}{2}-r} a_{n+2r+1}. \tag{3.3}$$

Proof. We have

$$\begin{aligned} a_{n+k+1} &= a_{n+k} + \sum_{r=1}^{k-1} a_{n+r} \geq 2 \sum_{r=1}^{k-1} a_{n+r} \\ &= 2a_{n+k-1} + 2_{a+k-2} + 2 \sum_{r=1}^{k-3} a_{n+r} \geq 2a_{n+k-1} + 4 \sum_{r=1}^{k-3} a_{n+r} \end{aligned}$$

$$
\begin{aligned}
&= 2a_{n+k-1} + 4a_{n+k-3} + 4a_{n+k-4} + 4\sum_{r=1}^{k-5} a_{n+r} \\
&\geq 2a_{n+k-1} + 4a_{n+k-3} + 8\sum_{r=1}^{k-5} a_{n+r} \\
&= \cdots \geq \sum_{r=i}^{\frac{k}{2}-1} 2^{\frac{k}{2}-r} a_{n+2r+1} + 2^{\frac{k}{2}-i+1} \sum_{r=1}^{2i-1} a_{n+r} \\
&= \sum_{r=i-1}^{\frac{k}{2}-1} 2^{\frac{k}{2}-r} a_{n+2r+1} + 2^{\frac{k}{2}-i+1} a_{n+2i-2} + 2^{\frac{k}{2}-i+1} \sum_{r=1}^{2i-3} a_{n+r} \\
&\geq \sum_{r=i-1}^{\frac{k}{2}-1} 2^{\frac{k}{2}-r} a_{n+2r+1} + 2^{\frac{k}{2}-i+2} \sum_{r=1}^{2i-3} a_{n+r} \\
&= \cdots \geq \sum_{r=0}^{\frac{k}{2}-1} 2^{\frac{k}{2}-r} a_{n+2r+1},
\end{aligned}
$$

which gives (3.3). □

The following lemma will allow us to determine the number of zeros of $f_{n,k}(x)$ for $x \geq 1$.

Lemma 3.2. *Suppose $k \geq 4$ is even and n is odd. If $1 \leq i \leq k-1$, then $f_{n,k}^{(i)}(1) < 0$ implies $f_{n,k}^{(i+1)}(1) < 0$, where $f_{n,k}^{(i)}$ denotes the i-th derivative of $f_{n,k}$.*

Proof. Let $f = f_{n,k}$ and $i = k - j$ for some $1 \leq j \leq k-1$. Then the assumption $f^{(k-j)}(1) < 0$ is equivalent to

$$
\frac{k!}{j!} a_{n+k} + \sum_{r=0}^{j-1} (-1)^r \frac{(k-r-1)!}{(j-r-1)!} a_{n+r} < \prod_{s=1}^{k-j} (n+j+s) + \prod_{s=1}^{k-j} (n+j+s+1). \quad (3.4)
$$

We will show that inequality (3.4) implies

$$
\frac{k!}{(j-1)!} a_{n+k} + \sum_{r=0}^{j-2} (-1)^r \frac{(k-r-1)!}{(j-r-2)!} a_{n+r} < \prod_{s=0}^{k-j} (n+j+s) + \prod_{s=0}^{k-j} (n+j+s+1). \quad (3.5)
$$

Observe first that the left-hand side of both inequalities (3.4) and (3.5) is positive as

$$
\begin{aligned}
&\frac{k!}{j!} a_{n+k} + \sum_{r=0}^{j-1} (-1)^r \frac{(k-r-1)!}{(j-r-1)!} a_{n+r} \\
&\qquad = \frac{k!}{j!} \sum_{r=j}^{k-1} a_{n+r} + \sum_{r=0}^{j-1} \left(\frac{k!}{j!} + (-1)^r \frac{(k-r-1)!}{(j-r-1)!} \right) a_{n+r} > 0,
\end{aligned}
$$

since $a_{n+k} = \sum_{r=0}^{k-1} a_{n+r}$ and $\frac{k!}{j!} > \frac{(k-r-1)!}{(j-r-1)!}$. Note also that

$$\frac{\prod_{s=0}^{k-j}(n+j+s) + \prod_{s=0}^{k-j}(n+j+s+1)}{\prod_{s=1}^{k-j}(n+j+s) + \prod_{s=1}^{k-j}(n+j+s+1)} > n+j,$$

so to show (3.5), it suffices to show

$$\frac{k!}{(j-1)!}a_{n+k} + \sum_{r=0}^{j-2}(-1)^r \frac{(k-r-1)!}{(j-r-2)!}a_{n+r} \leq (n+j)\left(\frac{k!}{j!}a_{n+k} + \sum_{r=0}^{j-1}(-1)^r \frac{(k-r-1)!}{(j-r-1)!}a_{n+r}\right). \tag{3.6}$$

For (3.6), it is enough to show

$$\frac{k!}{(j-1)!}a_{n+k} + \sum_{r=0}^{j-2}(-1)^r \frac{(k-r-1)!}{(j-r-2)!}a_{n+r} \leq (j+1)\left(\frac{k!}{j!}a_{n+k} + \sum_{r=0}^{j-1}(-1)^r \frac{(k-r-1)!}{(j-r-1)!}a_{n+r}\right). \tag{3.7}$$

Starting with the left-hand side of (3.7), we have

$$\begin{aligned}
&\frac{k!}{(j-1)!}a_{n+k} + \sum_{r=0}^{j-2}(-1)^r \frac{(k-r-1)!}{(j-r-2)!}a_{n+r} \\
&= \frac{k!}{(j-1)!}\sum_{r=j-1}^{k-1} a_{n+r} + \sum_{r=0}^{j-2}\left(\frac{k!}{(j-1)!} + (-1)^r\frac{(k-r-1)!}{(j-r-2)!}\right)a_{n+r} \\
&= \frac{k!}{(j-1)!}\sum_{r=j}^{k-1} a_{n+r} + \sum_{r=0}^{j-1}\left(j\frac{k!}{j!} + (-1)^r(j-r-1)\frac{(k-r-1)!}{(j-r-1)!}\right)a_{n+r} \\
&= \frac{k!}{(j-1)!}\sum_{r=j}^{k-1} a_{n+r} + \sum_{r=0}^{j-1} j\left(\frac{k!}{j!} + (-1)^r\frac{(k-r-1)!}{(j-r-1)!}\right)a_{n+r} \\
&\quad + \sum_{r=0}^{j-1}(-1)^{r+1}(r+1)\frac{(k-r-1)!}{(j-r-1)!}a_{n+r} \\
&\leq \frac{k!}{(j-1)!}\sum_{r=j}^{k-1} a_{n+r} + \sum_{r=0}^{j-1} j\left(\frac{k!}{j!} + (-1)^r\frac{(k-r-1)!}{(j-r-1)!}\right)a_{n+r} \\
&\quad + \sum_{r=0}^{\lfloor\frac{j-2}{2}\rfloor}(2r+2)\frac{(k-2r-2)!}{(j-2r-2)!}a_{n+2r+1}
\end{aligned}$$

$$= (j+1)\frac{k!}{j!}\sum_{r=j}^{k-1} a_{n+r} + \sum_{r=0}^{j-1} j\left(\frac{k!}{j!} + (-1)^r \frac{(k-r-1)!}{(j-r-1)!}\right) a_{n+r}$$
$$- \frac{k!}{j!}\sum_{r=j}^{k-1} a_{n+r} + \sum_{r=0}^{\lfloor \frac{j-2}{2} \rfloor} (2r+2)\frac{(k-2r-2)!}{(j-2r-2)!} a_{n+2r+1}.$$

Below we show

$$\sum_{r=0}^{\lfloor \frac{j-2}{2} \rfloor} (2r+2)\frac{(k-2r-2)!}{(j-2r-2)!} a_{n+2r+1} \le \frac{k!}{j!}\sum_{r=j}^{k-1} a_{n+r}. \tag{3.8}$$

Then from (3.8), we have

$$\frac{k!}{(j-1)!}a_{n+k} + \sum_{r=0}^{j-2}(-1)^r \frac{(k-r-1)!}{(j-r-2)!} a_{n+r}$$
$$\le (j+1)\frac{k!}{j!}\sum_{r=j}^{k-1} a_{n+r} + \sum_{r=0}^{j-1} j\left(\frac{k!}{j!} + (-1)^r \frac{(k-r-1)!}{(j-r-1)!}\right) a_{n+r}$$
$$\le (j+1)\frac{k!}{j!}\sum_{r=j}^{k-1} a_{n+r} + \sum_{r=0}^{j-1}(j+1)\left(\frac{k!}{j!} + (-1)^r \frac{(k-r-1)!}{(j-r-1)!}\right) a_{n+r}$$
$$= (j+1)\frac{k!}{j!}a_{n+k} + (j+1)\sum_{r=0}^{j-1}(-1)^r \frac{(k-r-1)!}{(j-r-1)!} a_{n+r},$$

which gives (3.7), as desired.

To finish the proof, we need to show (3.8). We may assume $j \ge 2$, since the inequality is trivial when $j = 1$. By Lemma 3.1 and the fact that $2^m \ge 2m$ if $m \ge 1$, we have

$$\sum_{r=j}^{k-1} a_{n+r} \ge a_{n+k-1}$$
$$\ge \sum_{r=0}^{\frac{k}{2}-2} 2^{\frac{k}{2}-r-1} a_{n+2r+1} \ge \sum_{r=0}^{\frac{k}{2}-2} (k-2r-2) a_{n+2r+1} \ge \sum_{r=0}^{\lfloor \frac{j-2}{2} \rfloor} (k-2r-2) a_{n+2r+1},$$

the last inequality holding since $j \le k-1$, with k even. So to show (3.8), it is enough to show

$$(k-2r-2)\frac{k!}{j!} \ge (2r+2)\frac{(k-2r-2)!}{(j-2r-2)!}, \qquad 0 \le r \le \lfloor (j-2)/2 \rfloor, \tag{3.9}$$

where $2 \le j \le k-1$. Since the ratio $\frac{k!/j!}{(k-2r-2)!/(j-2r-2)!}$ is decreasing in j for fixed k and r, one needs to verify (3.9) only when $j = k-1$, and it holds in this case since $2r+2 \le j < k$. This completes the proof. □

We now determine the number of zeros of $f_{n,k}(x)$ on the interval $[1,\infty)$.

Lemma 3.3. *Suppose $k \geq 4$ is even and n is odd. Then the polynomial $f_{n,k}(x)$ has exactly one zero on the interval $[1,\infty)$. Furthermore, this zero is simple.*

Proof. Let $f = f_{n,k}$, where we first assume $n \geq 3$. Then

$$f(1) = -2 + a_{n+k} + \sum_{r=0}^{k-1}(-1)^r a_{n+r} = -2 + 2\sum_{r=0}^{\frac{k}{2}-1} a_{n+2r} > 0,$$

since $a_{n+k-2} \geq a_{k+1} = 2$. Let ℓ be the smallest positive integer i such that $f^{(i)}(1) < 0$; note that $1 \leq \ell \leq k+1$ since $f^{(k+1)}(1) < 0$. Then

$$f^{(\ell+1)}(1), f^{(\ell+2)}(1), \ldots, f^{(k+1)}(1)$$

are all negative, by Lemma 3.2. Since $f^{(k+1)}(x) < 0$ for all $x \geq 1$, it follows that $f^{(\ell)}(x) < 0$ for all $x \geq 1$. To see this, note that if $\ell \leq k$, then $f^{(k)}(1) < 0$ implies $f^{(k)}(x) < 0$ for all $x \geq 1$, which in turn implies each of $f^{(k)}(x), f^{(k-1)}(x), \ldots, f^{(\ell)}(x)$ is negative for all $x \geq 1$.

If $\ell \geq 2$, then $f^{(\ell-1)}(1) \geq 0$ and $f^{(\ell)}(x) < 0$ for all $x \geq 1$. Since $f^{(\ell-1)}(1) \geq 0$ and $\lim_{x\to\infty} f^{(\ell-1)}(x) = -\infty$, we have either (i) $f^{(\ell-1)}(1) = 0$ and $f^{(\ell-1)}(x)$ has no zeros on the interval $(1,\infty)$ or (ii) $f^{(\ell-1)}(1) > 0$ and $f^{(\ell-1)}(x)$ has exactly one zero on the interval $(1,\infty)$. If $\ell \geq 3$, then $f^{(\ell-2)}(x)$ would also have at most one zero on $(1,\infty)$ since $f^{(\ell-2)}(1) \geq 0$, with $f^{(\ell-2)}(x)$ initially increasing up to some point $s \geq 1$ before it decreases monotonically to $-\infty$ (where $s = 1$ if $f^{(\ell-1)}(1) = 0$ and $s > 1$ if $f^{(\ell-1)}(1) > 0$). Note that each derivative of f of order ℓ or less is eventually negative. Continuing in this fashion, we then see that if $\ell \geq 2$, then $f'(x)$ has at most one zero on the interval $(1,\infty)$, with $f'(1) \geq 0$ and $f'(x)$ eventually negative. If $\ell = 1$, then $f'(x) < 0$ for all $x \geq 1$. Since $f(1) > 0$ and $\lim_{x\to\infty} f(x) = -\infty$, it follows in either case that f has exactly one zero on the interval $[1,\infty)$, which finishes the case when $n \geq 3$.

If $n = 1$, then $f_{1,k}(x) = -x^{k+1}(1+x) + 2x^k + x - 1$ so that $f_{1,k}(1) = 0$, with

$$\begin{aligned} f'_{1,k}(x) &= -(k+1)x^k - (k+2)x^{k+1} + 2kx^{k-1} + 1 \\ &\leq -(k+1)x^{k-1} - (k+2)x^{k-1} + 2kx^{k-1} + 1 = -3x^{k-1} + 1 < 0 \end{aligned}$$

for $x \geq 1$. Thus, there is exactly one zero on the interval $[1,\infty)$ in this case as well.

Let t be the root of the equation $f_{n,k}(x) = 0$ on $[1,\infty)$. We now show that t has multiplicity one. First assume $n \geq 3$. Then $t > 1$. We consider cases depending on the value of $f'(1)$. If $f'(1) < 0$, then $f'(x) < 0$ for all $x \geq 1$ and thus $f'(t) < 0$ is non-zero, implying t is a simple root. If $f'(1) > 0$, then $f'(t) < 0$ due to $f(1) > 0$ and the fact that $f'(x)$ would then have one root v on $(1,\infty)$ with $v < t$. Finally, if $f'(1) = 0$, then the proof of Lemma 3.2 above shows that $f''(1) < 0$ and thus $f''(x) < 0$ for all $x \geq 1$, which implies $f'(t) < 0$. If $n = 1$, then $t = 1$ and $f'_{1,k}(1) < 0$. Thus, t is a simple root in all cases, as desired, which completes the proof. □

We next consider the case when n is even.

Lemma 3.4. *Suppose $k \geq 4$ and n are even. Then $f_{n,k}(x)$ has no zeros on $[1, \infty)$.*

Proof. In this case, we have

$$f_{n,k}(x) = x^{n+k}(1+x) + a_{n+k}x^k + \sum_{r=0}^{k-1}(-1)^r a_{n+r}x^{k-r-1},$$

by (2.3) and (3.1). If $x \geq 1$, then $f_{n,k}(x) > 0$ since $a_{n+k} = \sum_{r=0}^{k-1} a_{n+r}$ and $x^k \geq x^{k-r-1}$ for $0 \leq r \leq k-1$. □

The main result of this section now follows rather quickly.

Theorem 3.5. *(i) If k is even and n is odd, then the polynomial $P_{n,k}(x)$ has one real zero q, and it is simple with $q \leq -1$.*
(ii) If k and n are even, then the polynomial $P_{n,k}(x)$ has no real zeros.

Proof. Note first that the preceding lemmas, where we assumed $k \geq 4$ is even, may be adjusted slightly and are also seen to hold in the case $k = 2$. First suppose n is odd. By Lemma 3.3, the polynomial $f_{n,k}(x)$, and hence $g_{n,k}(x)$, has one zero for $x \geq 1$, and it is simple. By [7, Lemma 2.3], the characteristic polynomial $c_k(x) = x^k - x^{k-1} - x^{k-2} - \cdots - 1$ has one negative real zero when k is even, and it is seen to lie in the interval $(-1, 0)$. Since $g_{n,k}(x) = c_k(-x)P_{n,k}(-x)$, it follows that $P_{n,k}(-x)$ has one zero for $x \geq 1$. Thus, $P_{n,k}(x)$ has one zero for $x \leq -1$, and it is simple. By Lemma 2.1, the polynomial $P_{n,k}(x)$ has exactly one real zero.

If n is even, then the polynomial $f_{n,k}(x)$, and hence $g_{n,k}(x)$, has no zeros for $x \geq 1$, by Lemma 3.4. By (2.1), it follows that $P_{n,k}(x)$ has no zeros for $x \leq -1$. By Lemma 2.1, $P_{n,k}(x)$ has no real zeros. □

4. The case k odd

Throughout this section, $k \geq 3$ will denote a positive odd integer. We study the zeros of the polynomial $g_{n,k}(x)$ when $x \geq 1$, and for this, it is again more convenient to consider the polynomial $f_{n,k}(x) := (1+x)g_{n,k}(x)$. First suppose n is odd. When k and n are both odd, note that

$$\begin{aligned} f_{n,k}(x) &= x^{n+k}(1+x) - a_{n+k}x^k - a_n x^{k-1} + a_{n+1}x^{k-2} - \cdots + a_{n+k-2}x - a_{n+k-1} \\ &= x^{n+k}(1+x) - a_{n+k}x^k + \sum_{r=0}^{k-1}(-1)^{r+1}a_{n+r}x^{k-r-1}, \end{aligned}$$

by (2.3) and the recurrence for a_n. In the lemmas below, we ascertain the number of zeros of the polynomial $f_{n,k}(x)$ when $x \geq 1$. We start with the following inequality.

Lemma 4.1. *Suppose $k \geq 3$ is odd and $n \geq 1$. If $1 \leq j \leq k-1$, then*

$$3\frac{k!}{j!}a_{n+k-1} \geq \sum_{r=1}^{\lfloor \frac{j}{2} \rfloor} 2r\frac{(k-2r)!}{(j-2r)!}a_{n+2r-1}. \tag{4.1}$$

Proof. First note that we have the inequality

$$a_{n+k-1} \geq \sum_{r=1}^{\frac{k-3}{2}} 2^{\frac{k-1}{2}-r}a_{n+2r}. \tag{4.2}$$

To show (4.2), proceed as in the proof of Lemma 3.1 above and write

$$\begin{aligned} a_{n+k-1} &\geq a_{n+k-2} + \sum_{r=2}^{k-3} a_{n+r} \\ &\geq 2a_{n+k-3} + 2\sum_{r=2}^{k-4} a_{n+r} \\ &= 2a_{n+k-3} + 2a_{n+k-4} + 2\sum_{r=2}^{k-5} a_{n+r} \\ &\geq 2a_{2n+k-3} + 4a_{n+k-5} + 4\sum_{r=2}^{k-6} a_{n+r} \\ &= \cdots \geq \sum_{r=1}^{\frac{k-3}{2}} 2^{\frac{k-1}{2}-r}a_{n+2r}. \end{aligned}$$

Since $2^m \geq 2m$ if $m \geq 1$, we have

$$a_{n+k-1} \geq \sum_{r=1}^{\frac{k-3}{2}} 2^{\frac{k-1}{2}-r}a_{n+2r} \geq \sum_{r=1}^{\frac{k-3}{2}} (k-2r-1)a_{n+2r}. \tag{4.3}$$

First suppose $j \leq k-2$. In this case, we show

$$\frac{k!}{j!}a_{n+k-1} \geq \sum_{r=1}^{\lfloor \frac{j}{2} \rfloor} r\frac{(k-2r)!}{(j-2r)!}a_{n+2r-1}, \tag{4.4}$$

which implies (4.1). And (4.4) is seen to hold since by (4.3),

$$\frac{k!}{j!}a_{n+k-1} \geq \sum_{r=1}^{\frac{k-3}{2}} \frac{(k-2r-1)k!}{j!}a_{n+2r} \geq \sum_{r=1}^{\lfloor \frac{j}{2} \rfloor} \frac{(k-2r-1)k!}{j!}a_{n+2r},$$

with $a_{n+2r} \geq a_{n+2r-1}$ and

$$\frac{(k-2r-1)k!}{r(k-2r)!} \geq \frac{(k-2)!}{(k-2r-2)!} \geq \frac{j!}{(j-2r)!}.$$

The $j = k-1$ case of (4.1) follows from noting

$$\begin{aligned} 3ka_{n+k-1} &\geq ka_{n+k-1} + \sum_{r=1}^{\frac{k-3}{2}} 2k(k-2r-1)a_{n+2r} \\ &\geq (k-1)a_{n+k-2} + \sum_{r=1}^{\frac{k-3}{2}} 2r(k-2r)a_{n+2r-1} = \sum_{r=1}^{\frac{k-1}{2}} 2r(k-2r)a_{n+2r-1}, \end{aligned}$$

since $k(k-2r-1) \geq r(k-2r)$ if $1 \leq r \leq \frac{k-3}{2}$. □

Lemma 4.2. *Suppose $k, n \geq 3$ are odd. If $1 \leq i \leq k-1$, then $f_{n,k}^{(i)}(1) > 0$ implies $f_{n,k}^{(i+1)}(1) > 0$.*

Proof. Let $f = f_{n,k}$ and $i = k-j$ for some $1 \leq j \leq k-1$. Then the assumption $f^{(k-j)}(1) > 0$ is equivalent to

$$\frac{(n+k)!}{(n+j)!} + \frac{(n+k+1)!}{(n+j+1)!} > \frac{k!}{j!}a_{n+k} + \sum_{r=0}^{j-1}(-1)^r \frac{(k-r-1)!}{(j-r-1)!}a_{n+r}. \tag{4.5}$$

Using (4.5), we will show $f^{(k-j+1)}(1) > 0$, i.e.,

$$\frac{(n+k)!}{(n+j-1)!} + \frac{(n+k+1)!}{(n+j)!} > \frac{k!}{(j-1)!}a_{n+k} + \sum_{r=0}^{j-2}(-1)^r \frac{(k-r-1)!}{(j-r-2)!}a_{n+r}. \tag{4.6}$$

Note that the right-hand side of both inequalities (4.5) and (4.6) is positive since $a_{n+k} = \sum_{r=0}^{k-1} a_{n+r}$. Since the left-hand side of (4.6) divided by the left-hand side of (4.5) is greater than $n+j$, it suffices to show

$$\begin{aligned} &\frac{k!}{(j-1)!}a_{n+k} + \sum_{r=0}^{j-2}(-1)^r \frac{(k-r-1)!}{(j-r-2)!}a_{n+r} \\ &\qquad \leq (n+j)\left(\frac{k!}{j!}a_{n+k} + \sum_{r=0}^{j-1}(-1)^r \frac{(k-r-1)!}{(j-r-1)!}a_{n+r}\right). \end{aligned} \tag{4.7}$$

For (4.7), it is enough to show

$$\begin{aligned} &\frac{k!}{(j-1)!}a_{n+k} + \sum_{r=0}^{j-2}(-1)^r \frac{(k-r-1)!}{(j-r-2)!}a_{n+r} \\ &\qquad \leq (j+3)\left(\frac{k!}{j!}a_{n+k} + \sum_{r=0}^{j-1}(-1)^r \frac{(k-r-1)!}{(j-r-1)!}a_{n+r}\right), \end{aligned} \tag{4.8}$$

since $n \geq 3$.

Starting with the left-hand-side of (4.8), and proceeding at this stage as in the proof of Lemma 3.2 above, we have

$$\begin{aligned}
&\frac{k!}{(j-1)!}a_{n+k}+\sum_{r=0}^{j-2}(-1)^r\frac{(k-r-1)!}{(j-r-2)!}a_{n+r}\\
&\quad\le\frac{k!}{(j-1)!}\sum_{r=j}^{k-1}a_{n+r}+\sum_{r=0}^{j-1}j\left(\frac{k!}{j!}+(-1)^r\frac{(k-r-1)!}{(j-r-1)!}\right)a_{n+r}\\
&\qquad+\sum_{r=1}^{\lfloor\frac{j}{2}\rfloor}2r\frac{(k-2r)!}{(j-2r)!}a_{n+2r-1}\\
&\quad=(j+3)\frac{k!}{j!}\sum_{r=j}^{k-1}a_{n+r}+\sum_{r=0}^{j-1}j\left(\frac{k!}{j!}+(-1)^r\frac{(k-r-1)!}{(j-r-1)!}\right)a_{n+r}\\
&\qquad-3\frac{k!}{j!}\sum_{r=j}^{k-1}a_{n+r}+\sum_{r=1}^{\lfloor\frac{j}{2}\rfloor}2r\frac{(k-2r)!}{(j-2r)!}a_{n+2r-1}\\
&\quad\le(j+3)\frac{k!}{j!}\sum_{r=j}^{k-1}a_{n+r}+\sum_{r=0}^{j-1}j\left(\frac{k!}{j!}+(-1)^r\frac{(k-r-1)!}{(j-r-1)!}\right)a_{n+r},
\end{aligned}$$

where the last inequality follows from Lemma 4.1. Thus,

$$\begin{aligned}
&\frac{k!}{(j-1)!}a_{n+k}+\sum_{r=0}^{j-2}(-1)^r\frac{(k-r-1)!}{(j-r-2)!}a_{n+r}\\
&\quad\le(j+3)\frac{k!}{j!}\sum_{r=j}^{k-1}a_{n+r}+\sum_{r=0}^{j-1}j\left(\frac{k!}{j!}+(-1)^r\frac{(k-r-1)!}{(j-r-1)!}\right)a_{n+r}\\
&\quad\le(j+3)\frac{k!}{j!}\sum_{r=j}^{k-1}a_{n+r}+\sum_{r=0}^{j-1}(j+3)\left(\frac{k!}{j!}+(-1)^r\frac{(k-r-1)!}{(j-r-1)!}\right)a_{n+r}\\
&\quad=(j+3)\frac{k!}{j!}a_{n+k}+(j+3)\sum_{r=0}^{j-1}(-1)^r\frac{(k-r-1)!}{(j-r-1)!}a_{n+r},
\end{aligned}$$

which gives (4.8) and completes the proof. □

We can now determine the number of zeros of $f_{n,k}(x)$ on the interval $[1,\infty)$.

Lemma 4.3. *Suppose $k\ge 3$ and n are odd. Then $f_{n,k}(x)$ has exactly one zero on the interval $[1,\infty)$ and it is simple.*

Proof. If $n\ge 3$, then use Lemma 4.2 and the same reasoning as in the proof of Lemma 3.3 above. Note that in this case we have

$$f_{n,k}(1)=2-a_{n+k}+\sum_{r=0}^{k-1}(-1)^{r+1}a_{n+r}=2-2\sum_{r=0}^{\frac{k-1}{2}}a_{n+2r}<0,$$

as $a_{n+k-1}, a_{n+k-3} > 0$. If $n = 1$, then $f_{1,k}(x) = x^{k+1}(1+x) - 2x^k + x - 1$ and the result also holds as $f_{1,k}(1) = 0$ with $f'_{1,k}(x) > 0$ if $x \geq 1$. □

We next consider the case when n is even.

Lemma 4.4. *If $k \geq 3$ is odd and n is even, then $f_{n,k}(x)$ has no zeros on $[1, \infty)$.*

Proof. In this case, we have

$$f_{n,k} = -x^{n+k}(1+x) - a_{n+k}x^k + \sum_{r=0}^{k-1}(-1)^{r+1}a_{n+r}x^{k-r-1}.$$

If $x \geq 1$, then $f_{n,k}(x) < 0$ since $a_{n+k} = \sum_{r=0}^{k-1} a_{n+r}$ and $-x^k \leq -x^{k-r-1}$ for $0 \leq r \leq k-1$. □

We now prove the main result of this section.

Theorem 4.5. *(i) If $k \geq 3$ and n are odd, then the polynomial $P_{n,k}(x)$ has one real zero q, and it is simple with $q \leq -1$.*
(ii) If $k \geq 3$ is odd and n is even, then the polynomial $P_{n,k}(x)$ has no real zeros.

Proof. First suppose n is odd. By Lemma 4.3, the polynomial $f_{n,k}(x)$, and hence $g_{n,k}(x)$, has one zero on $[1, \infty)$, and it is simple. By [7, Lemma 2.3], the characteristic polynomial $c_k(x) = x^k - x^{k-1} - x^{k-2} - \cdots - 1$ has no negative real zeros when k is odd. Since $g_{n,k}(x) = c_k(-x)P_{n,k}(-x)$, it follows that $P_{n,k}(x)$ has one zero for $x \leq -1$, and hence one real zero, by Lemma 2.1.

If n is even, then the polynomial $f_{n,k}(x)$, and hence $g_{n,k}(x)$, has no zeros for $x \geq 1$, by Lemma 4.4. Thus, neither does $P_{n,k}(-x)$, which implies it has no real zeros. □

5. Convergence of zeros

In this section, we show that for each fixed $k \geq 2$, the sequence of real zeros of $P_{n,k}(x)$ for n odd is convergent. Before proving this, we remind the reader of the following version of Rouché's Theorem which can be found in [4].

Theorem 5.1 (Rouché)**.** *If $p(z)$ and $q(z)$ are analytic interior to a simple closed Jordan curve $\mathcal{C}$, and are continuous on $\mathcal{C}$, with*

$$|p(z) - q(z)| < |q(z)|, \qquad z \in \mathcal{C},$$

then the functions $p(z)$ and $q(z)$ have the same number of zeros interior to $\mathcal{C}$.

We now give three preliminary lemmas.

Lemma 5.2. *(i) If $k \geq 2$, then the polynomial $c_k(x) = x^k - x^{k-1} - \cdots - x - 1$ has one positive real zero λ, with $\lambda > 1$. All of its other zeros have modulus strictly less than one.*
(ii) The zeros of $c_k(x)$, which we will denote by $\alpha_1 = \lambda, \alpha_2, \ldots, \alpha_k$, are distinct and thus

$$a_n = c_1\alpha_1^n + c_2\alpha_2^n + \cdots + c_n\alpha_k^n, \qquad n \geq 0, \tag{5.1}$$

where $c_1, c_2, \ldots, c_k$ are constants.
(iii) The constant c_1 is a positive real number.

Proof. (i) It is more convenient to consider the polynomial $d_k(x) := (1-x)c_k(x)$. Note that

$$d_k(x) = (1-x)\left(x^k - \frac{1-x^k}{1-x}\right) = 2x^k - x^{k+1} - 1.$$

We regard $d_k(z)$ as a complex function. Since on the circle $|z| = 3$ in the complex plane holds

$$|2z^k| = 2 \cdot 3^k < 3^{k+1} - 1 = |-z^{k+1}| - 1 \leq |-z^{k+1} - 1|,$$

it follows from Rouché's Theorem that $d_k(z)$ has $k+1$ zeros in the disc $|z| < 3$ since the function $-z^{k+1} - 1$ has all of its zeros there. On the other hand, on the circle $|z| = 1 + \epsilon$, we have

$$|-z^{k+1}| = (1+\epsilon)^{k+1} < 2(1+\epsilon)^k - 1 \leq |2z^k - 1|,$$

which implies that the polynomial $d_k(z)$ has exactly k zeros in the disc $|z| < 1+\epsilon$, for all $\epsilon > 0$ sufficiently small such that $-\frac{\ln(1-\epsilon)}{\ln(1+\epsilon)} < 2 \leq k$. Letting $\epsilon \to 0$, we see that there are k zeros for the polynomial $d_k(z)$ in the disc $|z| \leq 1$. But $z = 1$ is a zero of the polynomial $d_k(z) = (1-z)c_k(z)$ on the circle $|z| = 1$, and it is the only such zero since $d_k(z) = 0$ implies $|z|^k \cdot |2 - z| = 1$, or $|2 - z| = 1$, which is clearly satisfied by only $z = 1$. Hence, the polynomial $c_k(z)$ has $k-1$ zeros in the disc $|z| < 1$ and exactly one zero in the domain $1 < |z| < 3$. Finally, by Descartes' rule of signs and since $c_k(1) < 0$, we see that $c_k(x)$ has exactly one positive real zero λ, with $1 < \lambda < 3$.

(ii) We'll prove only the first statement, as the second one follows from the first and the theory of linear recurrences. For this, first note that $d_k'(x) = 0$ implies $x = 0, \frac{2k}{k+1}$. Now the only possible rational roots of the equation $d_k(x) = 0$ are ± 1, by the rational root theorem. Thus $d_k\left(\frac{2k}{k+1}\right) = 0$ is impossible as $k \geq 2$, which implies $d_k(x)$ and $d_k'(x)$ cannot share a zero. Therefore, the zeros of $d_k(x)$, and hence of $c_k(x)$, are distinct.

(iii) Substitute $n = 0, 1, \ldots, k-1$ into (5.1), and recall that $a_0 = a_1 = \cdots = a_{k-2} = 0$ with $a_{k-1} = 1$, to obtain a system of linear equations in the variables $c_1, c_2, \ldots, c_k$. Let A be the coefficient matrix for this system (where the equations are understood to have been written in the natural order) and let A' be the matrix obtained from A by replacing the first column of A with the vector $(0, \ldots, 0, 1)$ of

length k. Now the transpose of A and of the $(k-1)\times(k-1)$ matrix obtained from A' by deleting the first column and the last row are seen to be Vandermonde matrices. Therefore, by Cramer's rule, we have

$$c_1 = \frac{\det A'}{\det A} = \frac{(-1)^{k+1}\prod_{2\le i<j\le k}(\alpha_j-\alpha_i)}{\prod_{1\le i<j\le k}(\alpha_j-\alpha_i)}$$
$$= \frac{1}{(-1)^{k-1}\prod_{j=2}^{k}(\alpha_j-\alpha_1)} = \frac{1}{\prod_{j=2}^{k}(\alpha_1-\alpha_j)}.$$

If $j\ge 2$, then either $\alpha_j<0$ or α_j and α_ℓ are complex conjugates for some ℓ. Note that $\alpha_1-\alpha_j>0$ in the first case and

$$(\alpha_1-\alpha_j)(\alpha_1-\alpha_\ell) = (\alpha_1-a)^2+b^2>0$$

in the second, where $\alpha_j = a+bi$. Since all of the complex zeros of $c_k(x)$ which aren't real come in conjugate pairs, it follows that c_1 is a positive real number. □

We give the zeros of $c_k(z)$ for $2\le k\le 5$ as well as the value of the constant c_1 in Table 1 below, where $\overline{z}$ denotes the complex conjugate of z.

k	The zeros of $c_k(z)$	The constant c_1
2	1.61803, −0.61803	0.44721
3	1.83928, $r_1 = -0.41964+0.60629i$, $\overline{r_1}$	0.18280
4	1.92756, −0.77480, $r_1 = -0.07637+0.81470i$, $\overline{r_1}$	0.07907
5	1.96594, $r_1 = 0.19537+0.84885i$, $r_2 = -0.67835+0.45853i$, $\overline{r_1}$, $\overline{r_2}$	0.03601

Table 1: The zeros of $c_k(z)$ and the constant c_1.

The next lemma concerns the location of the positive zero of the k-th derivative of $f_{n,k}(x)$.

Lemma 5.3. *Suppose $k\ge 2$ is fixed and n is odd. Let s_n $(=s_{n,k})$ be the zero of $f_{n,k}(x)$ on $[1,\infty)$, where $f_{n,k}(x)$ is given by (3.1), and let t_n $(=t_{n,k})$ be the positive zero of the k-th derivative of $f_{n,k}(x)$. Let λ be the positive zero of $c_k(x)$. Then we have*

(i) $t_n<s_n$ for all odd n, and

(ii) $t_n\to\lambda$ as n odd increases without bound.

Proof. Suppose k is even, the proof when k is odd being similar. Then $f_{n,k}$ is given by (3.2) above. Throughout the following proof, n will always represent an odd integer and $f=f_{n,k}$. Recall from Lemma 3.3 that f has exactly one zero on the interval $[1,\infty)$.

(i) By Descartes' rule of signs, the polynomial $f^{(k)}(x)$ has one positive real zero t_n. If $t_n<1\le s_n$, then we are done, so let us assume $t_n\ge 1$. The condition $t_n\ge 1$,

or equivalently $f^{(k)}(1) \geq 0$, then implies $n \geq 3$, and thus $f(1) > 0$. (Indeed, $t_n \geq 1$ for all n sufficiently large since $a_{n+k} \sim c_1\lambda^{n+k}$, with $\lambda > 1$.)

Now observe that $f^{(k)}(1) \geq 0$ implies $f^{(i)}(1) > 0$ for $1 \leq i \leq k-1$, as the proof of Lemma 3.2 above shows in fact that $f^{(i)}(1) \leq 0$ implies $f^{(i+1)}(1) < 0$. Since $f^{(i)}(1) > 0$ for $0 \leq i \leq k-1$ and $f^{(k)}(1) \geq 0$, it follows that each of the polynomials $f(x), f'(x), \ldots, f^{(k)}(x)$ has exactly one zero on $[1, \infty)$ since $f^{(k+1)}(x) < 0$ for all $x \geq 1$. Furthermore, the zero of $f^{(i)}(x)$ on $[1, \infty)$ is strictly larger than the zero of $f^{(i+1)}(x)$ on $[1, \infty)$ for $0 \leq i \leq k-1$. In particular, the zero of $f(x)$ is strictly larger than the zero of $f^{(k)}(x)$, which establishes the first statement.

(ii) Let us assume n is large enough to ensure $t_n \geq 1$. Note that

$$\frac{f^{(k)}(x)}{k!} = -\binom{n+k}{k}x^n - \binom{n+k+1}{k}x^{n+1} + a_{n,k}$$

so that

$$-2\binom{n+k+1}{k}x^{n+1} + a_{n,k} \leq \frac{f^{(k)}(x)}{k!} \leq -2\binom{n+k}{k}x^n + a_{n,k}, \qquad x \geq 1. \quad (5.2)$$

Setting $x = t_n$ in (5.2), and rearranging, then gives

$$\left(\frac{a_{n+k}}{2\binom{n+k+1}{k}}\right)^{1/(n+1)} \leq t_n \leq \left(\frac{a_{n+k}}{2\binom{n+k}{k}}\right)^{1/n}. \quad (5.3)$$

The second statement then follows from letting n tend to infinity in (5.3) and noting $\lim_{n\to\infty}(a_{n+k})^{1/n} = \lambda$ (as $a_{n+k} \sim c_1\lambda^{n+k}$, by Lemma 5.2). □

We will also need the following formula for an expression involving the zeros of $c_k(x)$.

Lemma 5.4. *If $\alpha_1 = \lambda, \alpha_2, \ldots, \alpha_k$ are the zeros of $c_k(x)$, then*

$$\sum_{j=0}^{k-1}(-1)^j\lambda^{k-j-1}\mathcal{S}_j\{\alpha_2, \alpha_3, \ldots, \alpha_k\} = \frac{k\lambda^{k+2} - (2k-1)\lambda^{k+1} - (k-1)\lambda^k + 2k\lambda^{k-1} - \lambda - 1}{(\lambda-1)^2(\lambda+1)}, \quad (5.4)$$

where $\mathcal{S}_j\{\alpha_2, \alpha_3, \ldots, \alpha_k\}$ denotes the j-th symmetric function in the quantities $\alpha_2, \alpha_3, \ldots, \alpha_k$ if $1 \leq j \leq k-1$, with $\mathcal{S}_0\{\alpha_2, \alpha_3, \ldots, \alpha_k\} := 1$.

Proof. Let us assume k is even, the proof in the odd case being similar. First note that

$$(-1)^{i+1} = \mathcal{S}_i\{\alpha_1, \alpha_2, \ldots, \alpha_k\} = \mathcal{S}_i\{\alpha_2, \ldots, \alpha_k\} + \lambda\mathcal{S}_{i-1}\{\alpha_2, \ldots, \alpha_k\}, \quad 1 \leq i \leq k,$$

which gives the recurrences

$$\mathcal{S}_{2r}\{\alpha_2, \ldots, \alpha_k\} = -1 - \lambda\mathcal{S}_{2r-1}\{\alpha_2, \ldots, \alpha_k\}, \qquad 1 \leq r \leq (k-2)/2, \quad (5.5)$$

and

$$\mathcal{S}_{2r+1}\{\alpha_2,\ldots,\alpha_k\} = 1 - \lambda\mathcal{S}_{2r}\{\alpha_2,\ldots,\alpha_k\}, \qquad 0 \le r \le (k-2)/2. \tag{5.6}$$

Iterating (5.5) and (5.6) yields

$$\begin{aligned}\mathcal{S}_{2r}\{\alpha_2,\ldots,\alpha_k\} &= -(1+\lambda+\cdots+\lambda^{2r-1}) + \lambda^{2r} \\ &= -\frac{1-2\lambda^{2r}+\lambda^{2r+1}}{1-\lambda}, \quad 1 \le r \le (k-2)/2, \end{aligned}\tag{5.7}$$

and

$$\begin{aligned}\mathcal{S}_{2r+1}\{\alpha_2,\ldots,\alpha_k\} &= (1+\lambda+\cdots+\lambda^{2r}) - \lambda^{2r+1} \\ &= \frac{1-2\lambda^{2r+1}+\lambda^{2r+2}}{1-\lambda}, \quad 0 \le r \le (k-2)/2. \end{aligned}\tag{5.8}$$

Note that (5.7) also holds in the case when $r = 0$.

By (5.7) and (5.8), we then have

$$\begin{aligned}&\sum_{j=0}^{k-1}(-1)^j\lambda^{k-j-1}\mathcal{S}_j\{\alpha_2,\alpha_3,\ldots,\alpha_k\} \\ &\quad= -\sum_{r=0}^{\frac{k}{2}-1}\lambda^{k-2r-1}\left(\frac{1-2\lambda^{2r}+\lambda^{2r+1}}{1-\lambda}\right) - \sum_{r=0}^{\frac{k}{2}-1}\lambda^{k-2r-2}\left(\frac{1-2\lambda^{2r+1}+\lambda^{2r+2}}{1-\lambda}\right) \\ &\quad= \frac{1}{\lambda-1}\sum_{r=0}^{\frac{k}{2}-1}(\lambda^{k-2r-1}-2\lambda^{k-1}+\lambda^k) + \frac{1}{\lambda-1}\sum_{r=0}^{\frac{k}{2}-1}(\lambda^{k-2r-2}-2\lambda^{k-1}+\lambda^k) \\ &\quad= \frac{\lambda}{\lambda-1}\left(\frac{\lambda^k-1}{\lambda^2-1}\right) + \frac{1}{\lambda-1}\left(\frac{\lambda^k-1}{\lambda^2-1}\right) - \frac{2k\lambda^{k-1}}{\lambda-1} + \frac{k\lambda^k}{\lambda-1},\end{aligned}$$

which gives (5.4). □

We now can prove the main result of this section.

Theorem 5.5. *Suppose $k \ge 2$ and n is odd. Let r_n $(= r_{n,k})$ denote the real zero of the polynomial $P_{n,k}(x)$ defined by* (1.2) *above. Then $r_n \to -\lambda$ as $n \to \infty$.*

Proof. Let n denote an odd integer throughout. We first consider the case when k is even. Equivalently, we show that $s_n \to \lambda$ as $n \to \infty$, where s_n denotes the zero of $f_{n,k}(x)$ on the interval $[1,\infty)$. By Lemma 5.3, we have $t_n < s_n$ for all n with $t_n \to \lambda$ as $n \to \infty$, where t_n is the positive zero of the k-th derivative of $f_{n,k}(x)$. So it is enough to show $s_n < \lambda$ for all n sufficiently large, i.e., $f_{n,k}(\lambda) < 0$.

By Lemma 5.2, we have

$$f_{n,k}(\lambda) = -\lambda^{n+k}(1+\lambda) + a_{n,k}\lambda^k + \sum_{r=0}^{k-1}(-1)^r a_{n+r}\lambda^{k-r-1}$$

$$\sim -\lambda^{n+k}(1+\lambda) + c_1\lambda^{n+2k} + \sum_{r=0}^{k-1}(-1)^r c_1\lambda^{n+k-1}$$
$$= \lambda^{n+k}(-1-\lambda+c_1\lambda^k),$$

so that $f_{n,k}(\lambda) < 0$ for large n if $-1-\lambda+c_1\lambda^k < 0$, i.e.,

$$\lambda^k < \frac{1+\lambda}{c_1}. \tag{5.9}$$

So to complete the proof, we must show (5.9). By Lemmas 5.2 and 5.4, we have

$$\frac{1}{c_1} = \prod_{j=2}^{k}(\lambda-\alpha_j) = \sum_{j=0}^{k-1}(-1)^j\lambda^{k-j-1}\mathcal{S}_j\{\alpha_2,\alpha_3,\ldots,\alpha_k\}$$
$$= \frac{k\lambda^{k+2}-(2k-1)\lambda^{k+1}-(k-1)\lambda^k+2k\lambda^{k-1}-\lambda-1}{(\lambda-1)^2(\lambda+1)},$$

so that (5.9) holds if and only

$$\lambda^k(\lambda-1)^2 < k\lambda^{k+2}-(2k-1)\lambda^{k+1}-(k-1)\lambda^k+2k\lambda^{k-1}-\lambda-1,$$

i.e.,

$$1+\lambda+k\lambda^k+(2k-3)\lambda^{k+1} < 2k\lambda^{k-1}+(k-1)\lambda^{k+2}. \tag{5.10}$$

Recall from the proof of Lemma 5.2 that $2\lambda^k = 1+\lambda^{k+1}$. Substituting $\lambda^{k+1} = \frac{\lambda+\lambda^{k+2}}{2}$,

$$\lambda^k = \frac{1+\frac{\lambda+\lambda^{k+2}}{2}}{2} = \frac{2+\lambda+\lambda^{k+2}}{4},$$

and

$$\lambda^{k-1} = \frac{\lambda^k}{\lambda} = \frac{2+\lambda+\lambda^{k+2}}{4\lambda}$$

into (5.10), and rearranging, then gives

$$\left(1-\frac{\lambda}{2}-\frac{k}{\lambda}\right)+\frac{5k\lambda}{4} < \lambda^{k+2}\left(\frac{k}{2\lambda}-\frac{k}{4}+\frac{1}{2}\right). \tag{5.11}$$

For (5.11), note first that $c_k(2) > 0$ as $2^k > 2^k-1 = 2^{k-1}+\cdots+1$, which implies $\lambda < 2 \le k$ and thus $1-\frac{\lambda}{2}-\frac{k}{\lambda} < 0$. So to show (5.11), it is enough to show

$$\frac{5k}{4} < \lambda^{k+1}\left(\frac{k}{2\lambda}-\frac{k}{4}+\frac{1}{2}\right). \tag{5.12}$$

For (5.12), we'll consider the cases $k=2$ and $k \ge 4$. If $k=2$, then $\lambda = \theta = \frac{1+\sqrt{5}}{2}$, so that (5.12) reduces in this case to $\frac{5}{2} < \theta^2 = \theta+1$, which is true. Now suppose

$k \geq 4$ is even. First observe that $c_k\left(\frac{5}{3}\right) < 0$, whence $\lambda > \frac{5}{3}$, as $d_k\left(\frac{5}{3}\right) > 0$ since $\left(\frac{5}{3}\right)^k \left(2 - \frac{5}{3}\right) > 1$ for *all* $k \geq 3$. Thus, we have

$$\begin{aligned}\lambda^k &= \left(\lambda^{k-1} + 1\right) + \lambda^{k-2} + \lambda^{k-3} + \cdots + \lambda \\ &> 2\lambda^{\frac{k-1}{2}} + \lambda^{k-2} + \lambda^{k-3} + \cdots + \lambda > 2 \cdot \frac{5}{3} + \frac{5(k-2)}{3} = \frac{5k}{3}.\end{aligned}$$

So to show (5.12) when $k \geq 4$, it suffices to show

$$0 < \lambda\left(\frac{k}{2\lambda} - \frac{k}{4} + \frac{1}{2}\right) - \frac{3}{4} = \frac{k(2-\lambda)}{4} + \frac{2\lambda - 3}{4},$$

which is true as $\frac{5}{3} < \lambda < 2$. This completes the proof in the even case.

If k is odd, then we proceed in a similar manner. Instead of inequality (5.9), we get

$$\lambda^k + \frac{1}{\lambda} < \frac{1+\lambda}{c_1}, \tag{5.13}$$

which is equivalent to

$$\left(1 - \frac{\lambda}{2} - \frac{k}{\lambda} + \frac{(\lambda-1)^2}{\lambda}\right) + \frac{5k\lambda}{4} < \lambda^{k+2}\left(\frac{k}{2\lambda} - \frac{k}{4} + \frac{1}{2}\right). \tag{5.14}$$

Note that the sum of the first four terms on the left-hand side of (5.14) is negative since $1 - \frac{k}{\lambda} < 0$ and $-\frac{\lambda}{2} + \frac{(\lambda-1)^2}{\lambda} < 0$ as $\frac{5}{3} < \lambda < 2$ for $k \geq 3$. Thus, it suffices to show (5.12) in the case when $k \geq 3$ is odd, which has already been done since the proof given above for it applies to *all* $k \geq 3$. □

$n\backslash k$	2	3	4	5
1	1	1	1	1
5	1.39118	1.59674	1.61156	1.64627
9	1.48442	1.69002	1.73834	1.77122
49	1.59187	1.80885	1.88958	1.92625
99	1.60498	1.82403	1.90856	1.94605
199	1.61151	1.83165	1.91805	1.95599
λ	1.61803	1.83928	1.92756	1.96594

Table 2: Some real zeros of $P_{n,k}(-x)$, where λ is the positive zero of $c_k(x)$.

Perhaps the proofs presented here of Theorems 1.2 and 5.5 could be generalized to show comparable results for polynomials associated with linear recurrent sequences having various non-negative real weights, though the results are not true for all linear recurrences having such weights, as can be seen numerically in the case $k = 3$. Furthermore, numerical evidence (see Table 2 below) suggests that the sequence of zeros in Theorem 5.5 decreases monotonically for all k, as is true in the $k = 2$ case (see [2, Theorem 3.1]).

References

[1] Benjamin, A.T., Quinn, J.J., *Proofs that Really Count: The Art of Combinatorial Proof*, Mathematical Association of America, 2003.

[2] Garth, D., Mills, D., Mitchell, P., Polynomials generated by the Fibonacci sequence, *J. Integer Seq.* **10** (2007), Art. 07.6.8.

[3] Knuth, D.E., *The Art of Computer Programming: Sorting and Searching*, Vol. 3, Addison-Wesley, 1973.

[4] Marden, M., *Geometry of Polynomials*, Second Ed., Mathematical Surveys **3**, American Mathematical Society, 1966.

[5] Mátyás, F., On the generalization of the Fibonacci-coefficient polynomials, *Ann. Math. Inform.* **34** (2007), 71–75.

[6] Mátyás, F., Further generalizations of the Fibonacci-coefficient polynomials, *Ann. Math. Inform.* **35** (2008), 123–128.

[7] Mátyás, F., Liptai, K., Tóth, J.T., Filip, F., Polynomials with special coefficients, *Ann. Math. Inform.* **37** (2010), 101–106.

[8] Mátyás, F., Szalay, L., A note on Tribonacci-coefficient polynomials, *Ann. Math. Inform.* **38** (2011), 95–98.

[9] Munarini, E., A combinatorial interpretation of the generalized Fibonacci numbers, *Adv. in Appl. Math.* **19** (1998), 306–318.

[10] Munarini, E., Generalized q-Fibonacci numbers, *Fibonacci Quart.* **43** (2005), 234–242.

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

Annales Mathematicae et Informaticae
***40** (2012) pp. 77–92*
http://ami.ektf.hu

Implementing the GSOSM algorithm

Nikolett Fanni Menyhárt, Zoltán Hernyák

Eszterházy Károly College, Eger, Hungary
menyhart.nikolett@gmail.com
hz@aries.ektf.com

Submitted November 7, 2012 — Accepted December 11, 2012

Abstract

GSOSM algorithm is a method to reconstruct a surface from a set of scattered points. Implementing this algorithm on a sequential or parallel method contains several interesting questions. In this article we try to give some details on algorithms and problems implementing this method. The aim of the paper is to give ideas and details about the data structures and the implementation, and we draw the attention to possible problems the algorithm may run into. This may help those programmers who implement this type of algorithms for the first time, and will face these challenges.

Keywords: growing cell structures; surface reconstruction; mesh generation; shape modeling;

MSC: 65D17, 68T20, 97P50

1. Introduction

The GSOSM stands for Growing Self-Organizing Surface Maps. This method focuses on the problem, when we have a 3D body, its surface is scanned with a 3D scanner, and we have a set of points from its surface. These unordered, unconnected, unorganized set of points called the Mesh (throughout in this article it is referred as M).

What we want is to reconstruct the body from this scratch. We usually has no conception about the target body, however it is supposed it has a spherical topology. Usually we don't want to have as many points as the Mesh contains when we reach the final state. The reconstructed body's surface is build up with triangles.

The reconstruction starts from a small and simple object, for example from a triangulated cube. This is a proper 3D object; its surface is covered with triangles at the beginning. During the process we pull the points of this object towards the Mesh points, add some new points (and triangles) to make it more complex, until it becomes very similar to the target body. This object in this article is referred to as P.

The GSOSM method and the algorithm are used in this paper is discussed in several articles. In [5] the mesh was divided into subdomains, and was reconstructed in local parts using radial base functions, and was blended together at the end. This approaches was extended and modified in several ways e.q. in [6, 7, 8]. Another approach was presented using SOM (self-organizing map) methods based on Kohonen unsupervised artificial neural network (ANN) model, like GCS (growing cell structures) in [9], or GNG (growing neural gas) in [10]. In [11] the GCS model was transformed into NM (neural mesh) using statistical learning and the Laplacian-based smooth operator was also added. In [12] a GSOSM was introduced using a CCHL (competitive connection Hebbian learning) rule which produces a complete triangulation.

We use [1] as a basis, however other articles contain some modifications on the process (like [3] uses no Laplacian smoothing). We implemented the steps [2], but instead of the standard implementation of the Kohonen neural network, we choose to store the data in usual high level programming language collections, like lists and objects. We separated the code from these data elements, so we cannot say it is a standard neural network approach. In section 4 we give some details which kind of data structures we use.

We implemented the GSOSM steps from paper [2], as at first we wanted to reproduce the results on a sequential way. Here we discuss the problems we found during the implementation.

2. The GSOSM steps

During the preparation of the GSOSM method (described in [1]) we must load the points of M, the points and the surface definition of P, and all the settings, the parameters of the reconstruction from disk or other data source. We use the GeomView Object File Format (.off, see [13]) for reading M and P, as it is suitable for storing point clouds with or without surface information as well.

During the GSOSM we will process the points of the M in a random order, this is why we handle the list of points unordered (unorganized). The random order is important as we want P to grow every direction with the same probability. The main steps are:

1. let $s \in M$ a random point from the target space
2. find $w \in P$ the closest point (shortest distance from s) of the object

3. pull w and the topological neighbours of w towards to s, and set them "active" state with increasing a counter

4. sometimes add a new point to P to make it more complex by *vertex split*, inserting new triangles to the surface as well

5. sometimes delete the inactive points (and triangles) from P's surface using *edge collapse*.

The frequency of "sometimes" when we execute the vertex split or edge collapse is determined by the parameters of the process, usually based on the progress percentage.

3. GSOSM step 1,2: selecting s and find w

In the first step we select a random point $s \in M$ to be processed. We must find the winner point $w \in P$, which is the closest point of P to s.

The selection of w is based on the distance of s and the points of P which can be calculated the following way. Let $p \in P$ be any point, and calculate by

$$\mathrm{dist}(s,p) = \sqrt{(p.x - s.x)^2 + (p.y - s.y)^2 + (p.z - s.z)^2},$$

$$w := \{p : p \in P \wedge \nexists q \in P : \mathrm{dist}(q,s) < \mathrm{dist}(p,s)\}.$$

Note: as we don't need the final value of the distance, it is only compared to determine the minimum, we don't need to calculate the square root, only the expression inside the square root. However in 3D space it's not so simple. We need the winner to pick the closest point to s, to pull this winner and its neighbours towards to s. Let P be a large flat cuboid (as it can be seen if Figure 1) , and let s below the cuboid. A central point X on the other side is the closest point of the triangle's corner forming the surface. If we select this as the winner, and pull it towards to s, the edges around X will cross the lower plane of the cuboid, and after the transformation P loses its spherical topology.

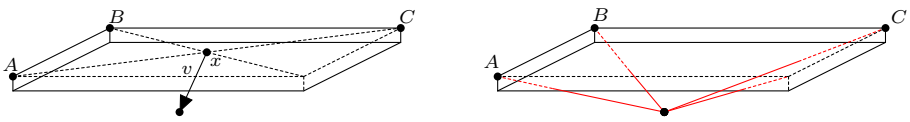


Figure 1: Selecting the winner from the wrong side

To prevent this behaviour, we must store the normal vectors of the triangles on the surface of P. As we use the .OFF file format (mentioned early) to read P, and in this file the normal vectors are not stored – after reading and reconstructing points and triangles of P, we must calculate the normal vectors by ourselves.

To do this, we suppose that at the start phase P is convex. In this case if we have $triangle(A, B, C)$, we can calculate $\overrightarrow{n}$ normal vector by the following way(see Figure 2):

$$\overrightarrow{AB} = B - A, \quad \overrightarrow{AC} = C - A, \quad \overrightarrow{n} = \overrightarrow{AB} \times \overrightarrow{AC}.$$

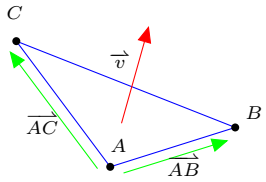


Figure 2: A-B-C triangle and its normal vector

The $\overrightarrow{n}$ is a normal vector, but its direction might be wrong. The normal vector must point out of the body, not to its inner parts. We suppose that at this phase P is convex, so all the points of P are on the same side of the plane defined by the $triangle(A, B, C)$. In this case any of them can be a good representative (except the points, which lays on this plane as well). Select any point x at the surface of P. Calculate $\overrightarrow{Ax}$ and its length. If it is not zero, this x is suitable to determine the correct direction of $\overrightarrow{n}$. If $\overrightarrow{n}$ points to the right direction, the α angle between $\overrightarrow{n}$ and $\overrightarrow{Ax}$ is a non-acute angle. Shift the two vectors to point A, and calculate the value of $\cos\alpha$:

$$\overrightarrow{Ax} := x - A, \parallel \overrightarrow{Ax} \parallel := \sqrt{(x.x - a.x)^2 + (x.y - a.y)^2 + (x.z - a.z)^2}$$

$$\cos\alpha := \frac{\overrightarrow{n} \cdot (x - A)}{\parallel \overrightarrow{n} \parallel \parallel x - A \parallel}$$

Note: as we know, when α is a non-acute angle, $\cos\alpha < 0$. As we can see in the formula, the sign of $\cos\alpha$ depends on the sign of the numerator, as the denominator always positive. So we need to evaluate the numerator expression only to determine the sign of $\cos\alpha$. If its sign is positive, we have to change the direction of $\overrightarrow{n}$ to point into the opposite direction.

At the beginning we suppose that P has a spherical geometry, and we want to keep this property during the progress at all costs. According to this geometry, each $p \in P$ surface point can be a corner point of several edges (and so a part of several triangles). We will notate these as $p.triangles$ and $p.edges$.

Notice, than an edge e can be attached to only two triangles at a time, according to the spherical topology of the object. When we examine any point p from the surface, we must check all the $p.triangles$, their normal vectors to determine if the p point can be the winner or not. When all the normal vectors of all the $p.triangles$ points away from s, then p cannot be selected as a winner. We must calculate the nominator of expression $\cos\alpha$ again as $(\overrightarrow{n} + p)(s - p)$ for all the triangles containing p.

During the initial phase of the program, we load points P, reconstruct the triangles of the surface, and calculates all the normal vectors. It is important to keep the normal vectors, as later the P will lose its convexity, and we won't be able to determine the correct direction of a $\overrightarrow{n}$. When we update the coordinates of any of the points of P, we must re-calculate of the normal vectors of the triangles based on this point. To do that, we can choose one of the following methods:

- At the beginning when the P was convex, and we determine the direction of the normal vector, we store the information that which was point A, and after using $\overrightarrow{AB}$ and $\overrightarrow{AC}$ to calculate the normal vector, we must switch its direction or not. With this extra information, we can recalculate the normal vector any time from now, after the changes of the coordinates.
- When any of the coordinates from A, B or C changes, we recalculate the normal vector immediately. The angle between the new normal vector and the old one must be a sharp angle (the coordinates change little), so $\cos\alpha \geq 0$.

At GSOSM step 4 new points and new triangles are added to the surface – we must take care about their normal vectors as well. At GSOSM step 5 triangles disappears, and points moves heavily. This step will update the normal vectors of the remaining and affected triangles. We will talk about the calculating the normal vectors when we examine these steps closely.

At step 1 we select an $s \in M$ randomly, then delete it from M to prevent selecting the same s later. At step 3 we pull points towards to s, but as we will see later very slightly. After processing each M points, the P won't be complex enough, and the surface of P won't fit tight. So we will process the points of M several times again-and-again. The number of iteration is controlled by a parameter. When the repeat counter is set to n, we might imagine as M that it owns every $s \in M$ points n times. As a set contains every element once, we might handle M as a list instead. But in that case we might select $s_1 \in M$, then $s_2 \in M$ to process, but it might happens that $s_1 = s_2$, and the winner w moves towards to these s twice in a short time. So we choose to store each point once in M, and construct an empty M' set. Select an $s \in M$ randomly, and delete it from M, and add it to M'. When M becomes empty, we switch the M and M', and continues the process with the full set again.

In [1] the points of P is organized into octree-based searching tree to speed up the searching process. Other possibility is to use the Point Cloud library itself, or the algorithm behind it. At the beginning of the implementation, we used a simple list to store the points. Inside this list, we use no special order, so to find the winner w, we must check all the elements of the list.

4. GSOSM data structures

According to section 2, we use the following data structures for the GSOSM process:

1. *Point3D(x,y,z)* is a base data for storing a point's coordinates in 3D space

2. *Mesh(lp)* holds a lists for the target body's points on its surface given by coordinates, *lp* is "list of Point3D"

3. *NVec(x,y,z)* is a normal vector, where x, y, z defines the triangle of the plane

4. *Triangle(e1,e2,e2,nv)* is a triangle on a 3D object's surface, given by three edges $e1$, $e2$ and $e3$, and nv stands for the normal vector of this triangle

5. *Edge(a,b,lt)* is an edge of a 3D object's surface, where a and b are Points (not Point3D, see later), and lt is a list of triangles based on this edge with exactly 2 elements on this list

6. *Point(x,y,z,le,T)* is an advanced point, which stores not only its coordinates, but le the list of edges (and along with the edges the triangles as well) which are connected to this point, and T is the signal (active) counter, its value is 0 at the beginning (see 6)

7. *Body(lp)* is the body described by the list of Points (so the edges and the triangles are given as well).

We found, that in our un-optimized data structure step 1 and 2 (winner find and pulling) is about 34% of total time, the vertex split is about 6%, and the edge collapse is about 58% of total computation time.

5. GSOSM step 2: pulling w towards s

When we select $s \in M$ to process, and $w \in P$ as the winner, we pull w towards s along the $p \to s$ vector. The percentage of the pulling defines the new position of w (marked as w') in the following way:

$$\overrightarrow{w}' := (1 - \lambda)\overrightarrow{w} + \lambda\overrightarrow{s}.$$

This λ value is the parameter of the algorithm. The more strong we pull, the faster the P fit tight to M. The more fast we pull, the P has less time to become complex enough, so the final shape of P won't be good enough.

Another problem appear as we test this part of the algorithm. If we pull strongly, the winner goes to s heavily. When we select another $s' \in M$, close to the previous s, the same winner will be the closest again. In this case a peak arises from a flat space, and the shape of the part of P cannot fit tight (see figure 3). Later, we will see that the fact the same w wins again and again means it becomes highly active, and other points of P turn into useless (inactive). We will erase them at step 5 using edge collapse, and we will lose a lot of points because the winner won't let other points to win.

The other reason to keep λ percentage low is that the more complex P is, the more likely that after a pull of w the w' will arrive inside the body of P (mainly when P loses its convexity). With a small value of λ it is not 100% chance that it

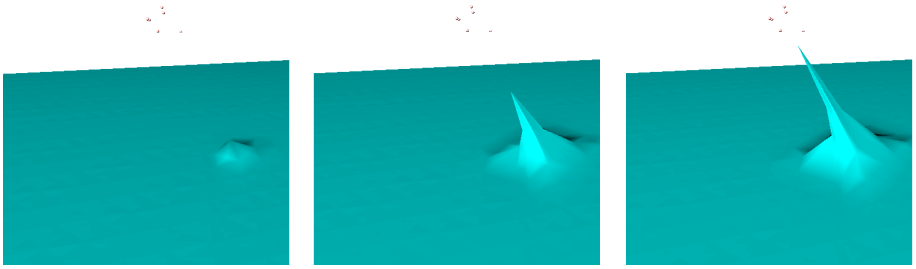


Figure 3: The peak arises

won't happen, but better chance to avoid this. Paper [1] advices the same, talks about unwanted effects, and chance to convergence to local minima or fold-overs.

The value λ is not a constant. At the beginning of the process, larger (but still small) values are better to let the P growing. Later we use smaller and smaller values. It is usual, that the value of λ is determined by a function, which argument is the progress percentage. This function converges to 0, to guarantee the convergence of the algorithm. Instead of a slowly calculable function we evaluate and fix these values in the parameters of the process, connecting the progress percentages with a constant value. When the progress percentage reaches the next limit, the λ changes its value to the next fix value. Paper [1] suggests using constant 6% for the whole process as an experimental value. The choice of this fraction is discussed in details in [4, 3].

5.1. Laplacian smoothing

After pulling the winner, [1] suggest using the Laplacian smoothing. In this case we select and move only the direct topological neighbours of the winner w. Let $R(p) = \{v_1, v_2, \ldots, v_n\}$ be the direct topological neighbours of any $p \in P$. Let us calculate for $v_i \in R(w)$ the Laplacian L as

$$L(v_i) = \frac{1}{valence(v_i)} \sum_{v_k \in R(v_i)} (v_k - v_i).$$

When $\overrightarrow{n_i}$ is one of the normal vectors of point v_i, the tangential component of L can be calculated as

$$L_t(v_i) = L(v_i) - (L(vi) \cdot \overrightarrow{n_i}) \overrightarrow{n_i}$$

and we can update the coordinates of v_i to v_i' as

$$v_i' = v_i + \alpha_l L_t(v_i).$$

In this expression α_l is a constant parameter of the process, [1] suggests using $\alpha_l = 0.06$ value, and suggest repeating this smoothing steps for 5 times.

5.2. Simple neighbours pulling

Another possibility is to update the direct or remote neighbours position is to use similar pulling towards to s as we used pulling w. We may use different λ values for pulling the neighbours as was used for pulling the winner, as we might use different λ values for different distances from w. We might describe the pulling at a given percentage of the process with $(p, \lambda_w, n, \lambda_1, \lambda_2, \ldots, \lambda_n)$ tuple, where p defines the process percentage, $perc_w$ describes how strong must pull the winner, n stands for how far we must walk from w winner, and λ_i $(i \in [1, n])$ sets how strong we must pull the i^{th} neighbour towards s. This method was suggested originally in [3].

5.3. Elastic pull

This method is the "elastic pull" model. This provides a more flexible way to handle the pulling of the neighbours. In the elastic model the surface of P can be imagined as the edges are made of a kind of material (soft rubber, hard and bold rubber, rubbed rope, or a stretchable metal). The material is described using a constant value $\gamma \in \mathbb{R},\ 0 \leqq \gamma \leqq 1$. Larger value means more flexibility. So 0.0 (0%) means no stretch at all, while 1.0 (100%) means that the material can be stretched infinite.

When we have an $edge(A, B)$ (a piece of rope), and we pull one end (A) of this edge towards a direction, we can calculate how much the edge become longer. For example after pulling endpoint A, the edge becomes 40% longer. When the rope is made from soft rubber, which has a value λ , this rope can stretch easily, most of the energy of pulling is absorbed, the power is passed to the other end of the rope is only $0.4(1.0 - 0.8) = 0.08$, which means 8%. This means, that the edges attached to endpoint B stretches with 8% (pulling a point of a spider web causes other points of the web moving along with). The next edge will become longer with 8%, which causes the next level edge become longer with $0.08(1.0 - 0.8) = 0.016$ (1.6%). As we can see, a rope with value 0.8 causes that each next edges will be less longer than the ones before. A rope with no ability to stretch (0.0) means that a M% stretching is passed to the next level with $M\%(1.0-0.0) = M\%$, so it forces the next level to stretch with the same amount. A value of 1.0 (100%, super elastic material) means no stretching force is passed, $M\%(1.0 - 1.0) = 0.0$.

We can use different settings for different value of progression easily. At each progression percentage we can describe the elastic pulling as (p, f, min), where p is the percentage, $f \in \mathbb{R},\ 0.0 \leqq f \leqq 1.0$ the flexibility value of the edges, and $min \in \mathbb{R},\ 0.0 \leqq min < f$ is the minimal value of stretching.

6. Checking the topology of P

As we encounter several unwanted effects, we set up a method to check if any cross-pull happened with P. A cross-pull is when a p point goes inside the body, and any of the $p.edges$ (a section) intersects any of the triangles of the surface. Let $e(P, Q)$ an edge with the endpoints P and Q, and $t(A, B, C)$ a triangle given by corner

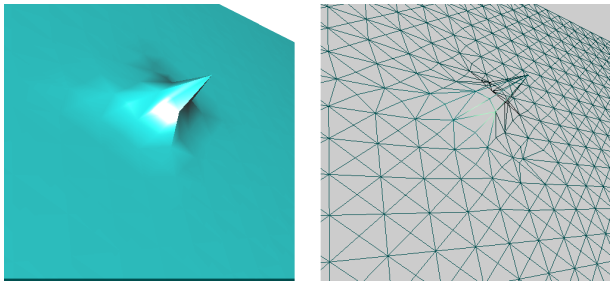


Figure 4: Elastic pull a 2D sheet – without Laplacian smoothing

points A, B and C, and $\overrightarrow{n}$ is the normal vector of the triangle. We can check if an edge e intersects the plane determined by triangle t. We can use the equation for the plane determined by its's normal vector and a point A of the plane:

$$n.x(x - A.x) + n.y(y - A.y) + n.z(z - A.z) = 0.$$

If we inserts both P and Q points into this equation, we can calculate the its final value. If one of the values is 0, that points rests in the plane, other values means the point is far from the plane. If the final values have different signs, it means the two points are in the different side of the plane, otherwise they are in the same side. So calculate the following final values:

$$p_1 = n.x(P.x - A.x) + n.y(P.y - A.y) + n.z(P.z - A.z)$$
$$p_2 = n.x(Q.x - A.x) + n.y(Q.y - A.y) + n.z(Q.z - A.z).$$

When $p_1 > 0 \wedge p_2 > 0$ or $p_1 < 0 \wedge p_2 < 0$ the edge e does not intersect the plane of triangle t, so does not the intersect triangle t itself. If the previous condition evaluates to false: edge e intersects the plane, but not necessarily intersects triangle t (maybe the intersection point is outside of the triangle). Another check must follow. First we calculate the intersection point coordinates (point q). We must need the direction vector of the line $\overrightarrow{pq} = Q - P$, then we must use the parametric equation of a line $S = P + t \cdot \overrightarrow{pq} : \forall t \in \mathbb{R}$ produces point S on the lin. We are searching for that $t \in \mathbb{R}$ which can be inserted into the equation of the plane and produces zero. So we must solve the equation:

$$n.x(P.x + t(Q.x - P.x) - A.x) + n.y(P.y + t(Q.y - P.y) - A.y) \\ + n.z(p.z + t(Q.z - P.z) - A.z) = 0.$$

After elementary steps we have:

$$t = \frac{n.x(P.x - A.x) + n.y(P.y - A.y) + n.z(P.z - A.z)}{-(n.x \cdot pq.x + n.y \cdot pg.y + n.z \cdot pq.z)}.$$

We use insert this value t back to the equation of the line, we give the intersection point F as the following: $F = A + t \cdot \overrightarrow{pq}$.

Using this intersection point F we can check if it is outside the triangle t: if $F.x \leqq A.x \wedge F.x \leqq B.x \wedge F.x \leqq C.x$ or $F.x \geqq A.x \wedge F.x \geqq B.x \wedge F.x \geqq C.x$ (and the same happens with $F.y$ and $F.z$) we can say that edge e won't intersect triangle t.

Otherwise, we calculate whether the point F and point A are on the same side of the line given by the two triangle points B-C. Using the same equations, we can write the equation of the line B-C based on point B as the following: $(x - B.x)(-C.y + B.y) + (y - B.y)(C.x - B.x) = 0$. Inserting point F and A into this equation, we can check if the values have the same signs or not, so are point F and A on the same side of the line. If for point F we got zero, it means the intersection point is on the line, which (in this case) means are on the same side. Checking this for point B (line based on points A and C), then for point C (based on points A and B) we can check if the intersection point is inside the triangle or not. If it is inside, we have an error in the surface.

We must further check the spherical topology property of P as well. It is done by:

- check that all the points $p \in P$ have at least three edges
- all the edges must be associated exactly to two triangles.

7. Setting up "active" state

After selecting w winner and pull towards to s, we increase the "active" counter of w with 1. Paper [1] says only the winner must be updated this way, we consider updating the selected neighbours as well. Otherwise, [1] says all the points of P (except for the winner) the value of this signal counter must be decreased by multiplying its value with α, where $\alpha \in \mathbb{R},\ 0 < \alpha < 1$. Paper [1] suggests using $\alpha = 0.95$ constant value.

This signal counter is used in section 9 to determine if a point $p \in P$ is active or not. When the value of the signal counter smaller than the required value, we delete point p using *edge collapse*.

Later this paper talks about the machine accuracy problem, and substitutes this method by a simple one: if a point p was not active since the last edge collapse – it must be handle as inactive point.

So we need only a logical value attached to each point p, which are initially set to $false$. When a point is selected as a winner, we set this flag to $true$. During the edge collapse phase, we handles all the points as inactive, which are still $false$. At the end of the edge collapse, we set all remaining points back to $false$. Another way is when we have a global iteration counter, when a p is selected to be a winner, we set the flag to the actual value of the counter (in which step was he selected winner). We can handle a point as inactive, when its "last winner flag" is too low, it has not been selected as a winner since the last edge collapse run.

8. Vertex split

As we described, steps 1–3 move the points of P towards the mesh M surface points. These steps will not increase (or decrease) the complexity of P, so applying these steps won't make P to be very similar to M. Step 4 targets to make P more complex by adding new points to it. Simply adding a new point won't help, as P holds not only points but edges and triangles as well. After adding a new point we must insert it into the edges and triangles properly, keeping the spherical topology of P.

First we select the environment where the new Q point can be added. Select the most active $A \in P$ point (with the highest signal counter value), and one of its direct topological neighbour $B \in P$, the most active neighbours of A.

Note: the standard method of vertex split suggests selecting point $A \in P$ with the most valence, as this point really need splitting. If we have several $A_1, \ldots, A_n \in P$ points with the same highest valence value, we can choose between them paying attention to its signal counter. An alternative way can be the following: select the $A \in P$ to apply the vertex split finding the longest edge, or one of a triangle base point with the largest area.

Vertex split will happen using the $edge(A, B)$ (there must be an edge between A and B as point B is a direct topological neighbour of point A). As P has spherical topology, there must be exactly two triangles based on this edge, so A and B must have two common direct topological neighbours (C_1 and C_2). Let's create a new point C. This new point C won't be on the section AB, as it is "edge split", and we want to use "vertex split". This new point C must be around A.

On a 2D space coordinates C might be chosen as an inner point either inside the $triangle(A, B, C_1)$ or inside $triangle(A, B, C_2)$.

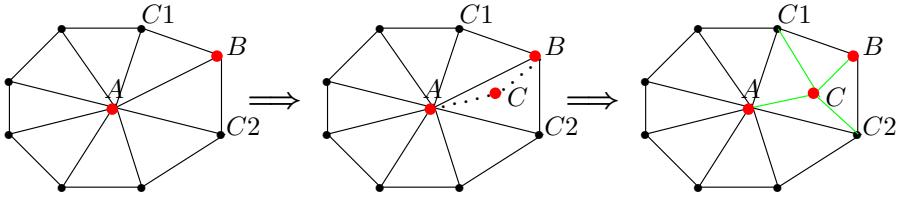


Figure 5: Vertex split – first three phase

In 3D we can put this new point around the $edge(A, B)$, not necessary on the triangles' plane. Let's calculate its coordinates as the following (for example):

$$C.x = (3/8A.x) + (3/8B.x) + (1/8C_1.x) + (1/8C_2.x)$$
$$C.y = (3/8A.y) + (3/8B.y) + (1/8C_1.y) + (1/8C_2.y)$$
$$C.z = (3/8A.z) + (3/8B.z) + (1/8C_1.z) + (1/8C_2.z).$$

The steps of vertex split are as follows:

1. add new point C to the point list of P with no edge and no triangle information
2. delete $edge(A, B)$ and the triangles based on this edge ($triangle(A, B, C_1)$ and $triangle(A, B, C_2)$
3. create new $edge(A, C)$ and $edge(C, B)$ (don't forget to add it to $A.edges$, $B.edges$ and $C.edges$)
4. create and insert $triangle(A, C, C_1)$, $triangle(A, C, C_2)$, then $triangle(C, B, C_1)$ and $triangle(C, B, C_2)$ into P properly.

Notice that at this point the valence of A does not decreased, nor the complexity of P increased, now we have 1 more edge, and 2 more triangles, and the long edge AB has been replaced by two short edges.

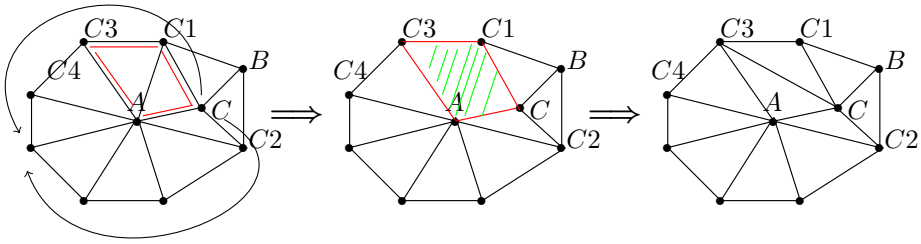


Figure 6: Vertex split – decreasing the valence of A

Further triangulation steps may be required. This time the steps are different (p is the selected point around A):

1. remove $edge(A, p)$, so remove $triangle(A, C, p)$ and $triangle(A, P, X)$ as well
2. define new $edge(X, C)$
3. define new triangles X, C, P and X, A, C
4. only when the new edge does not intersects any faces except for the removed ones.

At this phase we added a new point C properly into the surface of P. This surface has no hole anymore, has more point than before (more complex). To do a better job at this point, we might want to redirect some edges of A into C, to decrease the valence of A, and increase of C. If A has a valence of n, we might want to redirect $n/2$ edges into C.

To do that, first we gather all the direct neighbours points of A into set N_A^1. Notice: after execution of the previous steps, $C, C_1, C_2 \in N_A^1$, but B is not. Let's order these points as we could walk around A, starting from C, going to the direction towards C_1.

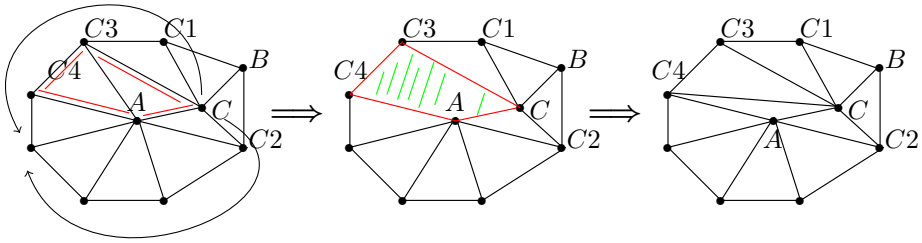


Figure 7: Vertex splitted cube – without decreased valences

We might think it is a good idea to decrease the valence of A, that processing this ordered list one-by-one we can redirect some edges to C. Select the $A-C-C_1-C_3$ quadrilateral, and follow the steps:

1. delete $edge(A, C_1)$ (decreasing the valence of A), and all the triangles based on this edge ($triangle(A, C, C_1)$ and $triangle(A, C_1, C_3)$)
2. add $edge(C, C_3)$ (increasing the valence of C)
3. add $triangle(C, C_1, C_3)$, and $triangle(A, C, C_3)$.

We can continue and repeat these steps walking around this direction for a few steps, and then we can turn around and walk on the other direction around A as well, redirecting the edges, until $n/2$ edges are attached to C.

First of all: notice, that there is another possibility, than creating new edge ($edge(C, C_3)$ for example) between remote points: the edges can intersect the surface of P. So after planning a redirect, we can check the integrity and correctness of the surface of P using the method described in 6. If any error encounters, we step back to the correct state.

Second: what is working on 2D, won't fit the 3D world. Let us suppose we have a cube, each square contains 2 triangles. One of the flats we vertex split, inserting a new C point and redirect the triangles on that square from the corner points to C. After this we find our $A - C - C_1 - C_3$ quadrilateral, and follow the steps. On Figure 8 we can see the schematics and on Figure 9 we can see how a cube can be deformed by this method.

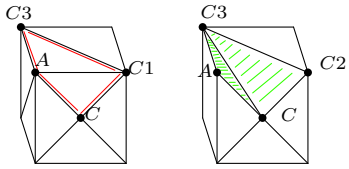


Figure 8: Vertex splitted cube – before and after redirection

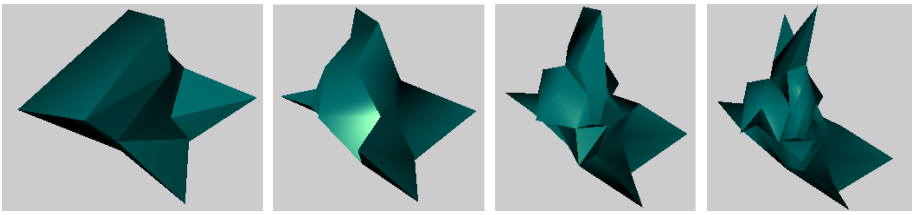


Figure 9: Vertex splitted cube after several steps

9. Edge collapse

According to the algorithm, we select s, find w, pull w towards s, somehow we pull the neighbours of w towards s as well, and sometimes we add new vertices to increase the complexity of P using vertex split. After a short time we will find, that P has inactive points, which never becomes a winner, they are far from the points of M, and are useless. Then we can clear them by another standard method called edge collapse.

We can set when to execute this step based on the number of the vertices of P (based on its complexity) or after every ν iterations. In [1] the suggested method is to execute after every νn iterations, where $\nu = 20$, and n is the size of P.

Selecting the less inactive or useless nodes we might select all the nodes which were not active (not selected as winner). We might expand the immunity against clearing to the ones which were selected and moved as a neighbour of any winner as well, however [1] suggests giving immunity only for the winners only.

When we select a node $a \in P$ to clear, we must select a direct neighbour $b \in P$ as well. We will clear node a redirecting its edges to b, so b's valence will become higher.

First we know that this step seems to be very easy in 2D, but in 3D it can yield unwanted effects, and the spherical geometry can fail, but this effect arises at the end of the collapsing. So we advise to save the entire state of P before the collapsing as it will be modified several ways, and we might roll back to the original state at the end. Another possibility is not to make any modification on the state of P, instead we collect the modification instructions into a list, then check the state of P according to this modifications steps, and if we find any failure, drop this list and do nothing.

In 2D the steps seems very easy and clear:

- delete $edge(a, b)$, and all the triangles based on this edge ($triangle(a, b, c)$ and $triangle(a, b, d)$)
- find every triangles containing point a, and replace this corner point to b (there is no triangle still exists which contains not only a and b as well, because we drop them at the previous step)

- delete point a, as it loses any connection to other points on the surface of P.

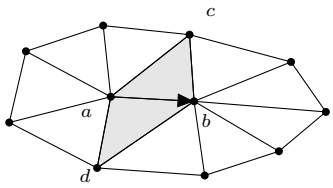


Figure 10: Edge collapse

The main problem is in step 2. When we have a $triangle(x, y, a)$ for example, we must replace $edge(x, a)$ to $edge(x, b)$. Might there is already an $edge(x, b)$ in P, so this step sometimes creates a new edge, sometimes not. The same is true for the triangles: $triangle(x, y, a)$ becomes $triangle(x, y, b)$, but sometimes this triangle already exists.

To demonstrate the problem, see figure 11. We have a tetrahedron $A-B-C-D$ with point E on the edge between BD. It is interesting, that collapsing $B \to E$ won't cause any problem, we would drop triangles $A - E - B$ and $E - B - C$, and triangle $A - B - C$ would become $A - E - C$ which will close the shape, and the tetrahedron still remain tetrahedron. But if we try to collapse edge $E \to A$, the whole side covered with $A - B - E$ and $A - E - B$ triangles would disappear. After that $B - E - C$ goes into $B - A - C$ which is already exists, and $E - C - B$ changes to $A - C - B$ which already exists as well. After the edge collapse steps we would have only two triangles, and the shape of this 3D object loses its spherical geometry and becomes a folded paper. This is the reason why we must prepare to roll back the edge collapse at any step we made.

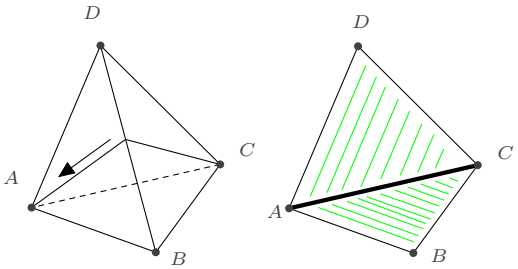


Figure 11: Edge collapse fails

Acknowledgements. First we want to thank to Annamária Stefán, who helped to develop and test the application, and gives several tips related to the topic. Besides, we want to thank László Balog who developed the useful tool, the *TurnOffTheWorld* application, which shows us the results of the GSOSM processing. Least but not

last we want to thank Miklós Hoffmann for supportive presence in this project, for his guidance, for his patience, and the constant pleasant atmosphere he created.

References

[1] I. P. Ivrissimitzis, W-K. Jeong, H-P. Seidel, Using Growing Cell Structures for Surface Reconstruction, *Shape Modeling International, 2003*, ISBN: 0-7695-1909-1, Page 78–86, 12–15 May 2003.

[2] The Point Cloud library (`http://www.pointclouds.org`).

[3] Hoffmann M., Numerical control of Kohonen neural network for scattered data approximation, Numerical Algorithms 39, 175–186, 2005.

[4] Hoffmann, M., Modified Kohonen Neural Network for Surface Reconstruction, Publ. Math., 54 suppl., 857–764., 1999.

[5] Tobor, I., Reuter, P., Schlick, C., Multi-scale reconstruction of implicit surfaces with attributes from large unorganized point sets, Shape Modeling Applications, 2004. Proceedings, pp. 19–30, 2004.

[6] Yutaka Ohtake, Alexander Belyaev, Hans-Peter Seidel, Sparse Surface Reconstruction with Adaptive Partition of Unity and Radial Basis Functions, Graphical Models, 68(1), pp. 15–24, 2006.

[7] Qi Xia, Shatin Wang, Xiaojun Wu, Orthogonal Least Squares in Partition of Unity Surface Reconstruction with Radial Basis Function, Geometric Modeling and Imaging–New Trends, 2006, pp. 28–33.

[8] Yi-Ling Chen, Shang-Hong Lai, A Partition-of-Unity Based Algorithm for Implicit Surface Reconstruction Using Belief Propagation, In Proceedings of Shape Modeling International 2007 (SMI'07), Lyon, France, June 2007.

[9] B. Fritzke, Growing Cell Structures – a self-organizing network for unsupervised and supervised learning, Neural Networks vol 7. no. 9, pp. 1441–1460, 1994.

[10] B. Fritzke, A growing neural gas network learns topologies, Advances in Neural Information Processing Systems, vol 7, MIT Press, pp. 625–632, 1995.

[11] Ivrissimtzis, Ioannis, Jeong, Won-Ki, Seidel, Hans-Peter, Neural meshes: statistical learning methods in surface reconstruction, Max-Planck-Institut für Informatik, MPI-I-2003-4-007, ISSN: 0946-011X, 2003.

[12] Vilson Luiz Dalle Mole, Aluizio Fausto Ribeiro Araújo, Growing Self-Organizing Surface Map: Learning a Surface Topology from a Point Cloud, Neural Computation Vol. 22, No. 3, pp. 689–729, 2010.

[13] `http://people.sc.fsu.edu/~jburkardt/data/off/off.html`

Annales Mathematicae et Informaticae
***40** (2012) pp. 93–103*
http://ami.ektf.hu

On unification of some weak separation properties*

Bishwambhar Roy[a], Ritu Sen[b]

[a]Department of Mathematics
Women's Christian College, India
bishwambhar_roy@yahoo.co.in

[b]Department of Mathematics
S. A. Jaipuria College, India
ritu_sen29@yahoo.co.in

Submitted May 14, 2012 — Accepted October 13, 2012

Abstract

In this paper, a new kind of sets called regular ψ-generalized closed (briefly $r\psi g$-closed) sets are introduced and studied in a topological space. Some of their properties are investigated. Finally, some unifications of some weak forms of almost regular, almost normal and mildly normal spaces have been given.

Keywords: ψ-open set, $r\psi g$-closed set, almost ψ-regular space, almost ψ-normal space, mildly ψ-normal space.

MSC: 54D15, 54A05, 54C08.

1. Introduction

The concept of generalized closed sets in a topological space was introduced by N. Levine [11]. After that, the concept of generalized closed sets has been investigated by many mathematicians. It is well known that separation axioms are one of the basic subjects of study in general topology and in several branches of mathematics. In 1973, Singal et al. introduced the concept of almost regular [25], almost normal

*The first author acknowledges the financial support from UGC, New Delhi.

[26] and mildly normal [27] spaces. Recently, Ekici, Noiri and Park [4, 5, 6, 19, 21, 22] continued the study of several weaker forms of separation axioms.

Throughout this paper (X, τ) always means a topological space on which no separation axioms are assumed unless explicitly stated. The closure and interior of a set $A(\subseteq X)$ are denoted by clA and $intA$ respectively. A subset A is said to be regular open (resp. regular closed) if $A = intclA$ (resp. $A = clintA$). The collection of all regular open (regular closed) sets in a topological space (X, τ) is denoted by $RO(X)$ (resp. $RC(X)$). The δ-closure [28] of a subset A of X is denoted by $cl_\delta A$ and is defined by $cl_\delta A = \{x : A \cap U \neq \emptyset$, for each $U \in RO(X)$ with $x \in U\}$. Let (X, τ) be a space and $A \subseteq X$. A point $x \in X$ is called a condensation point of A if for each $U \in \tau$ with $x \in U$, the set $U \cap A$ is uncountable. A is called ω-closed [8] if it contains all its condensation points. The complement of an ω-closed set is called ω-open. It is well known that the family of all ω-open subsets of a space (X, τ), denoted by τ_ω, forms a topology on X finer than τ. A subset A of a space X is said to be preopen [14] (resp. semi-open [10], δ-preopen [23], α-open [16], β-open [1]) if $A \subseteq intclA$ (resp. $A \subseteq clintA$, $A \subseteq intcl_\delta A$, $A \subseteq intclintA$, $A \subseteq clintclA$). The family of all preopen (resp. semi-open, δ-preopen, α-open, β-open) sets in a space X is denoted by $PO(X)$ (resp. $SO(X)$, δ-$PO(X)$, $\alpha O(X)$, $\beta O(X)$.

We now recall a few definitions and observe that many of the existing relevant definitions considered in various papers turn out to be special cases of the ones given below.

Definition 1.1. [3] Let (X, τ) be a topological space. A mapping $\psi : \mathcal{P}(X) \to \mathcal{P}(X)$ is called an operation on $\mathcal{P}(X)$, where $\mathcal{P}(X)$ denotes as usual the power set of X, if for each $A \in \mathcal{P}(X) \setminus \{\emptyset\}$, $intA \subseteq \psi(A)$ and $\psi(\emptyset) = \emptyset$.

The set of all operations on a space X will be denoted by $\mathcal{O}(X)$.

Observation 1.2. It is easy to check that some examples of operations on a space X are the well known operators viz. int, $intcl$, $intcl_\delta$, $clint$, $intclint$, $clintcl$.

Definition 1.3. [3] Let ψ be an operation on a space (X, τ). Then a subset A of X is called ψ-open if $A \subseteq \psi(A)$. Complements of ψ-open sets will be called ψ-closed sets. The family of all ψ-open (resp. ψ-closed) subsets of X is denoted by $\psi\mathcal{O}(X)$ (resp. $\psi\mathcal{C}(X)$).

Observation 1.4. It is clear that if ψ stands for any of the operators int, $intcl$, $intcl_\delta$, $clint$, $intclint$, $clintcl$, then ψ-openness of a subset A of X coincides with respectively the openness, preopenness, δ-preopenness, semi-openness, α-openness and β-openness of A (see [13, 23, 10, 12, 1]).

Definition 1.5. [3] Let (X, τ) be a topological space, $\psi \in \mathcal{O}(X)$ and $A \subseteq X$. Then the intersection of all ψ-closed sets containing A is called the ψ-closure of A, denoted by ψ-clA; alternately, ψ-clA is the smallest ψ-closed set containing A. The union of all ψ-open subsets of G is the ψ-interior of G, denoted by ψ-$intG$.

It is known from [9] that $x \in \psi$-clA iff $A \cap U \neq \emptyset$, for all U with $x \in U \in \psi\mathcal{O}(X)$ and $x \in \psi$-$intG$ iff $\exists\ x \in U \in \psi O(X)$ such that $x \in U \subseteq G$. In [9], it is also shown that $X \setminus \psi$-$clG = \psi$-$int(X \setminus G)$.

Observation 1.6. Obviously if one takes interior as the operation ψ, then ψ-closure becomes equivalent to the usual closure. Similarly, ψ-closure becomes pcl, pcl_δ, scl, α-cl, β-cl, if ψ is taken to stand for the operators $intcl$, $intcl_\delta$, $clint$, $intclint$ and $clintcl$ respectively (see [13, 23, 10, 12, 1] for details).

Definition 1.7. A subset A of a space (X, τ) is called
(a) generalized closed (briefly, g-closed) [11] if $clA \subseteq U$ whenever $A \subseteq U$ and $U \in \tau$;
(b) regular generalized closed (briefly, rg-closed) [20] if $clA \subseteq U$ whenever $A \subseteq U \in RO(X)$;
(c) generalized preregular closed [7] (briefly, gpr-closed), or preregular generalized closed [18] if $pclA \subseteq U$ whenever $A \subseteq U \in RO(X)$;
(d) $r\alpha g$-closed [19] if $cl_\alpha A \subseteq U$ whenever $A \subseteq U \in RO(X)$;
(e) $g\delta pr$-closed [6] if $pcl_\delta A \subseteq U$ whenever $A \subseteq U \in RO(X)$;
(f) $rg\omega$-closed [2] if $cl_\omega(A) \subseteq U$ whenever $A \subseteq U \in RO(X)$.

2. Properties of $r\psi g$-closed sets

Definition 2.1. Let ψ be an operation on a topological space (X, τ). A subset A of X is called a regular ψ-generalized closed set or simply an $r\psi g$-closed set (resp. $g\psi$-closed set [24]) if ψ-$cl(A) \subseteq U$ whenever $A \subseteq U \in RO(X)$ (resp. $A \subseteq U \in \tau$). The complement of an $r\psi g$-closed set (resp. $g\psi$-closed set) is called an $r\psi g$-open (resp. $g\psi$-open [24]) set.

Remark 2.2. Let ψ be an operation on a topological space (X, τ). Then we have the following relation between $r\psi g$-closed sets and other known sets :

$$\psi\text{-closed set} \Rightarrow g\psi\text{-closed set} \Rightarrow r\psi g\text{-closed set}$$

Example 2.3. Let $X = \{a, b, c\}$ and $\tau = \{\emptyset, \{a\}, \{a, b\}, X\}$. Then (X, τ) is a topological space. Consider the mapping $\psi : \mathcal{P}(X) \to P(X)$ defined by $\psi(A) = intA$ for all $A \subseteq X$. Then $\psi \in O(X)$. It can shown that $\{a\}$ is a $r\psi g$-closed set which is not $g\psi$-closed.

Remark 2.4. Obviously if on a space (X, τ) one takes the operation $\psi = int$, then $r\psi g$-closed sets become equivalent to rg-closed sets [7, 20]. Similarly, $r\psi g$-closed sets become gpr-closed sets [15, 18, 21], $r\alpha g$-closed sets [19], $g\delta pr$-closed sets [6], $rg\omega$-closed sets [2] if the role of ψ is taken to stand for $intcl$, $intclint$, $intcl_\delta$, τ_ω-int respectively.

Some characterizations of some weak separation properties via $g\gamma$-closed set with the operation were studied in [4].

Definition 2.5. A subset A of a space (X, τ) is said to be $g\gamma$-closed [4] if γ-$cl(A) \subseteq U$ whenever $A \subseteq U$ and $U \in \tau$. The complement of $g\gamma$-closed set is said to be $g\gamma$-open [4].

The next two examples show that union (intersection) of two $r\psi g$-closed sets is not in general an $r\psi g$-closed set.

Example 2.6. (a) Let $X = \{a, b, c, d, e\}$ and $\tau = \{\emptyset, X, \{a, b\}, \{c, d\}, \{a, b, c, d\}\}$. Then (X, τ) is a topological space with $RO(X) = \{\emptyset, X, \{a, b\}, \{c, d\}\}$. Consider the mapping $\psi : \mathcal{P}(X) \to \mathcal{P}(X)$ by $\psi(A) = intclA$ for all $A \subseteq X$. Then $\psi \in O(X)$. It is easy to check that $\{a\}$ and $\{b\}$ are two $r\psi g$-closed sets but their union $\{a, b\}$ is not $r\psi g$-closed.

(b) Consider the topological space (X, τ), where $X = \{a, b, c\}$ and $\tau = \{\emptyset, X, \{a\}, \{b\}, \{a, b\}\}$. Let $\psi : \mathcal{P}(X) \to \mathcal{P}(X)$ be a map defined by $\psi(A) = intA$ for all $A \subseteq X$. Then $\psi \in O(X)$. It is easy to check that $\{a, b\}$ and $\{a, c\}$ are two $r\psi g$-closed set but their intersection $\{a\}$ is not a $r\psi g$-closed set.

Theorem 2.7. *Let ψ be an operation on a topological space (X, τ). Let $A \subseteq X$ be an $r\psi g$-closed subset of X. Then ψ-$cl(A) \setminus A$ does not contain any non-empty regular closed set.*

Proof. Let F be a regular closed subset of (X, τ) such that $F \subseteq \psi$-$cl(A) \setminus A$. Then $F \subseteq X \setminus A$ and hence $A \subseteq X \setminus F \in RO(X)$. Since A is $r\psi g$-closed, ψ-$cl(A) \subseteq X \setminus F$ and hence $F \subseteq X \setminus \psi$-$cl(A)$. So $F \subseteq \psi$-$cl(A) \cap (X \setminus \psi$-$cl(A)) = \emptyset$. □

That the converse of the above theorem is false as shown by the next example.

Example 2.8. Let $X = \{a, b, c, d\}$ and $\tau = \{\emptyset, X, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}\}$. Then (X, τ) is a topological space. Consider the mapping $\psi : \mathcal{P}(X) \to P(X)$ defined by $\psi(A) = intA$ for all $A \subseteq X$. Then $\psi \in O(X)$. Let $A = \{a\}$. Then ψ-$cl(A) \setminus A = \{a, c, d\} \setminus \{a\} = \{c, d\}$ does not contain any non-empty regular closed set. But A is not $r\psi g$-closed.

Theorem 2.9. *Let ψ be an operation on a topological space (X, τ). Then a subset A is $r\psi g$-open iff $F \subseteq \psi$-$int(A)$ whenever F is a regular closed subset such that $F \subseteq A$.*

Proof. Let A be an $r\psi g$-open subset of X and F be a regular closed subset of X such that $F \subseteq A$. Then $X \setminus A$ is an $r\psi g$-closed set and $X \setminus A \subseteq X \setminus F \in RO(X)$. So ψ-$cl(X \setminus A) = X \setminus \psi$-$int(A) \subseteq X \setminus F$. Thus $F \subseteq \psi$-$int(A)$.

Conversely, let $F \subseteq \psi$-$int(A)$ whenever F is regular closed such that $F \subseteq A$. Let $X \setminus A \subseteq U$ where $U \in RO(X)$. Then $X \setminus U \subseteq A$ and $X \setminus U$ is regular closed. By the assumption, $X \setminus U \subseteq \psi$-$int(A)$ and hence ψ-$cl(X \setminus A) = X \setminus \psi$-$int(A) \subseteq U$. Hence $X \setminus A$ is $r\psi g$-closed and hence A is $r\psi g$-open. □

Theorem 2.10. *Let ψ be an operation on a topological space (X, τ) and A be an $r\psi g$-closed subset of X. If $B \subseteq X$ be such that $A \subseteq B \subseteq \psi$-$cl(A)$, then B is also an $r\psi g$-closed set.*

Proof. Let A be an $r\psi g$-closed set and $B \subseteq U \in RO(X)$. Then $A \subseteq U \in RO(X)$ and hence ψ-$cl(A) \subseteq U$. Thus by monotonicity and idempotent property of ψ-cl we have ψ-$cl(B) \subseteq U$, showing B to be $r\psi g$-closed. □

Theorem 2.11. *Let (X, τ) be a topological space and ψ be an operation on X. If A is an $r\psi g$-closed subset of X, then ψ-$cl(A) \setminus A$ is $r\psi g$-open.*

Proof. Let A be an $r\psi g$-closed subset of (X,τ) and F be a regular closed subset such that $F \subseteq \psi\text{-}cl(A) \setminus A$, so by Theorem 2.7, $F = \emptyset$ and thus $F \subseteq \psi\text{-}int(\psi\text{-}cl(A) \setminus A)$. So by Theorem 2.9, $\psi\text{-}cl(A) \setminus A$ is $r\psi g$-open. □

Example 2.12. Consider Example 2.8 once again. If we take $A = \{a\}$ then $\psi\text{-}cl(A) \setminus A = \{c, d\}$ is $r\psi g$-open but A is not $r\psi g$-closed.

Definition 2.13. Let ψ be an operation on a topological space (X,τ). Then (X,τ) is said to be $r\psi g$-$T_{1/2}$ if every $r\psi g$-closed set in (X,τ) is ψ-closed.

Theorem 2.14. *Let ψ be an operation on a topological space (X,τ). Then the following are equivalent :*
(i) (X,τ) is $r\psi g$-$T_{1/2}$.
(ii) Every singleton is either regular closed or ψ-open.

Proof. (i) ⇒ (ii) : Suppose $\{x\}$ is not regular closed for some $x \in X$. Then $X \setminus \{x\}$ is not regular open and hence X is the only regular open set containing $X \setminus \{x\}$. Thus $X \setminus \{x\}$ is $r\psi g$-closed. Hence $X \setminus \{x\}$ is ψ-closed (by (i)). Thus $\{x\}$ is ψ-open.

(ii) ⇒ (i) : Let A be any $r\psi g$-closed subset of (X,τ) and $x \in \psi\text{-}cl(A)$. We have to show that $x \in A$. If $\{x\}$ is regular closed and $x \notin A$, then $x \in \psi\text{-}cl(A) \setminus A$. Thus $\psi\text{-}cl(A) \setminus A$ contains a non-empty regular closed set $\{x\}$, a contradiction to Theorem 2.7. So $x \in A$. Again if $\{x\}$ is ψ-open, then since $x \in \psi\text{-}cl(A)$, it follows that $x \in A$. So in both the cases $x \in A$. Thus A is ψ-closed. □

Remark 2.15. Let ψ be an operation on a space (X,τ). Then every $r\psi g$-$T_{1/2}$ space reduces to preregular $T_{1/2}$ [7] (resp. δp-regular $T_{1/2}$ [6], $rg\omega$-$T_{1/2}$ [2]) if one takes ψ to be $PO(X)$ (resp. $\delta PO(X)$, τ_ω).

Theorem 2.16. *Let ψ be an operation on a topological space (X,τ). Then the following are equivalent :*
(i) Every regular open set of X is ψ-closed.
(ii) Every subset of X is $r\psi g$-closed.

Proof. (i) ⇒ (ii) : Let $A \subseteq U \in RO(X)$. Then by (i) U is ψ-closed and so $\psi\text{-}cl(A) \subseteq \psi\text{-}cl(U) = U$. Thus A is $r\psi g$-closed.

(ii) ⇒ (i) : Let $U \in RO(X)$. Then by (ii), U is $r\psi g$-closed and hence $\psi\text{-}cl(U) \subseteq U$, showing U to be a ψ-closed set. □

Theorem 2.17. *Let ψ be an operation on a topological space (X,τ). If A be $r\psi g$-open then $U = X$ whenever U is regular open and $\psi\text{-}int(A) \cup (X \setminus A) \subseteq U$.*

Proof. Let $U \in RO(X)$ and $\psi\text{-}int(A) \cup (X \setminus A) \subseteq U$ for an $r\psi g$-open set A. Then $X \setminus U \subseteq [X \setminus \psi\text{-}int(A)] \cap A$, i.e., $X \setminus U \subseteq \psi\text{-}cl(X \setminus A) \setminus (X \setminus A)$. Since $X \setminus A$ is $r\psi g$-closed by Theorem 2.7, $X \setminus U = \emptyset$ and hence $U = X$. □

The converse of the theorem above is not always true as shown by the following example.

Example 2.18. Let $X = \{a, b, c, d\}$ and $\tau = \{X, \emptyset, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}, \{a, b, d\}\}$. Consider the mapping $\psi : \mathcal{P}(X) \to \mathcal{P}(X)$ defined by $\psi(A) = intclA$ for all $A \subseteq X$. Then $\psi \in O(X)$. Let $A = \{b, c, d\}$. It can be easily verify that X is the only regular open set containing $\psi\text{-}int(A) \cup (X \setminus A)$ but A is not a $r\psi g$-open set in X.

3. Weak separation properties

Definition 3.1. Let (X, τ) be a topological space and ψ be an operation on X. Then (X, τ) is said to be almost ψ-regular if for each regular closed set F of X and each $x \notin F$ there exist disjoint ψ-open sets U and V such that $x \in U$, $F \subseteq V$.

Remark 3.2. Let ψ be an operation on a space (X, τ). Then every almost ψ-regular space reduces to an almost regular [25] (resp. almost p-regular [13], almost δp-regular [6], almost α-regular) space if one takes ψ to be int (resp. $intcl$, $intcl_\delta$, $intclint$).

Theorem 3.3. *Let ψ be an operation on a topological space (X, τ). Then the following statements are equivalent :*
(i) X is almost ψ-regular.
(ii) For each $x \in X$ and each $U \in RO(X)$ with $x \in U$ there exists $V \in \psi O(X)$ such that $x \in V \subseteq \psi\text{-}cl(V) \subseteq U$.
(iii) For each regular closed set F of X, $\cap\{\psi\text{-}cl(V) : F \subseteq V \in \psi O(X)\} = F$.
(iv) For each $A \subseteq X$ and each $U \in RO(X)$ with $A \cap U \neq \emptyset$, there exists $V \in \psi O(X)$ such that $A \cap V \neq \emptyset$ and $\psi\text{-}cl(V) \subseteq U$.
(v) For each non-empty subset A of X and each regular closed subset F of X with $A \cap F = \emptyset$, there exist $V, W \in \psi O(X)$ such that $A \cap V \neq \emptyset$, $F \subseteq W$ and $W \cap V = \emptyset$.
(vi) For each regular closed set F and $x \notin F$, there exist $U \in \psi O(X)$ and an $r\psi g$-open set V such that $x \in U$, $F \subseteq V$ and $U \cap V = \emptyset$.
(vii) For each $A \subseteq X$ and each regular closed set F with $A \cap F = \emptyset$, there exist $U \in \psi O(X)$ and an $r\psi g$-open set V such that $A \cap U \neq \emptyset$, $F \subseteq V$ and $U \cap V = \emptyset$.

Proof. (i) $\Rightarrow$ (ii) : Let $U \in RO(X)$ with $x \in U$. Then $x \notin X \setminus U \in RC(X)$. Thus by (i), there exist disjoint $G, V \in \psi O(X)$ such that $x \in V$, $X \setminus U \subseteq G$. So, $x \in V \subseteq \psi\text{-}cl(V) \subseteq \psi\text{-}cl(X \setminus G) = X \setminus G \subseteq U$.

(ii) $\Rightarrow$ (iii) : Let $X \setminus F \in RO(X)$ and $x \in X \setminus F$. Then by (ii), there exists $U \in \psi O(X)$ such that $x \in U \subseteq \psi\text{-}cl(U) \subseteq X \setminus F$. So $F \subseteq X \setminus \psi\text{-}cl(U) = V$ (say) $\in \psi O(X)$ and $U \cap V = \emptyset$. Then $x \notin \psi\text{-}cl(V)$. Thus $F \supseteq \cap\{\psi\text{-}cl(V) : F \subseteq V \in \psi O(X)\}$.

(iii) $\Rightarrow$ (iv) : Let A be a subset of X and $U \in RO(X)$ be such that $A \cap U \neq \emptyset$. Let $x \in A \cap U$. Then $x \notin X \setminus U$. Hence by (iii), there exists $W \in \psi O(X)$ such that $X \setminus U \subseteq W$ and $x \notin \psi\text{-}cl(W)$. Put $V = X \setminus \psi\text{-}cl(W)$. Then $V \in \psi O(X)$ contains x and hence $A \cap V \neq \emptyset$. Now $V \subseteq X \setminus W$, so $\psi\text{-}cl(V) \subseteq X \setminus W \subseteq U$.

(iv) $\Rightarrow$ (v) : Let F be a set as in the hypothesis of (v). Then $X \setminus F \in RO(X)$ with $A \cap (X \setminus F) \neq \emptyset$ and hence by (iv), there exists $V \in \psi O(X)$ such that $A \cap V \neq \emptyset$

and ψ-$cl(V) \subseteq X \setminus F$. If we put $W = X \setminus \psi$-$cl(V)$, then $W \in \psi O(X)$, $F \subseteq W$ and $W \cap V = \emptyset$.

(v) $\Rightarrow$ (i) : Let F be a regular closed set such that $x \notin F$. Then $F \cap \{x\} = \emptyset$. Thus by (v), there exist $U, V \in \psi O(X)$ such that $x \in U$, $F \subseteq V$ and $U \cap V = \emptyset$.

(i) $\Rightarrow$ (vi) : Trivial in view of Remark 2.2.

(vi) $\Rightarrow$ (vii) : Let $A \subseteq X$ and F be a regular closed set with $A \cap F = \emptyset$. Then for $a \in A$, $a \notin F$ and hence by (vi), there exist $U \in \psi O(X)$ and an $r\psi g$-open set V such that $a \in U$, $F \subseteq V$ and $U \cap V = \emptyset$. So $A \cap U \neq \emptyset$, $F \subseteq V$ and $U \cap V = \emptyset$.

(vii) $\Rightarrow$ (i) : Let $x \notin F$ where F is regular closed in X. Since $\{x\} \cap F = \emptyset$, by (vii) there exist $U \in \psi O(X)$ and an $r\psi g$-open set W such that $x \in U$, $F \subseteq W$ and $U \cap W = \emptyset$. Then $F \subseteq \psi$-$int(W) = V$ (say) $\in \psi O(X)$ (by Theorem 2.9) and hence $V \cap U = \emptyset$. □

Definition 3.4. A topological space X is said to be almost δp-regular [6] if for each regular closed set A of X and each point $x \in X \setminus A$, there exist disjoint δ-preopen sets U and V such that $x \in U$ and $A \subseteq V$.

Almost δp-regular spaces were introduced and studied in [6].

Remark 3.5. If in a topological space (X, τ) we take $\psi = intclint$, then an almost ψ-regular space reduces to an almost regular space [19].

Definition 3.6. Let ψ be an operation on a topological space (X, τ). Then (X, τ) is said to be almost ψ-normal if for each closed set A and each regular closed set B of X such that $A \cap B = \emptyset$, there exist disjoint ψ-open sets U and V such that $A \subseteq U$ and $B \subseteq V$.

Remark 3.7. Let ψ be an operation on a space (X, τ). Then an almost ψ-normal space reduces to an almost normal [26] (resp. almost p-normal [15, 21], almost δp-normal [5, 6], almost α-normal) space if one takes ψ to be int (resp. $intcl$, $intcl_\delta$, $intclint$).

We note that in a topological space (X, τ) with an operation ψ on X, A is $g\psi$-open iff $F \subseteq \psi$-$int(A)$ whenever $F \subseteq A$ and F is closed.

Theorem 3.8. *Let ψ be an operation on a topological space (X, τ). Then the following statements are equivalent :*
(i) X is almost ψ-normal.
(ii) For each closed set A and regular closed set B of X such that $A \cap B = \emptyset$, there exist disjoint $g\psi$-open sets U and V of X such that $A \subseteq U$ and $B \subseteq V$.
(iii) For each closed set A and each regular open set B containing A, there exists a $g\psi$-open set V of X such that $A \subseteq V \subseteq \psi$-$cl(V) \subseteq B$.
(iv) For each rg-closed set A and each regular open set B containing A, there exists a $g\psi$-open set V of X such that $clA \subseteq V \subseteq \psi$-$cl(V) \subseteq B$.
(v) For each rg-closed set A and each regular open set B containing A, there exists a ψ-open set V of X such that $clA \subseteq V \subseteq \psi$-$cl(V) \subseteq B$.
(vi) For each g-closed set A and each regular open set B containing A, there exists a ψ-open set V such that $cl(A) \subseteq V \subseteq \psi$-$cl(V) \subseteq B$.

(vii) For each g-closed set A and each regular open set B containing A, there exists a gψ-open set V such that $cl(A) \subseteq V \subseteq \psi\text{-}cl(V) \subseteq B$.

Proof. (i) ⇒ (ii) : Obvious by Remark 2.2.

(ii) ⇒ (iii) : Let A be a closed set and B be a regular open set containing A. Then $A \cap (X \setminus B) = \emptyset$, where A is closed and $X \setminus B$ is regular closed. So by (ii) there exist disjoint $g\psi$-open sets V and W such that $A \subseteq V$ and $X \setminus B \subseteq W$. Thus by Remark 2.2 and Theorem 2.9, $X \setminus B \subseteq \psi\text{-}int(W)$ and $V \cap \psi\text{-}int(W) = \emptyset$. Hence $\psi\text{-}cl(V) \cap \psi\text{-}int(W) = \emptyset$ and hence $A \subseteq V \subseteq \psi\text{-}cl(V) \subseteq X \setminus \psi\text{-}int(W) \subseteq B$.

(iii) ⇒ (iv) : Let A be rg-closed and B be a regular open set containing A. Then $clA \subseteq B$. The rest follows from (iii).

(iv) ⇒ (v) : This follows from (iv) and the fact that a subset A is $g\psi$-open iff $F \subseteq \psi\text{-}int(A)$ whenever $F \subseteq A$ and F is closed.

(v) ⇒ (vi) : Follows from (v) and the fact that every g-closed set is an rg-closed set.

(vi) ⇒ (vii) : Trivial by Remark 2.2.

(vii) ⇒ (i) : Let A be any closed set and B be a regular closed set such that $A \cap B = \emptyset$. Then $X \setminus B$ is a regular open set containing A where A is g-closed (as every closed set is g-closed). So there exists a $g\psi$-open set G of X such that $clA \subseteq G \subseteq \psi\text{-}cl(G) \subseteq X \setminus B$. Put $U = \psi\text{-}int(G)$ and $V = X \setminus \psi\text{-}cl(G)$. Then U and V are two disjoint ψ-open subsets of X such that $clA \subseteq U$ (as G is $g\psi$-open), i.e., $A \subseteq U$ and $B \subseteq V$. Hence X is almost ψ-normal. □

Remark 3.9. If in a topological space (X, τ) if we take $\psi = intclint$, then an almost ψ-normal space reduces to an almost normal space that follows from the next theorem.

Theorem 3.10. *A topological space* (X, τ) *is almost normal if and only if it is almost* α*-normal.*

Proof. One part of the Theorem is obvious as $\tau \subseteq \tau_\alpha$. We shall only show that if X is almost α-normal then it is normal. Let A be a closed set and B be a regular closed set such that $A \cap B = \emptyset$. Then by α-normality of X, there exist two disjoint α-open sets G and H such that $A \subseteq G$ and $B \subseteq H$. Let $U = int(cl(int(G)))$ and $V = int(cl(int(H)))$. Then U and V are two open subsets of X such that $A \subseteq U$, $B \subseteq V$ and $U \cap V = \emptyset$. Thus X is almost normal. □

Definition 3.11. Let ψ be an operation on a topological space (X, τ). Then (X, τ) is said to be mildly ψ-normal if for every pair of disjoint regular closed sets A and B of X, there exist two disjoint ψ-open sets U and V such that $A \subseteq U$ and $B \subseteq V$.

Remark 3.12. Let ψ be an operation on a space (X, τ). Then a mildly ψ-normal space reduces to a mildly normal [27, 17] (resp. mildly p-normal [15, 21], mildly δp-normal [5, 6], mildly α-normal) if one takes ψ to be int (resp. $intcl$, $intcl_\delta$, $intclint$).

Theorem 3.13. *Let ψ be an operation on a topological space (X, τ). Then the following are equivalent :*
(i) X is mildly ψ-normal.
(ii) For any disjoint $L, K \in RC(X)$, there exist $g\psi$-open sets U and V such that $L \subseteq U$ and $K \subseteq V$.
(iii) For $L, K \in RC(X)$ with $L \cap K = \emptyset$, there exist disjoint $r\psi g$-open sets U and V such that $L \subseteq U$ and $K \subseteq V$.
(iv) For any $L \in RC(X)$ and any $V \in RO(X)$ with $L \subseteq V$, there exists an $r\psi g$-open set U of X such that $L \subseteq U \subseteq \psi\text{-}cl(U) \subseteq V$.
(v) For any $L \in RC(X)$ and any $V \in RO(X)$ with $L \subseteq V$, there exists a ψ-open set U of X such that $L \subseteq U \subseteq \psi\text{-}cl(U) \subseteq V$.

Proof. (i) $\Rightarrow$ (ii) : Follows from Remark 2.2.

(ii) $\Rightarrow$ (iii) : Follows from Remark 2.2.

(iii) $\Rightarrow$ (iv) : Let $L \in RC(X)$ and $V \in RO(X)$ be such that $L \subseteq V$. Then by (iii) there exist disjoint $r\psi g$-open sets U and W such that $L \subseteq U$ and $X \setminus V \subseteq W$. Thus by Theorem 2.9, $X \setminus V \subseteq \psi\text{-}int(W)$ and $U \cap \psi\text{-}int(W) = \emptyset$. So $\psi\text{-}cl(U) \cap \psi\text{-}int(W) = \emptyset$ and hence $L \subseteq U \subseteq \psi\text{-}cl(U) \subseteq X \setminus \psi\text{-}int(W) \subseteq V$.

(iv) $\Rightarrow$ (v) : Let $L \in RC(X)$ and $V \in RO(X)$ be such that $L \subseteq V$. Thus by (iv) there exists an $r\psi g$-open set G of X such that $L \subseteq G \subseteq \psi\text{-}cl(G) \subseteq V$. Since $L \in RC(X)$, by Theorem 2.9, $L \subseteq \psi\text{-}int(G) = U$ (say). Hence $U \in \psi O(X)$ and $L \subseteq U \subseteq \psi\text{-}cl(U) \subseteq \psi\text{-}cl(G) \subseteq V$.

(v) $\Rightarrow$ (i) : Let $L, K \in RC(X)$ be such that $L \cap K = \emptyset$. Then $X \setminus K \in RO(X)$ with $L \subseteq X \setminus K$. Thus by (v) there exists a ψ-open set U of X such that $L \subseteq U \subseteq \psi\text{-}cl(U) \subseteq X \setminus K$. Put $V = X \setminus \psi\text{-}cl(U)$. Then U and V are disjoint ψ-open sets such that $L \subseteq U$ and $K \subseteq V$. □

By the similar arguments as shown in Theorem 3.10 we have

Remark 3.14. In a topological space (X, τ) if we take $\psi = intclint$, then a mildly ψ-normal space and mildly normal space are identical.

Acknowledgements. The authors are thankful to the referees for some comments to improve the paper.

References

[1] Abd El-Monsef, M. E., El-Deeb, S. N., Mahmoud, R. A. β-open sets and β-continuous mappings, *Bull. Fac. Sci. Assiut Univ.*, 12 (1983) 77–90.

[2] Al-Omari, A., Md Noorani, M. S., Regular generalized ω-closed sets, *Inter. J. Math. Math. Sci.*, (2007) 1–11.

[3] Császár, *Á.*, Generalized open sets, *Acta. Math. Hungar.*, 75 (1-2) (1997) 65–87.

[4] Ekici, E., On γ-normal spaces, *Bull. Math. Soc. Sci. Math. Roumanie*, Tome 50(98)(3)(2007), 259–272.

[5] Ekici, E., Noiri, T., On a generalization of normal, almost normal and mildly normal spaces-I, *Math. Moravica*, 10 (2006) 9–20.

[6] Ekici, E., Noiri, T., On a generalization of normal, almost normal and mildly normal spaces II, *Filomat*, 20(2)(2006) 67–80.

[7] Gnanambal, Y., On generalized regular closed sets in topological spaces, *Indian J. Pure Appl. Math.*, 28(3)(1997) 351–360.

[8] Hdeib, H. Z., ω-closed mappings, *Rev. Colombiana Mat.*, 16(1-2)(1982) 65–78.

[9] Küçü, M., Zorlutuna, $\dot{I}$., A unified theory for weak separation properties, *Internat. J. Math and Math. Sci.* 24(11)(2000) 765–772.

[10] Levine, N., Semi-open sets and semi-continuity in topological spaces, *Amer. Math. Monthly*, 70(1963) 36–41.

[11] Levine, N., Generalized closed sets in topology, *Rend. Circ. Mat. Palermo*, 19(2)(1970) 89–96.

[12] Maki, H., Devi R., Balachandran, K., Generalized α-closed sets in topology, *Bull. Fukuoka Univ. Ed.*, Part-III 42(1993) 13–21.

[13] Malghan, S. R., Navalagi, G. B., Almost p-regular, p-completely regular and almost p-completely regular spaces, *Bull. Math. Soc. Sci. Math. R. S. Roumanie*, 34(82) (1990) 417–326.

[14] Mashhour, A. S., El-Monsef, Abd., El. Deeb, S. N., On pre-continuous and weak pre-continuous mappings, *Math. Phys. Soc. Egypt*, 53(1982) 47–53.

[15] Navalagi, G. B., p-normal, almost p-normal and midly p-normal spaces, *Topology Atlas*, (Preprint).

[16] Njastad, O. On some classes of nearly open sets, *Pacific J. Math.*, 15(3)(1979) 961–970.

[17] Noiri, T., Mildly normal spaces and some functions, *Kyungpook Math. J.*, 36(1996) 183–190.

[18] Noiri, T. Almost p-regular spaces and some functions, *Acta Math. Hungar.*, 79(1998) 207–216.

[19] Noiri, T. Almost αg-closed functions and separation axioms, *Acta Math. Hungar.*, 82(3)(1999) 193–205.

[20] Palaniappan, N., Chandrasekhara Rao, K., Regular generalized closed sets, *Kyungpook Math. J.*, 33(2)(1993) 211–219.

[21] Park, J. H. Almost p-normal, mildly p-normal spaces and some functions, *Chaos Solitons and Fractals*, 18(2003) 267–274.

[22] Park, J. K., Park, J. H., Mildly generalized closed sets, almost normal and mildly normal spaces, *Chaos Solitons and Fratals*, 20(2004) 1103–1111.

[23] Raychaudhuri, S., Mukherjee, M. N., On δ-almost continuity and δ-preopen sets, *Bull. Inst. Math. Acad. Sinica*, 21(1993), 357–366.

[24] Roy, B., Sen, R., Noiri, T., Separation axioms on topological spaces - a unified version, (communicated).

[25] Singal, M. K., Arya, S. P. On almost-regular spaces, *Glasnik Mat.*, 4(24) (1969), 89–99.

[26] Singal, M. K., Arya, S. P., Almost normal, almost completely regular spaces, *Glasnik Mat.*, 5(25)(1970), 141–152.

[27] Singal, M. K., Singal, A. R., Mildly normal spaces, *Kyungpook Math. J.*, 13(1973), 27–31.

[28] Veličko, N. V., H-closed topological spaces, *Mat. Sb.*, 70(1966), 98–112.

Annales Mathematicae et Informaticae
***40** (2012) pp. 105–112*
http://ami.ektf.hu

The signed Roman domatic number of a graph

Seyed Mahmoud Sheikholeslami[a], Lutz Volkmann[b]

[a]Research Group of Processing and Communication
and
Department of Mathematics
Azarbaijan Shahid Madani University
Tabriz, I.R. Iran
s.m.sheikholeslami@azaruniv.edu

[b]Lehrstuhl II für Mathematik
RWTH Aachen University
52056 Aachen, Germany
volkm@math2.rwth-aachen.de

Submitted March 30, 2012 — Accepted November 28, 2012

Abstract

A *signed Roman dominating function* (SRDF) on a graph G is a function $f\colon V(G) \to \{-1, 1, 2\}$ such that $\sum_{u \in N[v]} f(u) \geq 1$ for every $v \in V(G)$, and every vertex $u \in V(G)$ for which $f(u) = -1$ is adjacent to at least one vertex w for which $f(w) = 2$. A set $\{f_1, f_2, \ldots, f_d\}$ of distinct signed Roman dominating functions on G with the property that $\sum_{i=1}^{d} f_i(v) \leq 1$ for each $v \in V(G)$, is called a *signed Roman dominating family* (of functions) on G. The maximum number of functions in a signed Roman dominating family on G is the *signed Roman domatic number* of G, denoted by $d_{sR}(G)$. In this paper we initiate the study of signed Roman domatic number in graphs and we present some sharp bounds for $d_{sR}(G)$. In addition, we determine the signed Roman domatic number of some graphs.

Keywords: signed Roman dominating function, signed Roman domination number, signed Roman domatic number

MSC: 05C69

1. Introduction

In this paper, G is a simple graph with vertex set $V = V(G)$ and edge set $E = E(G)$. The order $|V|$ of G is denoted by $n = n(G)$. For every vertex $v \in V$, the *open neighborhood* $N(v)$ is the set $\{u \in V(G) \mid uv \in E(G)\}$ and the *closed neighborhood* of v is the set $N[v] = N(v) \cup \{v\}$. The *degree* of a vertex $v \in V$ is $d(v) = |N(v)|$. The *minimum* and *maximum degree* of a graph G are denoted by $\delta = \delta(G)$ and $\Delta = \Delta(G)$, respectively. A graph G is *k-regular* if $d(v) = k$ for each vertex v of G. The *open neighborhood* of a set $S \subseteq V$ is the set $N(S) = \cup_{v \in S} N(v)$, and the *closed neighborhood* of S is the set $N[S] = N(S) \cup S$. A *tree* is an acyclic connected graph. The complement of a graph G is denoted by $\overline{G}$. A *cactus graph* is a connected graph in which any two cycles have at most one vertex in common. We write K_n for the *complete graph* of order n and C_n for a *cycle* of length n.

A *Roman dominating function* (RDF) on a graph $G = (V, E)$ is defined in [6, 8] as a function $f\colon V \to \{0, 1, 2\}$ satisfying the condition that every vertex v for which $f(v) = 0$ is adjacent to at least one vertex u for which $f(u) = 2$. The *weight* of an RDF f is the value $\omega(f) = \sum_{v \in V} f(v)$. The *Roman domination number* of a graph G, denoted by $\gamma_R(G)$, equals the minimum weight of an RDF on G. The Roman domination number has been studied by several authors (see for example [2, 3, 4]). A set $\{f_1, f_2, \ldots, f_d\}$ of distinct Roman dominating functions on G with the property that $\sum_{i=1}^{d} f_i(v) \leq 2$ for each $v \in V(G)$, is called a *Roman dominating family* (of functions) on G. The maximum number of functions in a Roman dominating family (RD family) on G is the *Roman domatic number* of G, denoted by $d_R(G)$. The Roman domatic number was introduced by Sheikholeslami and Volkmann [7] and has been studied by several authors (see for example [5]).

A *signed Roman dominating function* (SRDF) on a graph $G = (V, E)$ is defined in [1] as a function $f\colon V \to \{-1, 1, 2\}$ such that $\sum_{u \in N[v]} f(u) \geq 1$ for each $v \in V(G)$, and such that every vertex $u \in V(G)$ for which $f(u) = -1$ is adjacent to at least one vertex w for which $f(w) = 2$. The *weight* of an SRDF f is the value $\omega(f) = \sum_{v \in V} f(v)$. The *signed Roman domination number* of a graph G, denoted by $\gamma_{sR}(G)$, equals the minimum weight of an SRDF on G. A $\gamma_{sR}(G)$*-function* is a signed Roman dominating function of G with weight $\gamma_{sR}(G)$. A signed Roman dominating function $f\colon V \to \{-1, 1, 2\}$ can be represented by the ordered partition (V_{-1}, V_1, V_2) (or (V_{-1}^f, V_1^f, V_2^f) to refer f) of V, where $V_i = \{v \in V \mid f(v) = i\}$. In this representation, its weight is $\omega(f) = |V_1| + 2|V_2| - |V_{-1}|$.

A set $\{f_1, f_2, \ldots, f_d\}$ of distinct signed Roman dominating functions on G with the property that $\sum_{i=1}^{d} f_i(v) \leq 1$ for each $v \in V(G)$, is called a *signed Roman dominating family* (of functions) on G. The maximum number of functions in a signed Roman dominating family (SRD family) on G is the *signed Roman domatic number* of G, denoted by $d_{sR}(G)$. The signed Roman domatic number is well-defined and

$$d_{sR}(G) \geq 1 \tag{1.1}$$

for all graphs G since the set consisting of the SRDF with constant value 1 forms an SRD family on G. If $G_1, G_2, \ldots, G_k$ are the connected components of G, then

obviously $d_{sR}(G) = \min\{d_{sR}(G_i) \mid 1 \leq i \leq k\}$. Hence, we only consider connected graphs.

Our purpose in this paper is to initiate the study of signed Roman domatic number in graphs. We first study basic properties and bounds for the signed Roman domatic number of a graph. In addition, we determine the signed Roman domatic number of some classes of graphs.

We make use of the following results in this paper.

Proposition A ([1]). *If K_n is the complete graph of order $n \geq 1$, then $\gamma_{sR}(K_n) = 1$, unless $n = 3$ in which case $\gamma_{sR}(K_n) = 2$.*

Proposition B ([1]).

1. *For $n \geq 3$, $\gamma_{sR}(C_n) = \lceil \frac{2n}{3} \rceil$,*
2. *For $n \geq 2$, $\gamma_{sR}(P_n) = \lfloor \frac{2n}{3} \rfloor$.*

Proposition C ([1]). *Let G be a graph of order $n \geq 1$. Then $\gamma_{sR}(G) = n$ if and only if $G = \overline{K_n}$.*

Proposition D ([1]). *If G is a δ-regular graph of order n with $\delta \geq 1$, then $\gamma_{sR}(G) \geq \lceil n/(\delta + 1) \rceil$.*

2. Properties of the signed Roman domatic number

In this section we present basic properties of $d_{sR}(G)$ and sharp bounds on the signed Roman domatic number of a graph.

Theorem 2.1. *For every graph G,*

$$d_{sR}(G) \leq \delta(G) + 1.$$

Moreover, if $d_{sR}(G) = \delta(G) + 1$, then for each SRD family $\{f_1, f_2, \ldots, f_d\}$ on G with $d = d_{sR}(G)$ and each vertex v of minimum degree, $\sum_{u \in N[v]} f_i(u) = 1$ for each function f_i and $\sum_{i=1}^{d} f_i(u) = 1$ for all $u \in N[v]$.

Proof. If $d_{sR}(G) = 1$, the result is immediate. Let now $d_{sR}(G) \geq 2$ and let $\{f_1, f_2, \ldots, f_d\}$ be an SRD family on G such that $d = d_{sR}(G)$. Assume that v is a vertex of minimum degree $\delta(G)$. We have

$$d \leq \sum_{i=1}^{d} \sum_{u \in N[v]} f_i(u) = \sum_{u \in N[v]} \sum_{i=1}^{d} f_i(u) \leq \sum_{u \in N[v]} 1 = \delta(G) + 1.$$

Thus $d_{sR}(G) \leq \delta(G) + 1$.

If $d_{sR}(G) = \delta + 1$, then the two inequalities occurring in the proof become equalities. Hence for the SRD family $\{f_1, f_2, \ldots, f_d\}$ on G and for each vertex v of minimum degree, $\sum_{u \in N[v]} f_i(u) = 1$ for each function f_i and $\sum_{i=1}^{d} f_i(u) = 1$ for all $u \in N[v]$. □

The next results are immediate consequences of Proposition C and Theorem 2.1.

Corollary 2.2. *For $n \geq 1$, $d_{sR}(\overline{K_n}) = 1$.*

Corollary 2.3. *For any tree T of $n \geq 3$, $d_{sR}(T) \leq 2$. The bound is sharp for a double star obtained from two vertex disjoint stars $K_{1,3}$ by connecting their centers.*

Problem 2.4. *Characterize all trees T for which $d_{sR}(T) = 2$.*

Corollary 2.5. *For $n \geq 2$, $d_{sR}(K_{1,n}) = 1$.*

Proof. It follows from Theorem 2.1 that $d_{sR}(K_{1,n}) \leq 2$. Suppose to the contrary that $d_{sR}(K_{1,n}) = 2$ and assume that $\{f_1, f_2\}$ is an SRD family on $K_{1,n}$. Let $V(K_{1,n}) = \{v, u_1, \ldots, u_n\}$ and $E(K_{1,n}) = \{vu_i \mid 1 \leq i \leq n\}$. Theorem 2.1 implies that $f_1(v) + f_2(v) = 1$. Since $f_j(x) \in \{-1, 1, 2\}$ for each j and each vertex x, we deduce that $f_1(v) = -1$ and $f_2(v) = 2$ or $f_1(v) = 2$ and $f_2(v) = -1$. Assume, without loss of generality, that $f_1(v) = -1$ and $f_2(v) = 2$. By Theorem 2.1, we must have $f_2(u_i) + f_2(v) = 1$ for each $1 \leq i \leq n$ and therefore $f_2(u_i) = -1$ for each $1 \leq i \leq n$. Since $n \geq 2$, we obtain the contradiction $1 \leq \sum_{x \in N[v]} f_2(x) = 2 - n \leq 0$. Thus $d_{sR}(K_{1,n}) = 1$. □

Theorem 2.6. *If G is a graph of order n, then*

$$\gamma_{sR}(G) \cdot d_{sR}(G) \leq n.$$

Moreover, if $\gamma_{sR}(G) \cdot d_{sR}(G) = n$, then for each SRD family $\{f_1, f_2, \ldots, f_d\}$ on G with $d = d_{sR}(G)$, each function f_i is a $\gamma_{sR}(G)$-function and $\sum_{i=1}^{d} f_i(v) = 1$ for all $v \in V$.

Proof. Let $\{f_1, f_2, \ldots, f_d\}$ be an SRD family on G such that $d = d_{sR}(G)$ and let $v \in V$. Then

$$d \cdot \gamma_{sR}(G) = \sum_{i=1}^{d} \gamma_{sR}(G) \leq \sum_{i=1}^{d} \sum_{v \in V} f_i(v) = \sum_{v \in V} \sum_{i=1}^{d} f_i(v) \leq \sum_{v \in V} 1 = n.$$

If $\gamma_{sR}(G) \cdot d_{sR}(G) = n$, then the two inequalities occurring in the proof become equalities. Hence for the SRD family $\{f_1, f_2, \ldots, f_d\}$ on G and for each i, $\sum_{v \in V} f_i(v) = \gamma_{sR}(G)$. Thus each function f_i is a $\gamma_{sR}(G)$-function, and $\sum_{i=1}^{d} f_i(v) = 1$ for all $v \in V$. □

The next two results are immediate consequences of Propositions B, C and Theorem 2.6.

Corollary 2.7. *For $n \geq 3$, $d_{sR}(C_n) = 1$.*

Corollary 2.8. *Let G be a graph of order $n \geq 1$. Then $\gamma_{sR}(G) = n$ and $d_{sR}(G) = 1$ if and only if $G = \overline{K_n}$.*

Corollary 2.9. *For $n \geq 1$, $d_{sR}(P_n) = 1$, unless $n = 2$ in which case $d_{sR}(P_n) = 2$.*

Proof. If follows from Proposition B and Theorem 2.6 that $d_{sR}(P_n) = 1$, unless $n = 2$ or $n = 4$. Let $P_n := v_1v_2 \dots v_n$. First let $n = 2$. Define the functions $f_i \colon \{v_1, v_2\} \to \{-1, 1, 2\}$ for $i = 1, 2$ by $f_1(v_1) = 2, f_1(v_2) = -1, f_2(v_1) = -1$ and $f_2(v_2) = 2$. Obviously f_1 and f_2 are signed Roman dominating functions of P_2 and $\{f_1, f_2\}$ is a signed Roman dominating family on P_2. Hence $d_{sR}(P_2) \geq 2$. Therefore $d_{sR}(P_2) = 2$ by Theorem 2.1.

Now let $n = 4$. It follows from Theorem 2.1 that $d_{sR}(P_4) \leq 2$. Suppose to the contrary that $d_{sR}(P_4) = 2$ and let $\{f_1, f_2\}$ be a signed Roman dominating family on P_4. By Theorem 2.1, we must have $f_i(v_1) + f_i(v_2) = 1$ for $i = 1, 2$ and $f_1(v_2) + f_2(v_2) = 1$. By Theorem 2.1, $f_1(v_1) + f_2(v_1) = 1$. Similarly, we have $f_1(v_4) + f_2(v_4) = 1$. Thus $f_1(v_i) + f_2(v_i) = 1$ for $1 \leq i \leq 4$. Since $f_1(v_i), f_2(v_i) \in \{-1, 1, 2\}$ and $f_1(v_i) + f_2(v_i) = 1$, we deduce that $f_1(v_i) = -1, f_2(v_i) = 2$ or $f_1(v_i) = 2, f_2(v_i) = -1$ for $1 \leq i \leq 4$. Assume, without loss of generality, that $f_1(v_1) = 2$ and $f_2(v_1) = -1$. Since $f_i(v_1) + f_i(v_2) = 1$ for $i = 1, 2$, we must have $f_1(v_2) = -1$ and $f_2(v_2) = 2$. If $f_1(v_3) = -1$, then we have $\sum_{u \in N[v_2]} f_1(u) \leq 0$ which is a contradiction. Thus, $f_1(v_3) = 2$ and hence $f_2(v_3) = -1$ which implies that $\sum_{u \in N[v_2]} f_2(u) \leq 0$ which is a contradiction again. Therefore $d_{sR}(P_4) = 1$ and the proof is complete. □

Theorem 2.10. *If K_n is the complete graph of order $n \geq 1$, then $d_{sR}(K_n) = n$, unless $n = 3$ in which case $d_{sR}(K_n) = 1$.*

Proof. If $n = 3$, the the result follows from Proposition A and Theorem 2.6. Now let $n \neq 3$ and let $V(K_n) = \{v_0, v_1, \dots, v_{n-1}\}$ be the vertex set of K_n. Consider two cases.

Case 1. Assume that n is even. Define the functions $f_1, f_2, \dots, f_n$ as follows. $f_1(v_{n-1}) = 2$, $f_1(v_i) = -1$ if $0 \leq i \leq \frac{n-2}{2}$ and $f_1(v_i) = 1$ if $\frac{n}{2} \leq i \leq n-2$, and for $2 \leq j \leq q$ and $0 \leq i \leq n-1$,

$$f_j(v_i) = f_{j-1}(v_{i+j-1}),$$

where the sum is taken modulo n. It is easy to see that f_j is a signed Roman dominating function of K_n of weight 1 and for each $1 \leq j \leq n$ and $\{f_1, f_2, \dots, f_n\}$ is a signed Roman dominating family on K_n. Hence $d_{sR}(K_n) \geq n$. Therefore $d_{sR}(K_n) = n$ by Proposition A and Theorem 2.6.

Case 2. Assume that n is odd. Define the functions $f_1, f_2, \dots, f_n$ as follows. $f_1(v_{n-1}) = f(v_{n-2}) = 2$, $f_1(v_i) = -1$ if $0 \leq i \leq \frac{n-1}{2}$ and $f_1(v_i) = 1$ if $\frac{n+1}{2} \leq i \leq n-3$, and for $2 \leq j \leq q$ and $0 \leq i \leq n-1$,

$$f_j(v_i) = f_{j-1}(v_{i+j-1}),$$

where the sum is taken modulo n. It is easy to see that f_j is a signed Roman dominating function of K_n of weight 1, for each $1 \leq j \leq n$ and $\{f_1, f_2, \dots, f_n\}$ is a signed Roman dominating family on K_n. Hence $d_{sR}(K_n) \geq n$. Therefore $d_{sR}(K_n) = n$ by Proposition A and Theorem 2.6. □

For some regular graphs we will improve the upper bound given in Theorem 2.1.

Theorem 2.11. *Let G be a δ-regular graph of order n such that $\delta \geq 1$. If $n \not\equiv 0 \pmod{(\delta+1)}$, then $d_{sR}(G) \leq \delta$.*

Proof. Since $n \not\equiv 0 \pmod{(\delta+1)}$, we deduce that $n = p(\delta+1)+r$ with integers $p \geq 1$ and $1 \leq r \leq \delta$. Let $\{f_1, f_2, \ldots, f_d\}$ be an SRD family on G such that $d = d_{sR}(G)$. It follows that

$$\sum_{i=1}^{d} \omega(f_i) = \sum_{i=1}^{d} \sum_{v \in V} f_i(v) = \sum_{v \in V} \sum_{i=1}^{d} f_i(v) \leq \sum_{v \in V} 1 = n.$$

Proposition D implies $\omega(f_i) \geq \gamma_{sR}(G) \geq p+1$ for each $i \in \{1, 2, \ldots, d\}$. If we suppose to the contrary that $d \geq \delta + 1$, then the above inequality chain leads to the contradiction

$$n \geq \sum_{i=1}^{d} \omega(f_i) \geq d(p+1) \geq (\delta+1)(p+1) = p(\delta+1) + \delta + 1 > n.$$

Thus $d \leq \delta$, and the proof is complete. □

Theorem 2.10 demonstrates that Theorem 2.11 is not valid in general when $n \equiv 0 \pmod{(\delta+1)}$.

Theorem 2.12. *If G is a graph of order $n \geq 1$, then*

$$\gamma_{sR}(G) + d_{sR}(G) \leq n + 1 \tag{2.1}$$

with equality if and only if $G \simeq \overline{K_n}$ or $G \simeq K_n$ $(n \neq 3)$.

Proof. It follows from Theorem 2.6 that

$$\gamma_{sR}(G) + d_{sR}(G) \leq \frac{n}{d_{sR}(G)} + d_{sR}(G). \tag{2.2}$$

According to Theorem 2.1, we have $1 \leq d_{sR}(G) \leq n$. Using these bounds, and the fact that the function $g(x) = x + n/x$ is decreasing for $1 \leq x \leq \sqrt{n}$ and increasing for $\sqrt{n} \leq x \leq n$, the last inequality leads to the desired bound immediately.

If $G \simeq K_n$ $(n \neq 3)$ then it follows from Proposition A and Theorem 2.10 that $\gamma_{sR}(G) + d_{sR}(G) = n + 1$. If $G \simeq \overline{K_n}$, then it follows from Proposition C and Corollary 2.2 that $\gamma_{sR}(G) + d_{sR}(G) = n + 1$.

Conversely, let equality hold in (2.1). It follows from (2.2) that

$$n + 1 = \gamma_{sR}(G) + d_{sR}(G) \leq \frac{n}{d_{sR}(G)} + d_{sR}(G) \leq n + 1,$$

which implies that $\gamma_{sR}(G) = \frac{n}{d_{sR}(G)}$ and $d_{sR}(G) = 1$ or $d_{sR}(G) = n$. If $d_{sR}(G) = n$, then $\delta(G) = n - 1$ by Theorem 2.1 and hence G is a complete graph K_n. Since also $\gamma_{sR}(G) = 1$, we deduce that $n \neq 3$ and hence $G \simeq K_n$ $(n \neq 3)$. If $d_{sR}(G) = 1$, then $\gamma_{sR}(G) = n$, and it follows from Proposition C that $G \simeq \overline{K_n}$. This completes the proof. □

As an application of Theorems 2.1 and 2.11, we will prove the following Nordhaus-Gaddum type result.

Theorem 2.13. *For every graph G of order n,*

$$d_{sR}(G) + d_{sR}(\overline{G}) \leq n + 1. \tag{2.3}$$

Furthermore, $d_{sR}(G) + d_{sR}(\overline{G}) = n + 1$ if and only if $n \neq 3$ and $G \simeq K_n$ or $G \simeq \overline{K_n}$.

Proof. It follows from Theorem 2.1 that

$$\begin{aligned} d_{sR}(G) + d_{sR}(\overline{G}) &\leq (\delta(G) + 1) + (\delta(\overline{G}) + 1) \\ &= (\delta(G) + 1) + (n - \Delta(G) - 1 + 1) \leq n + 1. \end{aligned}$$

If G is not regular, then $\Delta(G) - \delta(G) \geq 1$, and hence the above inequality chain implies the better bound $d_{sR}(G) + d_{sR}(\overline{G}) \leq n$.

If $n \neq 3$ and $G \simeq K_n$ or $G \simeq \overline{K_n}$, then Corollary 2.2 and Theorem 2.10 lead to $d_{sR}(G) + d_{sR}(\overline{G}) = n + 1$.

Conversely, assume that $d_{sR}(G) + d_{sR}(\overline{G}) = n + 1$. Then G is δ-regular and thus $\overline{G}$ is $(n - \delta - 1)$-regular. If $\delta = 0$ or $\delta = n - 1$, then $G \simeq K_n$ or $G \simeq \overline{K_n}$, and we obtain the desired result.

Next assume that $1 \leq \delta \leq n - 2$ and $1 \leq \delta(\overline{G}) = n - \delta - 1 \leq n - 2$. We assume, without loss of generality, that $\delta \leq (n-1)/2$. If $n \not\equiv 0 \pmod{(\delta+1)}$, then it follows from Theorems 2.1 and 2.11 that

$$\begin{aligned} d_{sR}(G) + d_{sR}(\overline{G}) &\leq \delta(G) + (\delta(\overline{G}) + 1) \\ &= \delta(G) + (n - \delta(G) - 1 + 1) = n, \end{aligned}$$

a contradiction. Next assume that $n \equiv 0 \pmod{(\delta + 1)}$. Then $n = p(\delta + 1)$ with an integer $p \geq 2$. If $n \not\equiv 0 \pmod{(n - \delta)}$, then it follows from Theorems 2.1 and 2.11 that

$$\begin{aligned} d_{sR}(G) + d_{sR}(\overline{G}) &\leq (\delta(G) + 1) + \delta(\overline{G}) \\ &= \delta(G) + 1 + (n - \delta(G) - 1) = n, \end{aligned}$$

a contradiction. Therefore assume that $n \equiv 0 \pmod{(n - \delta)}$. Then $n = q(n - \delta)$ with an integer $q \geq 2$. Since $\delta \leq (n - 1)/2$, this leads to the contradiction

$$n = q(n - \delta) \geq \left(n - \frac{n-1}{2}\right) = \frac{q(n+1)}{2} \geq n + 1,$$

and the proof is complete. □

The next result is a generalization of Corollary 2.3.

Theorem 2.14. *If G is a connected cactus graph, then $d_{sR}(G) \leq 2$.*

Proof. Let $d = d_{sR}(G)$. If $\delta(G) \leq 1$, then Theorem 2.1 implies the desired bound $d \leq 2$ immediately.

It remains the case that $\delta(G) = 2$. If G is a cycle, then the result follows from Corollary 2.7. Otherwise, the cactus graph G contains a cycle $v_1v_2 \ldots v_tv_1$ as an end block with exactly one cut vertex, say v_1. Applying Theorem 2.1, we see that $d \leq 3$. Suppose to the contrary that $d = 3$. Let $\{f_1, f_2, f_3\}$ be a signed Roman dominating family on G.

Claim. If $f_i(v_j) = 2$ for $1 \leq i \leq 3$ and $2 \leq j \leq t$, then $d \leq 2$.

Proof of claim. Assume, without loss of generality, that $f_1(v_2) = 2$. Because of $f_1(v_2) + f_2(v_2) + f_3(v_2) \leq 1$, we deduce that $f_2(v_2) = f_3(v_2) = -1$. Since f_i is a signed Roman dominating function, we see that $f_i(v_1) = 2$ or $f_i(v_3) = 2$ for $2 \leq i \leq 3$. Assume, without loss of generality, that $f_2(v_1) = 2$. It follows as above that $f_1(v_1) = f_3(v_1) = -1$. Hence we obtain the contradiction $1 \leq \sum_{x \in N[v_2]} f_3(x) = -2 + f_3(v_3) \leq 0$, and the claim is proved.

Thus we assume that $f_i(v_j) \leq 1$ for $1 \leq i \leq 3$ and $2 \leq j \leq t$. If $t \geq 4$, then we conclude that $f_i(v_3) = 1$ for $1 \leq i \leq 3$, a contradiction to $f_1(v_3) + f_2(v_3) + f_3(v_3) \leq 1$. Finally, assume that $t = 3$. If $f_i(v_1) \leq 1$ for $1 \leq i \leq 3$, then $f_i(v_2) = 1$ for $1 \leq i \leq 3$, a contradiction. Now assume, without loss of generality, that $f_1(v_1) = 2$. This implies that $f_2(v_1) = f_3(v_1) = -1$ and therefore $f_2(v_2) = f_3(v_2) = f_2(v_3) = f_3(v_3) = 1$. This leads to $f_1(v_2) = f_1(v_3) = -1$. Thus we obtain the contradiction $1 \leq \sum_{x \in N[v_2]} f_1(x) = f_1(v_1) + f_1(v_2) + f_1(v_3) = 0$, and the proof is complete. □

References

[1] Ahangar, H. A., Henning, M. A., Zhao, Y., Löwenstein, C., Samodivkin, V., Signed Roman domination in graphs, *J. Comb. Optim.*, (to appear).

[2] Chambers, E. W., Kinnersley, B., Prince, N., West, D. B., Extremal problems for Roman domination, *SIAM J. Discrete Math.*, 23 (2009) 1575–1586.

[3] Cockayne, E. J., Dreyer Jr., P. M., Hedetniemi, S. M., Hedetniemi, S. T., On Roman domination in graphs, *Discrete Math.*, 278 (2004) 11–22.

[4] Favaron, O., Karami, H., Sheikholeslami, S. M., On the Roman domination number in graphs, *Discrete Math.*, 309 (2009) 3447–3451.

[5] Kazemi, A. P., Sheikholeslami, S. M., Volkmann, L., Roman (k, k)-domatic number of a graph, *Ann. Math. Inform.*, 38 (2011) 45–57.

[6] ReVelle, C. S., Rosing, K. E., Defendens imperium romanum: a classical problem in military strategy, *Amer. Math. Monthly*, 107 (2000) 585–594.

[7] Sheikholeslami, S. M., Volkmann, L., Roman domatic number of a graph, *Appl. Math. Letters*, 23 (2010) 1295–1300.

[8] Stewart, I., Defend the Roman Empire, *Sci. Amer.*, 281 (1999) 136–139.

[9] West, D.B, Introduction to Graph Theory, Prentice-Hall, Inc, 2000.

Annales Mathematicae et Informaticae
40 *(2012) pp. 113–123*
http://ami.ektf.hu

The rank and Hanna Neumann property of some submonoids of a free monoid

Shubh Narayan Singh, K. V. Krishna

Department of Mathematics
Indian Institute of Technology Guwahati
Guwahati, India
{shubh,kvk}@iitg.ac.in

Submitted October 27, 2011 — Accepted January 28, 2012

Abstract

This work aims at further investigations on the work of Giambruno and Restivo [5] to find the rank of the intersection of two finitely generated submonoids of a free monoid. In this connection, we obtain the rank of a finitely generated submonoid of a free monoid that is accepted by semi-flower automaton with two bpi's. Further, when the product automaton of two deterministic semi-flower automata with a unique bpi is semi-flower with two bpi's, we obtain a sufficient condition on the product automaton in order to satisfy the Hanna Neumann property.

Keywords: Finitely generated monoids, semi-flower automata, rank, Hanna Neumann property.

MSC: 68Q70, 68Q45, 20M35.

1. Introduction

In [6], Howson obtained an upper bound for the rank of intersection of two finitely generated subgroups of a free group in terms of the individual ranks of subgroups. Thus, it is known that the intersection of two finitely generated subgroups of a free group is finitely generated. In 1956, Hanna Neumann proved that if H and K are finite rank subgroups of a free group, then

$$\widetilde{rk}(H \cap K) \leq 2\widetilde{rk}(H)\widetilde{rk}(K),$$

where $\widetilde{rk}(N) = \max(0, rk(N) - 1)$ for a subgroup N of rank $rk(N)$. This is an improvement on Howson's bound. Further, Neumann conjectured that

$$\widetilde{rk}(H \cap K) \leq \widetilde{rk}(H)\widetilde{rk}(K), \qquad (\star)$$

which is known as Hanna Neumann conjecture [10]. In 1990, Walter Neumann proposed a stronger form of the conjecture called strengthened Hanna Neumann conjecture (SHNC) [11]. Meakin and Weil proved SHNC for the class of positively generated subgroups of a free group [7]. The conjecture has recently been settled by Mineyev (cf. [8, 9]) and announced independently by Friedman (cf. [2, 3]).

In contrast, it is not true that the intersection of two finitely generated submonoids of a free monoid is finitely generated. Since Tilson's work [12] in 1972, through the work of Giambruno and Restivo [5] in 2008, there are several contributions in the literature on the topic. Using automata-theoretic approach, Giambruno and Restivo have investigated an upper bound for the rank of the intersection of two submonoids of special type in a free monoid. In fact, for the special case, they have proved the Hanna Neumann property. Two submonoids H and K are said to satisfy *Hanna Neumann property* (in short, HNP), if H and K satisfy the inequality $(\star)$.

This work extends the work of Giambruno and Restivo [5] to another special class of submonoids. Here, we find the rank of a finitely generated submonoid of a free monoid that is accepted by semi-flower automaton with two bpi's. Further, we obtain a condition to extend HNP for the submonoids of a free monoid which satisfy the following condition C.

> Two submonoids of a free monoid are said to satisfy the condition C, if they are accepted by deterministic semi-flower automata, each with a unique bpi and their product automaton is semi-flower with two bpi's.

Rest of the paper is organized as follows. In Section 2, we present some preliminary concepts and results that are used in this work. Section 3 is dedicated to present the main results of the paper. We conclude the paper in Section 4.

2. Preliminaries

In this section, we present some background material from [1, 4, 5]. We try to confine to the terminology and notations given there so that one may refer to [1, 4, 5] for those notions that are not presented here, if any.

Let A be a finite set called an *alphabet* with its elements as *letters*. The free monoid over A is denoted by A^* and ε denotes the empty word – the identity element of A^*. It is known that every submonoid of A^* is generated by a unique minimal set of generators. Thus, the *rank* of a submonoid H, denoted by $rk(H)$, of A^* is defined as the cardinality of the minimal set of generators X of H, i.e. $rk(H) = |X|$. Further, the *reduced rank* of a submonoid H of A^* is defined as $\max(0, rk(H) - 1)$ and it is denoted by $\widetilde{rk}(H)$.

An *automaton* $\mathcal{A}$ over an alphabet A is a quadruple $(Q, I, T, \mathcal{F})$, where Q is a finite set called the set of *states*, I and T are subsets of Q called the sets of *initial* and *final* states, respectively, and $\mathcal{F} \subseteq Q \times A \times Q$ called the set of *transitions*. Clearly, by denoting the states as vertices/nodes and the transitions as labeled arcs, an automaton can be represented by a digraph in which initial and final states shall be distinguished appropriately.

A *path* in $\mathcal{A}$ is a finite sequence of consecutive arcs in its digraph. For $q_i \in Q$ $(0 \leq i \leq k)$ and $a_j \in A$ $(1 \leq j \leq k)$, let

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \xrightarrow{a_3} \cdots \xrightarrow{a_{k-1}} q_{k-1} \xrightarrow{a_k} q_k$$

be a path P in an automaton $\mathcal{A}$ that is starting at q_0 and ending at q_k. In this case, we write $i(P) = q_0$ and $f(P) = q_k$. The word $a_1 \cdots a_k \in A^*$ is the *label of the path* P. For each state $q \in Q$, the *null path* is a path from q to q labeled by ε.

A path in $\mathcal{A}$ is called *simple* if all the states on the path are distinct. A path that starts and ends at the same state is called as a *cycle*, if it is not a null path. A cycle with all its intermediate states are distinct is called a *simple cycle*. A cycle that starts and ends in a state q is called simple in q, if no intermediate state is equal to q. Other notions related to paths, viz. subpath, prefix and suffix, can be interpreted with their literal meaning or one may refer to [5].

Let $\mathcal{A}$ be an automaton. The *language accepted/recognized by* $\mathcal{A}$, denoted by $L(\mathcal{A})$, is the set of words that are labels of paths from an initial state to a final state. A state $q \in Q$ is *accessible* (respectively, *coaccessible*) if there is a path from an initial state to q (respectively, a path from q to a final state). An automaton is called *trim* if all the states of the automaton are accessible and coaccessible. An automaton $\mathcal{A} = (Q, I, T, \mathcal{F})$ is *deterministic* if it has a unique initial state, i.e. $|I| = 1$, and there is at most one transition defined for a state and a letter.

An automaton is called a *semi-flower automaton* if it is trim with a unique initial state that is equal to a unique final state such that all the cycles visit the unique initial-final state.

If an automaton $\mathcal{A} = (Q, I, T, \mathcal{F})$ is semi-flower, we denote the initial-final state by 1. In which case, we simply write $\mathcal{A} = (Q, 1, 1, \mathcal{F})$. Further, let us denote by $C_{\mathcal{A}}$ the set of cycles that are simple in 1 and by $Y_{\mathcal{A}}$ the set of their labels.

Now, in the following we state the correspondence between semi-flower automata and finitely generated submonoids of a free monoid.

Theorem 2.1 ([5]). *If $\mathcal{A}$ is a semi-flower automaton over an alphabet A, then $Y_{\mathcal{A}}$ is finite and $\mathcal{A}$ recognizes the submonoid generated by $Y_{\mathcal{A}}$ in A^*. Moreover, if $\mathcal{A}$ is deterministic, then $Y_{\mathcal{A}}$ is the minimal set of generators of the submonoid recognized by $\mathcal{A}$.*

In addition to the above result, given a finitely generated submonoid H of the free monoid A^*, one can easily construct a semi-flower automaton $\mathcal{A}$ such that $L(\mathcal{A}) = H$. Here, to construct $\mathcal{A}$, one may choose a initial-final state and connect a petal to the initial-final state that corresponds to each word of a (finite) generating set of H.

With this basic information, we now present the two results of Giambruno and Restivo which will be generalized/extended in the present paper.

Theorem 2.2 ([5]). *If $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ is a semi-flower automaton with a unique bpi, then*

$$rk(L(\mathcal{A})) \leq |\mathcal{F}| - |Q| + 1.$$

Moreover, if $\mathcal{A}$ is deterministic, then

$$rk(L(\mathcal{A})) = |\mathcal{F}| - |Q| + 1.$$

Here, a state q of an automaton is called a *branch point going in*, in short *bpi*, if the indegree of q (i.e. the number of arcs coming into q) is at least 2.

Theorem 2.3 ([5]). *If H and K are the submonoids accepted by deterministic semi-flower automata $\mathcal{A}_H$ and $\mathcal{A}_K$, respectively, each with a unique bpi such that $\mathcal{A}_H \times \mathcal{A}_K$ is a semi-flower automaton with a unique bpi, then*

$$\widetilde{rk}(H \cap K) \leq \widetilde{rk}(H)\widetilde{rk}(K).$$

Here, for automata $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ and $\mathcal{A}' = (Q', 1', 1', \mathcal{F}')$ both over an alphabet A, $\mathcal{A} \times \mathcal{A}'$ is the *product automaton* $(Q \times Q', (1, 1'), (1, 1'), \widetilde{\mathcal{F}})$ over the alphabet A such that

$$((p, p'), a, (q, q')) \in \widetilde{\mathcal{F}} \iff (p, a, q) \in \mathcal{F} \text{ and } (p', a, q') \in \mathcal{F}'$$

for all $p, q \in Q$, $p', q' \in Q'$ and $a \in A$.

Notice that if $\mathcal{A}$ and $\mathcal{A}'$ are deterministic then so is $\mathcal{A} \times \mathcal{A}'$. But if $\mathcal{A}$ and $\mathcal{A}'$ are trim, then $\mathcal{A} \times \mathcal{A}'$ need not be trim. However, by considering only those states which are accessible and coaccessible, we can make the product automaton $\mathcal{A} \times \mathcal{A}'$ trim. This process does not alter the language accepted by $\mathcal{A} \times \mathcal{A}'$. In fact, we have

$$L(\mathcal{A} \times \mathcal{A}') = L(\mathcal{A}) \cap L(\mathcal{A}').$$

Hence, if we state a product automaton $\mathcal{A} \times \mathcal{A}'$ is semi-flower, we assume that the trim part of $\mathcal{A} \times \mathcal{A}'$, without any further explanation.

In the hypothesis of Theorem 2.3, if the product automaton has more than one bpi, then it is not true that H and K satisfy HNP. This has been shown through certain examples in [4, 5]. In the present work, first we observe that HNP fails if the product automaton has two bpi's. We demonstrate this in Example 3.7. Then we proceed to investigate on the conditions to achieve HNP in case the product automaton has two bpi's.

We would require the following supplementary results from [5] in our main results. Instead of reworking the details, we simply state in the required form. In these results, let the automata be over an alphabet A of cardinality n; and for an automaton $\mathcal{A} = (Q, I, T, \mathcal{F})$ and $i \geq 0$

$$BPO_i(\mathcal{A}) = \{q \in Q \mid \text{out degree of } q = i\}.$$

Proposition 2.4. *If $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ is a deterministic semi-flower automaton over A, then*

$$|\mathcal{F}| - |Q| = \sum_{i=2}^{n} |BPO_i(\mathcal{A})|(i-1).$$

Proposition 2.5. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two deterministic automata over A. If $c_i = |BPO_i(\mathcal{A}_1)|$ and $d_i = |BPO_i(\mathcal{A}_2)|$, for each $i = 1, \ldots, n$, then*

$$|BPO_t(\mathcal{A}_1 \times \mathcal{A}_2)| \leq \sum_{t \leq r,s \leq n} c_r d_s.$$

Proposition 2.6. *Let $\langle c_1, \ldots, c_n \rangle$ and $\langle d_1, \ldots, d_n \rangle$ be two finite sequences of natural numbers; then*

$$\sum_{t=2}^{n}(t-1)\left(\sum_{t \leq r \leq n} c_r \sum_{t \leq s \leq n} d_s\right) \leq \left(\sum_{i=2}^{n}(i-1)c_i\right)\left(\sum_{j=2}^{n}(j-1)d_j\right).$$

3. Main Results

In this section we present two results. First we obtain the rank of a finitely generated submonoid of a free monoid, if it is accepted by a semi-flower automaton with two bpi's. This generalizes the result of Giambruno and Restivo for semi-flower automata with a unique bpi. Then we proceed to obtain HNP for the submonoids of a free monoid that satisfy the condition C.

We begin with introducing a concise notation for a semi-flower automaton in which only the initial-final state, bpi's and the respective paths between them will be represented along with their labels. We call this as *bpi's and paths representation*, in short *BPR*, of an automaton. For example, the BPR of the semi-flower automaton given in Figure 1 is shown in Figure 2.

The following lemma is useful for obtaining the rank of a semi-flower automaton with two bpi's.

Lemma 3.1. *If $\mathcal{A}$ is a semi-flower automaton with two bpi's, say p and q such that the distance from q to the 1 is not more than that of p, then*

(i) *there is a unique simple path from q to 1, and*

(ii) *every cycle in $\mathcal{A}$ visits q.*

Proof. (i) If $q = 1$, then we are done. If not, by the choice of q, the initial-final state 1 is not a bpi. Moreover, since q is coaccessible, there is a path from q to 1. Now suppose there are two different simple paths P_1 and P_2 with labels u and v, respectively, from q to 1. Note that P_1 and P_2 are not one suffix of the other. Let w be the label of longest suffix path P' which is in common between the paths P_1 and P_2. As 1 is not a bpi, $w \neq \varepsilon$. But then $i(P')$ will be a bpi different from q.

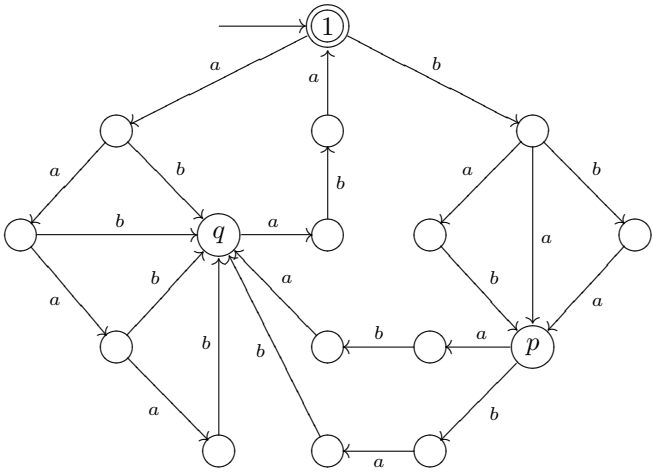


Figure 1: A semi-flower automaton

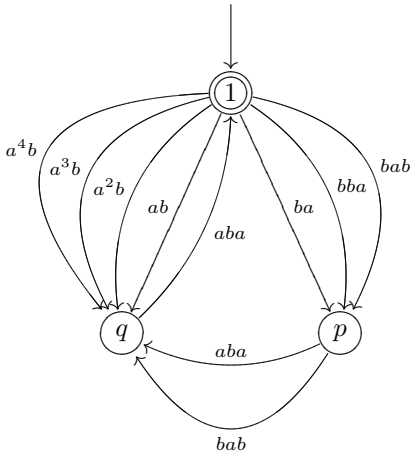


Figure 2: BPR of the semi-flower automaton given in Figure 1

This a contradiction to the choice of q. Thus, there is a unique simple path from q to 1.

(ii) Since every cycle in $\mathcal{A}$ passes through 1, if $q = 1$, then we are done. If not, 1 is not a bpi. Now suppose there is a cycle that is not visiting q. Then the cycle contributes one to the indegree of the state 1. Also, from above (i), there is a path from q to 1. This implies that the state 1 is a bpi; a contradiction. □

Since every cycle that visits p also visits q, it follows that

Corollary 3.2. *If p and q are distinguishable, the distance from p to 1 is more than that of q.*

Notation 3.3. In what follows, if a semi-flower automaton has two bpi's, say p and q, then we consider that the distance from q to 1 is not more than that of p. Moreover, we assume that the indegree of p is m and the indegree of q is $(l + k)$, where k is the number of edges ending at q that are not in any of the paths from p to q. With this information, the BPR of such an automaton will be as shown in Figure 3.

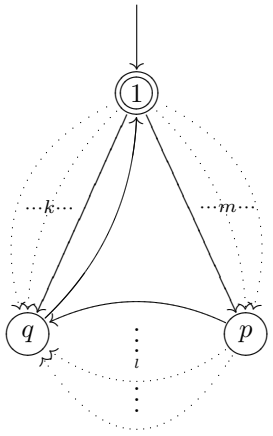


Figure 3: BPR of a semi-flower automaton with two bpi's

Now we are ready to present our first result of the paper.

Theorem 3.4. *If $\mathcal{A}$ is a semi-flower automaton with two bpi's p and q, then*

$$rk(L(\mathcal{A})) \leq ml + k.$$

Moreover, if $\mathcal{A}$ is deterministic, then

$$rk(L(\mathcal{A})) = ml + k.$$

Proof. As the number of simple cycles passing through the initial-final state 1 (i.e. in $C_{\mathcal{A}}$) gives us an upper bound for the rank $rk(L(\mathcal{A}))$, we count these cycles using indegree of p and q. The number of cycles in $C_{\mathcal{A}}$ that are passing through q but not p is k. Also, as each path entering the state p will split into l number of paths and enter in the state q, we have ml number of cycles in $C_{\mathcal{A}}$ that are passing through p. Thus, the total number of cycles in $C_{\mathcal{A}}$ is $ml+k$. Hence, as $L(\mathcal{A})$ is the submonoid generated by $Y_{\mathcal{A}}$, we have

$$rk(L(\mathcal{A})) \leq |Y_{\mathcal{A}}| = |C_{\mathcal{A}}| = ml + k.$$

If $\mathcal{A}$ is deterministic, then by Theorem 2.1, we have

$$rk(L(\mathcal{A})) = ml + k.$$

□

In a semi-flower automaton with two indistinguishable bpi's, i.e. with a unique bpi, we have the following corollary.

Corollary 3.5. *Theorem 2.2 follows.*

Proof. In the hypothesis of Theorem 3.4, if $p = q$ (i.e. p and q are indistinguishable), then $\mathcal{A}$ has a unique bpi. In which case, $l = 0$ and consequently, $rk(L(\mathcal{A}))$ is less than or equal to the indegree k of the unique bpi. And in case $\mathcal{A}$ is deterministic, $rk(L(\mathcal{A})) = k$. Now the number of transitions $|\mathcal{F}|$ in $\mathcal{A}$ can be counted by the number of arcs entering all the states of $\mathcal{A}$. As $\mathcal{A}$ is trim, every state of $\mathcal{A}$ has an arc into it. Further, since $\mathcal{A}$ has a unique bpi, except the bpi, all other states have indegree one. Thus, we have

$$|\mathcal{F}| = |Q| - 1 + k,$$

so that

$$rk(L(\mathcal{A})) \leq |\mathcal{F}| - |Q| + 1.$$

Moreover, if $\mathcal{A}$ is deterministic, then the equality holds. □

Remark 3.6. Theorem 3.4 generalizes Theorem 2.2.

Before proceeding to our second result, it is appropriate to note the following example.

Example 3.7. Consider the submonoids $H = \{aa, aba, ba, bb\}^*$ and $K = \{a, bab\}^*$ of the free monoid $\{a, b\}^*$. We give the automata $\mathcal{A}_H$ and $\mathcal{A}_K$ which accept H and K, respectively, in Figure 4. Note that $\mathcal{A}_H$ and $\mathcal{A}_K$ are deterministic semi-flower automata, each with unique bpi. The (trim form of) product automaton $\mathcal{A}_H \times \mathcal{A}_K$ is shown in Figure 5. Clearly, $\mathcal{A}_H \times \mathcal{A}_K$ is semi-flower with two bpi's, viz. $(1, 1')$ and $(1, 3')$ and hence $rk(H \cap K) = 5$. Whereas, $rk(H) = 4$ and $rk(K) = 2$. Thus, H and K do not satisfy HNP, i.e.

$$\widetilde{rk}(H \cap K) > \widetilde{rk}(H)\widetilde{rk}(K).$$

The following lemma is useful in proving our second result of the paper.

Lemma 3.8. *If $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ is a semi-flower automaton with two bpi's p and q, then*

$$rk(L(\mathcal{A})) - (m-1)(l-1) \leq |\mathcal{F}| - |Q| + 1.$$

Moreover, if $\mathcal{A}$ is deterministic, then the equality holds.

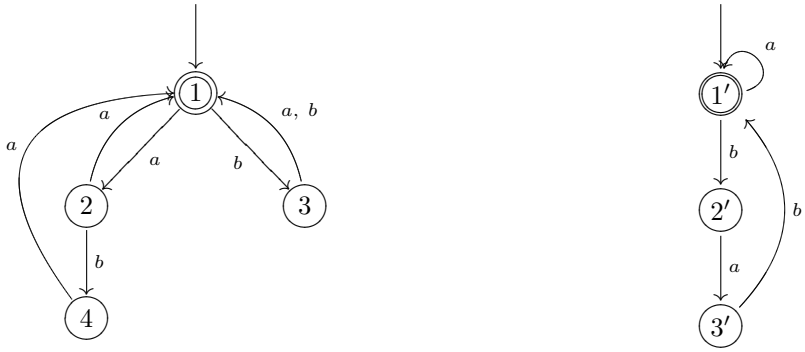


Figure 4: $\mathcal{A}_H$ (in the left) and $\mathcal{A}_K$ (in the right) of Example 3.7

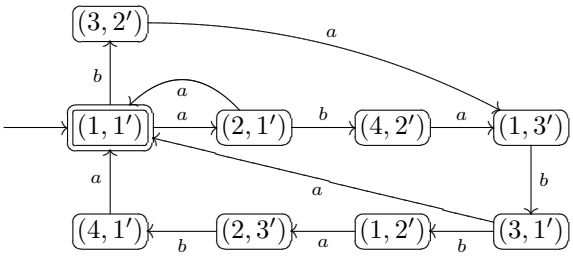


Figure 5: $\mathcal{A}_H \times \mathcal{A}_K$ of Example 3.7

Proof. Since the number of transitions $|\mathcal{F}|$ of $\mathcal{A}$ is the total indegree (i.e. the sum of indegrees of all the states) of the digraph of $\mathcal{A}$, we have

$$|\mathcal{F}| = m + l + k + |Q| - 2.$$

Consequently,

$$\begin{aligned} & |\mathcal{F}| - |Q| + 1 &&= m + l + k - 1 \\ \implies\ & |\mathcal{F}| - |Q| + 1 &&= (ml + k) - (ml - m - l + 1) \\ \implies\ & |\mathcal{F}| - |Q| + 1 &&= (ml + k) - (m - 1)(l - 1). \end{aligned}$$

Hence, by Theorem 3.4, $|\mathcal{F}| - |Q| + 1 \geq rk(L(\mathcal{A})) - (m-1)(l-1)$. □

Now, by Proposition 2.4, we have the following corollary.

Corollary 3.9. *If $\mathcal{A}$ is a deterministic semi-flower automaton with two bpi's p and q, then $rk(L(\mathcal{A})) = (m-1)(l-1) + \sum_{t=2}^{n} |BPO_t(\mathcal{A})|(t-1) + 1$.*

Theorem 3.10. *If H and K are the submonoids accepted by deterministic semi-flower automata $\mathcal{A}_H$ and $\mathcal{A}_K$, respectively, each with a unique bpi such that the product automaton $\mathcal{A}_H \times \mathcal{A}_K$ is semi-flower with two bpi's p and q, then*

$$\widetilde{rk}(H \cap K) \leq \widetilde{rk}(H)\widetilde{rk}(K) + (m-1)(l-1).$$

Proof. Note that

$$\begin{aligned}\widetilde{rk}(H \cap K) &= rk(L(\mathcal{A}_H \times \mathcal{A}_K)) - 1 \\ &= (m-1)(l-1) + \sum_{t=2}^{n} |BPO_t(\mathcal{A}_H \times \mathcal{A}_K)|(t-1) \text{ by Corollary 3.9} \\ &\leq (m-1)(l-1) + \sum_{t=2}^{n}(t-1)\left(\sum_{t\leq r,s\leq n} c_r d_s\right) \text{ by Proposition 2.5,}\end{aligned}$$

where $c_r = |BPO_r(\mathcal{A}_H)|$ and $d_s = |BPO_s(\mathcal{A}_K)|$. Consequently, by Proposition 2.6

$$\begin{aligned}\widetilde{rk}(H \cap K) &\leq (m-1)(l-1) + \left(\sum_{i=2}^{n}(i-1)c_i\right)\left(\sum_{j=2}^{n}(j-1)d_j\right) \\ &= (m-1)(l-1) + \widetilde{rk}(H)\widetilde{rk}(K) \text{ by Theorem 2.2 and Proposition 2.4.}\end{aligned}$$

Hence the result. □

Corollary 3.11. *In addition to the hypothesis of Theorem 3.10, if there is a unique path from p to q in $\mathcal{A}_H \times \mathcal{A}_K$, then*

$$\widetilde{rk}(H \cap K) \leq \widetilde{rk}(H)\widetilde{rk}(K).$$

4. Conclusion

In this work we have obtained the rank of a finitely generated submonoid of a free monoid that is accepted by a semi-flower automaton with two bpi's. This generalizes the rank result (cf. Theorem 2.2) for semi-flower automata with unique bpi by Giambruno and Restivo [5]. In fact, the present proof of Theorem 2.2 is shorter and elegant than that of the original proof by Giambruno and Restivo. In [5], Giambruno and Restivo obtained HNP for submonoids of a free monoid that are accepted by deterministic semi-flower automata, each with a unique bpi such that their product automaton is semi-flower with a unique bpi. Further, by keeping the former automata as they are, if the latter automaton has more than one bpi, they provided examples which fail to satisfy HNP. In the present work, we give an example which fails to satisfy HNP when the product automaton has two bpi's. In case the product automaton has two bpi's, we reported a sufficient condition to obtain HNP. The techniques introduced in this work shall give a scope to one in extending our work to a general scenario.

References

[1] Berstel, J., Perrin, D., *Theory of codes*, Academic Press (1985).

[2] Friedman, J., Linear algebra and the Hanna Neumann Conjecture, *Preprint.*

[3] Friedman, J., Sheaves on graphs, their homological invariants, and a proof of the Hanna Neumann conjecture, *Preprint.* arXiv:1105.0129v2.

[4] Giambruno, L., *Automata-theoretic methods in free monoids and free groups*, PhD thesis, Universitŕ degli Studi di Palermo, Palermo, Italy (2007).

[5] Giambruno, L., Restivo, A., An automata-theoretic approach to the study of the intersection of two submonoids of a free monoid. *RAIRO - Theor. Inform. Appl.*, 42 (2008) 503–524.

[6] Howson A. G., On the intersection of finitely generated free groups, *J. London Math. Soc.*, 29 (1954) 428–434.

[7] Meakin, J., Weil, P., Subgroups of free groups: a contribution to the Hanna Neumann conjecture, *Geom. Dedicata*, 94 (2002) 33–43.

[8] Mineyev, I., Groups, graphs, and the Hanna Neumann Conjecture, *Preprint.*

[9] Mineyev, I., Submultiplicativity and the Hanna Neumann Conjecture, *Ann. of Math.*, 175 (2012) 393–414.

[10] Neumann, H., On the intersection of finitely generated free groups, *Publ. Math. Debrecen*, 4 (1956) 186–189.

[11] Neumann, W. D., On intersections of finitely generated subgroups of free groups, *Lecture Notes in Math.*, 1456 (1990) 161–170.

[12] Tilson, B., The intersection of free submonoids of a free monoid is free, *Semigroup Forum*, 4 (1972) 345–350.

Annales Mathematicae et Informaticae
***40** (2012) pp. 125–134*
http://ami.ektf.hu

Parallel algorithm for determining the "small solutions" of Thue equations*

Gergő Szekrényesi

University of Miskolc, Department of Applied Mathematics
gergosz5@gmail.com

Submitted July 17, 2011 — Accepted March 2, 2012

Abstract

A typical research field in number theory is determining the solutions of Diophantine equations. One of the earliest topic amongst these are the topics of Thue equations and inequalities. These equations are bivariate homogenous forms, with integer coeffictients.

The aim of the article is to create a parallel program, that can solve arbitrary Thue equations in finite time, and also gives good running times with the classical families. Naturally with the computational packages such as Maple, Magma or Kant a Thue equation can be solved, but in practice they can't handle problems given with relatively large coefficients. The parallel version of the algorithm provides an alternative to this problem.

Another application of the parallel program lies in determining the solutions of simultaneos Pellian equations. According to the work of L. Szalay the solutions of the system of equations can be traced back to the solutions of Thue equations, which can be found with the method given.

Keywords: Thue equations, parallel algorithm, simultaneous Pellian equations

MSC: 11Y40; 11D75

*This research was carried out as part of the TAMOP-4.2.1.B-10/2/KONV-2010-0001 project with support by the European Union, co-financed by the European Social Fund.

1. Basic terms

1.1. Thue equations

Let $F(x, y)$ be a bivariate, homogenous, over $\mathbb{Q}[x, y]$ irreducible polinomial with integer coefficients of at least third degree. Explicitly:

$$F(x, y) = a_n x^n + a_{n-1} x^{(n-1)} y + \ldots + a_0 y^n \in \mathbb{Z}[x, y]$$
$$n \geq 3 \quad \text{and} \quad m \in \mathbb{Z}, m \neq 0.$$

Then the

$$F(x, y) = m \tag{1.1}$$

equation is called Thue equation, of which's $x, y \in \mathbb{Z}$ are sought. Thue [10] proved that, (1.1) has only finitely many solutions for all m. Later, Baker [1] showed, that there exists an effectively computable constant based only on m and the coefficients of F which serves as an upper bound for the solutions.

Thue equations can be classified into parametric families. The first parametric family was introduced by Thue himself. In 1990 Thomas was the first to investigate families of fixed degrees, specifically a family of degree three. Furthermore several authors investigated certain families E.g. A. Pethő, Gaál, Lettl, Heuberger, Togbé and Ziegler.

There are a number of algorithms which provide the solutions of Thue equations. Unfortunately Baker's bound is too large, and does not make it possible to handle the problem by simple enumeration. For a similar problem Baker and Davenport created a method to reduce this bound drastically. A. Pethő and Schulenberg used a method of continued fraction reduction, and later the LLL-algorithm.

Among these algorithms a prominent one is the method described by A. Pethő, which provides the "small solutions" of such equations. An advantage of this algorithm is that it can be relatively easily implemented (it is based on the continued fraction algorithm). Furthermore it can benefit from many parallelization techniques. The most important definitions and theorems like the approximation theorems and their uses regarding the solutions of Thue equations are part of the article.

A method to find all the solutions of (1.1) for $m = 1$ and arbitrary n also exists. Finally Bilu and Hanrot [3] gave a much more efficient continued fraction method to reduce the bound, and they were able to solve equations up to the degree of 1000.

1.2. Simultaneous Pell equations

Simultaneous Pell equations are defined as follows:

$$a_1 x^2 + b_1 y^2 = c_1, \tag{1.2}$$

$$a_2x^2 + b_2z^2 = c_2, \tag{1.3}$$

where the x, y and z are nonnegative integers, and the coefficients satisfy the following conditions:

$$a_1b_1 < 0, \quad a_2b_2 < 0, \quad c_1c_2 \neq 0, \quad a_1c_2 - a_2c_1 \neq 0.$$

There is a method due to L. Szalay [9] with which one can trace back the solutions of such equations to the solving finitely many Thue equations.

1.3. Balancing numbers

Definition 1.1. An n positive integer is called a balancing number, if

$$1 + 2 + \cdots + (n-1) = (n+1) + (n+2) + \cdots + (n+r)$$

holds, for a suitable positive integer r.

Liptai published some results [5, 6] on special kinds of balancing numbers. For example a balancing number is called a Fibonacci balancing number if it is a Fibonacci number as well. He also proved that such balancing numbers are solutions of the equation

$$x^2 - 5y^2 = \pm 4.$$

If we are to find such common numbers in the Fibonacci sequence and the balancing numbers we have to provide an additional equation regarding the current balancing number, and solve the system. Another special type of balancing numbers are (a, b)-type balancing numbers.

Definition 1.2. Let $a, b \in \mathbb{N}$. The $an + b$ natural number is called an (a, b)-type balancing number if the following equation holds for a suitable $r \in \mathbb{N}$:

$$(a+b) + (2a+b) + \cdots + (a(n-1)+b) = (a(n+1)+b) + \cdots + (a(n+r)+b)$$

Kovács, Liptai and Olajos published about (a, b)-type balancing numbers, see [4]. Consider the following simultaneous Pellian equation system:

$$x^2 - 5y^2 = 4 \tag{1.4}$$

$$8x^2 - z^2 = 167. \tag{1.5}$$

The solutions of this system provide the common elements of the Fibonacci series and a (a, b)-type balancing numbers.

2. Theory of Thue equations

2.1. Continued Fraction Algorithm

The continued fractions play an important role in finding the solutions of Thue equations and inequalities.

Input: $\alpha \in \mathbb{R}, \alpha \neq 0, A \in \mathbb{Z}$
Output: $a_0 \in \mathbb{Z}, a_1, \ldots a_n \in \mathbb{N}, p_n, q_n \in \mathbb{Z}$
1. $i \leftarrow 0,\ \alpha_0 \leftarrow \alpha,\ p_{-2} \leftarrow 0,\ p_{-1} \leftarrow 1,\ q_{-2} \leftarrow 1,\ q_{-1} \leftarrow 0$
2. DO
3. $\quad a_i \leftarrow \lfloor \alpha_i \rfloor$
4. $\quad p_i \leftarrow a_i p_{i-1} + p_{i-2}$
5. $\quad q_i \leftarrow a_i q_{i-1} + q_{i-2}$
6. $\quad$ IF $\alpha_i - a_i = 0$ THEN STOP
7. $\quad i \leftarrow i + 1$
8. $\quad \alpha_i \leftarrow 1/(\alpha_{i-1} - a_{i-1})$
9. WHILE $q_i \leq A$

The algorithm provides the continued fraction expansion of the α real number, the values a_i, and the convergents, the $\frac{p_i}{q_i}$. It is easy to see, that the algorithm stops at the 6th step if and only if α is rational. Furthermore, for the convergents the following inequality holds for every $n \geq 0$:

$$\frac{1}{(a_{n+1}+2)q_n^2} \leq \left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2} \tag{2.1}$$

If α is irrational, then (2.1) implies $\lim_{n\to\infty} \frac{p_n}{q_n} = \alpha$.

The right side of (2.1) characterizes the continued fractions. The following lemma is due to Legendre

Lemma 2.1. (Legendre) *Let $\alpha \in \mathbb{R}$ and $x, y \in \mathbb{Z}, y \neq 0$. If*

$$\left|\alpha - \frac{x}{y}\right| < \frac{1}{2y^2}$$

holds, then $\dfrac{x}{y}$ is a convergent to α.

2.2. Fast algorithm for finding the "small solutions"

Consider the following Thue inequality

$$|F(x,y)| \leq m. \tag{2.2}$$

The aim is to find the solutions of (2.2) where $|y| < C$ and $(x, y) = 1$ (coprime integers).

Remark 2.2. Using the coprime solutions it is easy to determine all the other solutions:
Let $|F(x,y)| \leq m$ and $(x,y) = d : x = x_0 d, y = y_0 d, (x_0, y_0) = 1$. The method computes (x_0, y_0) for which $|F(x_0, y_0)| \leq m$ and $|F(x_0 d, y_0 d)| \leq m \Rightarrow d^n |F(x_0, y_0)| \leq m \Rightarrow d^n \leq \dfrac{m}{|F(x_0, y_0)|}$, thus the possible values of d can be checked individually.

Let α be the root of $F(x,1) = 0$. Its conjugates: $\alpha^{(1)}, \ldots, \alpha^{(n)}$. To find the solutions we need to calculate some constants and bounds from the roots of $F(x,1) = 0$, the $\alpha^{(i)}$ numbers. These are as follows:

$$c_1 = \frac{2|m|^{\frac{1}{n}}}{|\alpha^{(j)} - \alpha^{(i)}|},$$
$$c_2 = \frac{|m|2^{n-1}}{\prod\limits_{j \neq i} |\alpha^{(j)} - \alpha^{(i)}|}$$

If $\alpha^{(i)} = a + bI$ is a complex number, then

$$c_3 = \left(\frac{c_2}{|b|}\right)^{\frac{1}{n}}.$$

Furthermore

$$c_4 = (2c_2)^{\frac{1}{n-2}}.$$

Lemma 2.3. *(Thue) The Thue equation $F(x,y) = m$ has only finitely many $(x,y) \in \mathbb{Z}^2$ solutions.*

Remark 2.4. This lemma does not give any method on how to find these solutions. Experience shows, that Thue equations usually only have few number and small in absolute value solutions.

The following lemma gives described by A. Pethő provides a method to find the solutions of a given Thue equation or inequality, up to a certain prescribed bound.

Lemma 2.5. *(Pethő) For all (x,y) solutions of the* (2.2) *Thue-inequality either $|y| < \max(c_1, c_3, c_4)$ or $\frac{x}{y}$ is a convergent in the continued fraction expansion of one of the real conjugates of the α roots of $F(x,1) = 0$, for which*

$$|y| < (c_2(A+2))^{\frac{1}{n-2}}, \text{ where}$$

$A = \max\limits_{j \leq j_0} a_j$, *where j_0 is the index for which the denominator of the convergent of $\alpha^{(i)}$ $q_{j_0} < C$, $q_{j_0+1} > C$ holds.*

The main point of the algorithm is, that if an $(x,y) \in \mathbb{Z}^2$ pair is a solution of (2.2) then $\frac{x}{y}$ is a good approximation of the roots of $F(x,1)$. We calculate the "medium large" solutions with the simple continued fraction expansion of the roots of $F(x,1)$, then we find the smaller solutions by elementary bounds and equations.

Remark 2.6.

- The constant C is is taken to be considerably large, for example 10^{500}.
- The lemma provides the $|y| < C$ solutions fast in this case as well, but does not give any proof that no other solutions exist in the range $|y| > C$.

- In the case of $C = 10^{500}$ the $\alpha^{(i)}$ number must be at least 1000 digits precise for the continued fraction algorithm to work accurately.
- The value of A is generally small, as the continued fraction digits are small.
- The algorithm provies a way to reduce the C upper bound, by substituting it with $(c_2(A+2))^{\frac{1}{n-2}}$, and reapplying the calculation. This can be done as long as it yields new values for the upper bound, or until we get a value below $\max(c_1, c_3, c_4)$.

3. Main result: the Algorithm

3.1. Steps of the Algorithm

1. Calculate the α_i roots of the polinomial.
2. For all roots perform the followings (parallelly):
   (a) Calculate the constants c_1, c_2, c_3, c_4.
   (b) For only the real α_i-s calculate the continued fraction expansion and the convergents, using the above described algorithm, and reduce the upper bound as much as possible.
   (c) Up to the newly calculated upper bound subsitute the convergents p_k, q_k to see if they satisfy the $|F(p_k, q_k)| \leq m$ inequality.
   (d) If yes, add (p_k, q_k) to the set of solutions.
   (e) In the interval $|y| < \max(c_1, c_3, c_4)$ find the solutions using an exhaustive search.

Remark 3.1. The algorithm can be used to solve equations as well besides inequalites. In this case replace the $\leq$ operator with equality. The case is similar with the absolute value too.

3.2. Parallelization

The implementation of the above described algorithm is rather inefficient in terms of time, especially in the case of equations with large coefficients. Based on the steps described above, the algorithm can be parallellized.

After calculating the roots of $F(x, 1) = 0$, the α_i numbers, it is easy to see that the operations that must be performed with the current α_i root are independent and such a dedicated process can be started for every α_i root. This way the continued fraction algorithm can be run parallelly, reducing the running time of the algorithm. This gives the best speed-up if we have n processors for the n different roots.

For each of the roots the exhaustive search must be performed in the interval $|y| < \max(c_1, c_3, c_4)$. On the other hand if we calculate this interval for every root,

and then calculate their maximum we get the interval where we must perform the search regardless of which root are we processing. Furthermore this search interval can be divided to smaller parts as well, exactly as many parts as many processes we have. By using this method not only we speed-up the algorithm, but eliminate some otherwise redundant calculation.

It is obvious that the data parallellism of the algorithm is beneficial from many aspects: not only the time to find the solutions is reduced, but the strain on the individual processes is much less than in the sequential case. Furthermore the practical implementation of the algorithm can be done relatively easily, as the stages are totally independent and synchronization is only needed at the end of the algorithm, to collect the solutions found by the individual processes.

4. Simultaneous Pell equations

Consider again the system given with the equations (1.2) and (1.3):

$$a_1x^2 + b_1y^2 = c_1,$$
$$a_2x^2 + b_2z^2 = c_2,$$

satisfiying the natural conditions $a_1b_1 < 0$, $a_2b_2 < 0$, $c_1c_2 \neq 0$, $a_1c_2 - a_2c_1 \neq 0$. The algorithm to solve such equations can be derived from a combination of (1.2) and (1.3), which leads to the solution of Thue equations of degree four. These types of equations can be solved for instance with the method described above. Unfortunately the number of Thue equations needed to solve increases as the coefficients, the $a_i, b_i, c_i (i = 1, 2)$ numbers get larger.

Introduce the following equation by combining (1.2) and (1.3):

$$a_3x^2 + b_3y^2 + c_3y^2 = 0, \tag{4.1}$$

where a_3, b_3 and c_3 are non-zero integers. For this equation the following lemma holds.

Lemma 4.1. *Assume that* (x_0, y_0, z_0) *is a solution of* (4.1) *with* $z_0 \neq 0$. *Then every integer* (x, y, z) $z \neq 0$ *solution of* (4.1) *can be parametrized in the following way:*

$$x = \pm\frac{D}{d}(-ax_0s^2 - 2by_0rs + bx_0r^2),$$
$$y = \pm\frac{D}{d}(-ay_0s^2 - 2ax_0rs + by_0r^2),$$
$$z = \pm\frac{D}{d}(-az_0s^2 + bz_0r^2),$$

where r *and* $s > 0$ *are coprime integers,* D *is a non negative integer and* $d \mid 2a^2bcz_0^3$ *is a non negative integer.*

Using this lemma we gain the following equation from the simultaneous Pell equation system:

$$a_1c_y^2(\alpha_{i_1}s^2+\beta_{i_1}sr+\gamma_{i_1}r^2)^2+b_1c_x^2(\alpha_{i_2}s^2+\beta_{i_2}sr+\gamma_{i_2}r^2)^2=c_1c_x^2c_y^2\Big(\frac{d}{D}\Big)^2. \quad (4.2)$$

Expanding this equation we gain Thue equations of degree four. The number of Thue equations we need to solve is exactly the number of divisors of the right hand side of the equation. The solutions of these Thue equations are used to determine the solutions of the Pellian system, using the given parametrization.

Furthermore, this step of the algorithm can be optimized. Only one Thue inequality is needed to be solved instead of a series of Thue equations, as the coefficients of the equations remain the same, only the right hand changes. By this, the right hand side of the inequality has to be chosen so it corresponds to the greatest of the original Thue equations. After calculating the solutions of the inequality the solutions have to be tested to filter out the solutions that provide different right sides than the original ones.

5. Running times

To demonstrate the program we have tested it on several examples. To compare the results we ran the same examples using Maple 12. The examples were run with the following configuration: Intel Pentium 4 3.2GHz dual core processor, 1GB RAM. The speed-up is defined as the quotient of the parallel running time and the sequential running time.

Comparison:

	Maple 12	Own Program	Speed-up
$\lvert f(x,y)\rvert \le 200$	1671 ms	34 ms	49.14
$\lvert g(x,y)\rvert \le 27$	843 ms	82 ms	10.28

$$f(x,y)=x^3+x^2y-2xy^2-y^3,$$

$$g(x,y)=x^6+20000x^5y-50015x^4y^2-\\ -20x^3y^3+50000x^2y^4+\\ +20006xy^5+y^6.$$

Considering another case, the $\left|x^{19}+2y^{19}\right|=2$ Thue equation was solved with both a program written in C programming language using the PARI/GP computational package and the program created in this essay. The results are the following:

	PARI/GP	Own Program	Speed-up
$\left\lvert x^{19}+2y^{19}\right\rvert =2$	11.7 sec	63.3 sec	0.18

The difference is due to the program using the PARI/GP package uses the method of Bilu and Hanrot, thus calculated the set of fundamental units of the algebraic number field, and got better running times, but the order of magnitude is the same.

We tested the program with the following simultaneous Pell equation system as well.

$$x^2 - 5y^2 = \pm 4,$$
$$z^2 - 8x^2 = 1.$$

From this system we gain the

$$\left|s^4 - 30s^3r + 195s^2r^2 - 150sr^3 + 25r^4\right| \leq 96100$$

Thue inequality. The running time of Maple 12 was 573.921 seconds, while our program solved the task in 395 milliseconds, which means a speed-up of about 1453. The program in L. Szalay's article which was written in MAGMA [7] solved the same inequality in about 4 seconds.

5.1. Further Pellian system examples

The following table shows some other simultaneous Pellian equation systems that we have tested and compared with the program created by L. Szalay.

System	Own Program	MAGMA	Speed-up
$-11x^2 + y^2 = 1$ $-56x^2 + z^2 = 1$	104ms	26sec	250
$-7x^2 + y^2 = 2$ $-32x^2 + z^2 = -23$	21sec	40sec	1.9
$-8x^2 + y^2 = 1$ $-2x^2 + 3z^2 = 1$	1.5sec	5sec	3.3
$-5x^2 + y^2 = -20$ $-2x^2 + z^2 = 1$	381ms	3sec	7.8
$x^2 - 2y^2 = -1$ $x^2 - 10z^2 = -9$	113ms	1sec	8.8

The table shows that the program proved to be faster in these cases, however the difference is not always significant.

Consider now again the system belonging to (a, b)-type balancing numbers:

$$x^2 - 5y^2 = 4$$
$$8x^2 - z^2 = 167.$$

From this system the generated Thue inequality is:

$$|27889s^4 + 37408s^3r + 9046s^2r^2 - 972sr^3 + 9r^4| \leq 1.008 \cdot 10^{16}.$$

The Thue inequality has 58 solutions and those provide the sole solution of the original system: (7, 3, 15). The running time in this case was approximately 6 minutes, whereas the MAGMA was unable to solve the system because the right hand side of the inequality causes an overflow in the system.

References

[1] Baker, A., Transcendental Number Theory, *Cambridge*, 1990.

[2] Baker, A., Davenport, H., The equations $3x^2 - 2 = y^2$ and $8x^2 - 7 = z^2$, *Quart. J. Math. Oxford*, 20 (1969), 129–137.

[3] Bilu, Y. and Hanrot, G., Solving Thue equations of high degree, *J. Number Theory*, 60 (1996), 373–390.

[4] Kovács, T., Liptai, K. and Olajos, P. On (a, b) balancing numbers, *Publ. Math. Debrecen*, 77, No. 3-4. (2010) 485–498.

[5] Liptai, K. Fibonacci balancing numbers, *Fibonacci Quart.*, 42 ,No. 4, (2004) 330–340.

[6] Liptai, K. Lucas balancing numbers, *Acta Math. Univ. Ostrav.*, 14, No. 1, (2006) 43–47.

[7] MAGMA Computational Algebra System, Computational Algebra Group School of Mathematics and Statistics, *University of Sydney*, NSW 2006, Australia. `http://magma.maths.usyd.edu.au/magma/`

[8] Pethő, A., Schulenberg, R., Effektives Lösen von Thue Gleichungen, *Publ. Math.* Debrecen (1987)

[9] Szalay, L., On the resolution of simultaneous Pell equations, *Annales Mathematicae et Informaticae*, 34 (2007) 77–87.

[10] Thue, A., Über Annäherungenswerte algrebraischen Zahlen, *J. Reine Angew. Math.*, 135 (1909) 284–305.

Annales Mathematicae et Informaticae
***40** (2012) pp. 135–144*
http://ami.ektf.hu

Two applications of the theorem of Carnot

Zoltán Szilasi

Institute of Mathematics, MTA-DE Research Group *Equations, Functions and Curves*
Hungarian Academy of Sciences and University of Debrecen
szilasi.zoltan@science.unideb.hu

Submitted May 20, 2012 — Accepted December 11, 2012

Abstract

Using the theorem of Carnot we give elementary proofs of two statements of C. Bradley. We prove his conjecture concerning the tangents to an arbitrary conic from the vertices of a triangle. We give a synthetic proof of his theorem concerning the "Cevian conic", and we also give a projective generalization of this result.

Keywords: Carnot theorem; Pascal theorem; Menelaos theorem; barycentric coordinates; Cevian conic.

MSC: 51M04; 51A20; 51N15

1. Preliminaries

Throughout this paper we work in the Euclidean plane and in its projective closure, the real projective plane. By XY we denote the *signed* distance of points X, Y of the Euclidean plane. This means that we suppose that on the line $\overleftrightarrow{XY}$ an orientation is given, and $XY = d(X,Y)$ or $XY = -d(X,Y)$ depending on the direction of the vector $\overrightarrow{XY}$. The *simple ratio* of the collinear points X, Y, Z (where $Y \neq Z$ and $X \neq Y$) is defined by

$$(XYZ) := \frac{XZ}{ZY}$$

and it is independent of the choice of orientation on the line $\overleftrightarrow{XY}$, thus in our formulas we can use the notation XY without mentioning the orientation.

We recall here the most important tools that we use in our paper. The proofs of these theorems can be found in [4].

Theorem 1.1. *Let ABC be an arbitrary triangle in the Euclidean plane, and let A_1, B_1, C_1 be points (different from the vertices) on the sides $\overleftrightarrow{BC}$, $\overleftrightarrow{CA}$, $\overleftrightarrow{AB}$, respectively. Then*

- *(Menelaos) A_1, B_1, C_1 are collinear if and only if*

$$(ABC_1)(BCA_1)(CAB_1) = -1,$$

- *(Ceva) $\overleftrightarrow{AA_1}$, $\overleftrightarrow{BB_1}$, $\overleftrightarrow{CC_1}$ are concurrent if and only if*

$$(ABC_1)(BCA_1)(CAB_1) = 1.$$

Referring to the theorem of Ceva, if P is a point that is not incident to any side of the triangle, we call the lines $\overleftrightarrow{AP}$, $\overleftrightarrow{BP}$, $\overleftrightarrow{CP}$ *Cevians*, and we call the points $\overleftrightarrow{AP} \cap \overleftrightarrow{BC}$, $\overleftrightarrow{BP} \cap \overleftrightarrow{AC}$, $\overleftrightarrow{CP} \cap \overleftrightarrow{AB}$ the *feet of the Cevians through P*.

Now we formulate the most important theorem on projective conics, the *theorem of Pascal* (together with its converse). We note that this theorem is valid not only in the real projective plane, but in any projective plane over a field (i.e. in any Pappian projective plane).

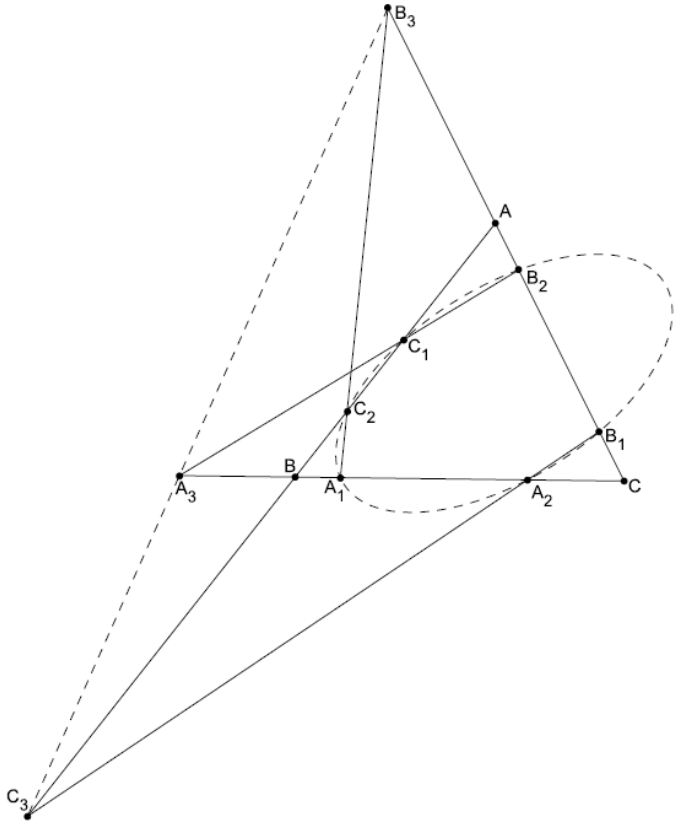
Theorem 1.2. *(Pascal) Suppose that the points A, B, C, D, E, F of the real projective plane are in general position (i.e. no three of them are collinear). Then there is a conic incident with these points if and only if the points $\overleftrightarrow{AB} \cap \overleftrightarrow{DE}$, $\overleftrightarrow{BC} \cap \overleftrightarrow{EF}$ and $\overleftrightarrow{CD} \cap \overleftrightarrow{FA}$ are collinear.*

2. The theorem of Carnot

The theorem of Menelaos gives a necessary and sufficient condition for points on the sides of a triangle to be collinear. The theorem of Carnot is a natural generalization of this theorem, and gives a necessary and sufficient condition for two points on each side of a triangle to form a conic. The proof ([4]) depends on the theorems of Menelaos and Pascal. For completeness we recall it here.

Theorem 2.1. *(Carnot) Let ABC be an arbitrary triangle in the Euclidean plane, and let (A_1, A_2), (B_1, B_2), (C_1, C_2) be pairs points (different from the vertices) on the sides $\overleftrightarrow{BC}$, $\overleftrightarrow{CA}$, $\overleftrightarrow{AB}$, respectively. Then the points A_1, A_2, B_1, B_2, C_1 and C_2 are on a conic if and only if*

$$(ABC_1)(ABC_2)(BCA_1)(BCA_2)(CAB_1)(CAB_2) = 1.$$



Proof. Let $A_3 := \overleftrightarrow{BC} \cap \overleftrightarrow{B_2C_1}$, $B_3 := \overleftrightarrow{AC} \cap \overleftrightarrow{A_1C_2}$ and $C_3 := \overleftrightarrow{AB} \cap \overleftrightarrow{A_2B_1}$. By the theorem of Pascal A_1, A_2, B_1, B_2, C_1, C_2 are on a conic if and only if A_3, B_3, C_3 are collinear. Thus we have to prove that the collinearity of these points is equivalent to the condition above.

Since A_3, B_2, C_1 are collinear, by the theorem of Menelaos

$$(ABC_1)(BCA_3)(CAB_2) = -1.$$

Similarly,

$$(ABC_2)(BCA_1)(CAB_3) = -1$$

and

$$(ABC_3)(BCA_2)(CAB_1) = -1.$$

Multiplying these equalities we get

$$(ABC_1)(ABC_2)(ABC_3)(BCA_1)(BCA_2)(BCA_3)(CAB_1)(CAB_2)(CAB_3) = -1.$$

Using the theorem of Menelaos again, A_3, B_3, C_3 are collinear if and only if

$$(ABC_3)(BCA_3)(CAB_3) = -1.$$

By our previous relation this holds if and only if

$$(ABC_1)(ABC_2)(BCA_1)(BCA_2)(CAB_1)(CAB_2) = 1. \qquad \square$$

A similar generalization of the theorem of Menelaos can be formulated not only for curves of second order (i.e., for conics), but also for the more general class of algebraic curves of order n. By the general theorem, if we consider n points on each side of a triangle (different from the vertices), these $3n$ points are on an algebraic curve of order n if and only if the product of the $3n$ simple ratios as above is $(-1)^n$. The most general version of this theorem has been obtained by B. Segre, cf. [5].

3. The theorem of Carnot from the point of view of barycentric coordinates

In this section we work in the real projective plane and represent its points by homogeneous coordinates. It is well-known that any four points A, B, C, D of general position (no three of the points are collinear) can be transformed by collineation to the points $A'[1,0,0]$, $B'[0,1,0]$, $C'[0,0,1]$, $D'[1,1,1]$. Thus working with the images under this collineation instead of the original points, we may assume for any four points of general position that their coordinates are $[1,0,0]$, $[0,1,0]$, $[0,0,1]$, $[1,1,1]$, respectively.

Let us choose the four-point above such that D is the centroid of the triangle ABC. Then, using the mentioned collineation we call the coordinates of the image of any point P the *barycentric coordinates* of P with respect to the triangle ABC.

Then $[0,1,\alpha]$, $[\beta,0,1]$ and $[1,\gamma,0]$ are the barycentric coordinates of the points A_1, B_1, C_1 such that $(BCA_1) = \alpha$, $(CAB_1) = \beta$ és $(ABC_1) = \gamma$.

We prove this claim for the point A_1 of barycentric coordinates $[0,1,\alpha]$. Let A_M be the midpoint of $\overline{BC}$. Since D is the centroid of ABC, $A_M = \overleftrightarrow{AD} \cap \overleftrightarrow{BC}$, so an easy calculation shows that the barycentric coordinates of A_M are $[0,1,1]$. Since the original points are sent to the points determined by the barycentric coordinates by a collineation, and collineations preserve cross-ratio, it means that $(BCA_1A_F) = \alpha$. Otherwise, since A_M is the midpoint of $\overline{BC}$, $(BCA_M) = 1$, so

$$(BCA_1A_M) = \frac{(BCA_1)}{(BCA_M)} = (BCA_1).$$

Thus we indeed have $(BCA_1) = \alpha$.

In terms of barycentric coordinates the theorem of Menelaos states that the points $[0,1,\alpha]$, $[\beta,0,1]$, $[1,\gamma,0]$ are collinear if and only if $\alpha\beta\gamma = -1$. Similarly, we have the following reformulation of the theorem of Ceva: the lines of $[0,1,\alpha]$ and $[1,0,0]$, $[\beta,0,1]$ and $[0,1,0]$, $[1,\gamma,0]$ and $[0,0,1]$ are concurrent if and only if $\alpha\beta\gamma = 1$.

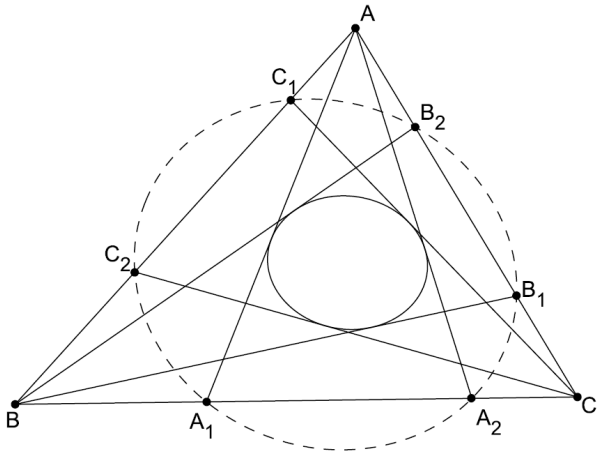
Finally, the theorem of Carnot takes the following form: $[0,1,\alpha_1]$, $[0,1,\alpha_2]$, $[\beta_1,0,1]$, $[\beta_2,0,1]$, $[1,\gamma_1,0]$ and $[1,\gamma_2,0]$ *are on a conic if and only if*

$$\alpha_1\alpha_2\beta_1\beta_2\gamma_1\gamma_2 = 1.$$

4. Tangents to a conic from the vertices of a triangle

The next result was formulated by C. Bradley [1] as a conjecture. In this section we prove Bradley's conjecture applying the theorem of Carnot and using barycentric coordinates. We note that our proof remains valid in any projective plane coordinatized by a field, so we may state our theorem in any Pappian projective plane.

Theorem 4.1. *Let a triangle ABC and a conic $\mathcal{C}$ in the real projective plane be given. The tangent lines from the vertices of ABC to $\mathcal{C}$ intersect the opposite sides of the triangle in six points that are incident to a conic.*



Proof. Let the vertices of the triangle be $A = [1,0,0]$, $B = [0,1,0]$ and $C = [0,0,1]$. Suppose that the tangents of $\mathcal{C}$ incident to A intersect $\overleftrightarrow{BC}$ in $A_1[0,1,\alpha_1]$ and $A_2[0,1,\alpha_2]$; the tangents incident to B intersect $\overleftrightarrow{AC}$ in $B_1[\beta_1,0,1]$ and $B_2[\beta_2,0,1]$, the tangents incident to C intersect $\overleftrightarrow{AB}$ in $C_1[1,\gamma_1,0]$ and $C_2[1,\gamma_2,0]$.

If $[0,1,\alpha]$ is an arbitrary point of $\overleftrightarrow{BC}$, then the points of the line of A and $[0,1,\alpha]$ have coordinates of the form $[1,\lambda,\alpha\lambda]$, where $\lambda \in \mathbb{R}$. If this line is a tangent of c, then there is exactly one λ such that $[1,\lambda,\alpha\lambda]$ satisfies the equation

$$a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 = 0$$

of $\mathcal{C}$. This condition implies that the equation

$$\lambda^2(a_{22} + 2a_{23}\alpha + a_{33}\alpha^2) + \lambda(2a_{12} + 2a_{23}\alpha) + a_{11} = 0$$

has exactly one solution λ. This holds if and only if the discriminant of this quadratic equation vanishes, i.e.,

$$4(a_{12} + \alpha a_{13})^2 - 4a_{11}(a_{22} + 2a_{23}\alpha + a_{33}\alpha^2) = 0.$$

From this an easy calculation leads to the following equation:

$$\alpha^2(a_{13}^2 - a_{11}a_{33}) + \alpha(2a_{12}a_{13} - 2a_{11}a_{23}) + (a_{12}^2 - a_{11}a_{22}) = 0.$$

The solutions of this equation are the α_1 and α_2 coordinates of A_1 and A_2. The product of the roots of this quadratic equation is the quotient of the constant and the coefficient of the second order term, i.e.,

$$\alpha_1\alpha_2 = \frac{a_{12}^2 - a_{11}a_{22}}{a_{13}^2 - a_{11}a_{33}}.$$

By similar calculations we find that

$$\beta_1\beta_2 = \frac{a_{23}^2 - a_{22}a_{33}}{a_{12}^2 - a_{11}a_{22}}$$

and

$$\gamma_1\gamma_2 = \frac{a_{13}^2 - a_{11}a_{33}}{a_{23}^2 - a_{22}a_{33}}.$$

Thus

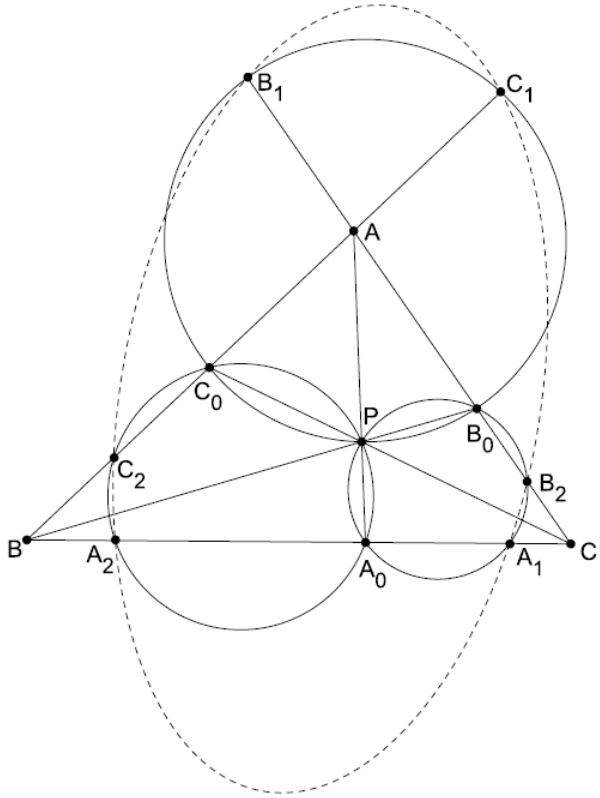
$$\alpha_1\alpha_2\beta_1\beta_2\gamma_1\gamma_2 = \frac{a_{12}^2 - a_{11}a_{22}}{a_{13}^2 - a_{11}a_{33}} \cdot \frac{a_{23}^2 - a_{22}a_{33}}{a_{12}^2 - a_{11}a_{22}} \cdot \frac{a_{13}^2 - a_{11}a_{33}}{a_{23}^2 - a_{22}a_{33}} = 1,$$

and by theorem of Carnot, this implies our claim. □

5. The Cevian conic

In [2] C. Bradley proved the following theorem using barycentric coordinates. We give here a purely synthetic proof, applying again the theorem of Carnot.

Theorem 5.1. *Let ABC be an arbitrary triangle in the Euclidean plane, and let P be an arbitrary point not incident to any of the sides of ABC. Denote the feet of the Cevians through P by A_0, B_0 and C_0. Suppose that the circle through A_0, B_0 and P intersect $\overleftrightarrow{BC}$ in A_1 and $\overleftrightarrow{AC}$ in B_2; the circle through B_0, C_0 and P intersect $\overleftrightarrow{AB}$ in C_1 and $\overleftrightarrow{AC}$ in B_1; the circle through A_0, C_0 and P intersect $\overleftrightarrow{BC}$ in A_2 and $\overleftrightarrow{AB}$ in C_2. Then A_1, A_2, B_1, B_2, C_1, C_2 are on a conic (called the* Cevian conic of P with respect to ABC*).*



Proof. Let the circle through B_0, C_0 and P be c_a; the circle through A_0, C_0 and P be c_b; and the circle through A_0, B_0 and P be c_c. The power of the point A with respect to the circle c_a is

$$AC_1 \cdot AC_0 = AB_1 \cdot AB_0,$$

whence

$$AC_1 = AB_1 \cdot \frac{AB_0}{AC_0}. \tag{5.1}$$

Similarly we get

$$BA_2 = BC_2 \cdot \frac{BC_0}{BA_0}$$

and

$$CB_2 = CA_1 \cdot \frac{CA_0}{CB_0}.$$

The point A is on the power line of c_b and c_c, thus

$$AC_2 \cdot AC_0 = AB_2 \cdot AB_0.$$

Hence

$$AC_2 = AB_2 \cdot \frac{AB_0}{AC_0}.$$

Similarly we get

$$BA_1 = BC_1 \cdot \frac{BC_0}{BA_0}$$

and

$$CB_1 = CA_2 \cdot \frac{CA_0}{CB_0}.$$

Using these results,

$$AC_1 \cdot AC_2 \cdot BA_1 \cdot BA_2 \cdot CB_1 \cdot CB_2 =$$
$$= \frac{(AB_0)^2 \cdot (AB_1) \cdot (AB_2) \cdot (BC_0)^2 \cdot BC_1 \cdot BC_2 \cdot (CA_0)^2 \cdot (CA_1) \cdot (CA_2)}{(AC_0)^2 \cdot (BA_0)^2 \cdot (CB_0)^2}.$$

Applying the theorem of Ceva to the Cevians through P, we get

$$\frac{(C_0B)^2}{(AC_0)^2} \cdot \frac{(A_0C)^2}{(BA_0)^2} \cdot \frac{(B_0A)^2}{(CB_0)^2} = 1,$$

thus

$$AC_1 \cdot AC_2 \cdot BA_1 \cdot BA_2 \cdot CB_1 \cdot CB_2 = C_1B \cdot C_2B \cdot A_1C \cdot A_2C \cdot B_1A \cdot B_2A,$$
$$\frac{AC_1}{C_1B} \cdot \frac{AC_2}{C_2B} \cdot \frac{BA_1}{A_1C} \cdot \frac{BA_2}{A_2C} \cdot \frac{CB_1}{B_1A} \cdot \frac{CB_2}{B_2A} = 1,$$
$$(ABC_1)(ABC_2)(BCA_1)(BCA_2)(CAB_1)(CAB_2) = 1.$$

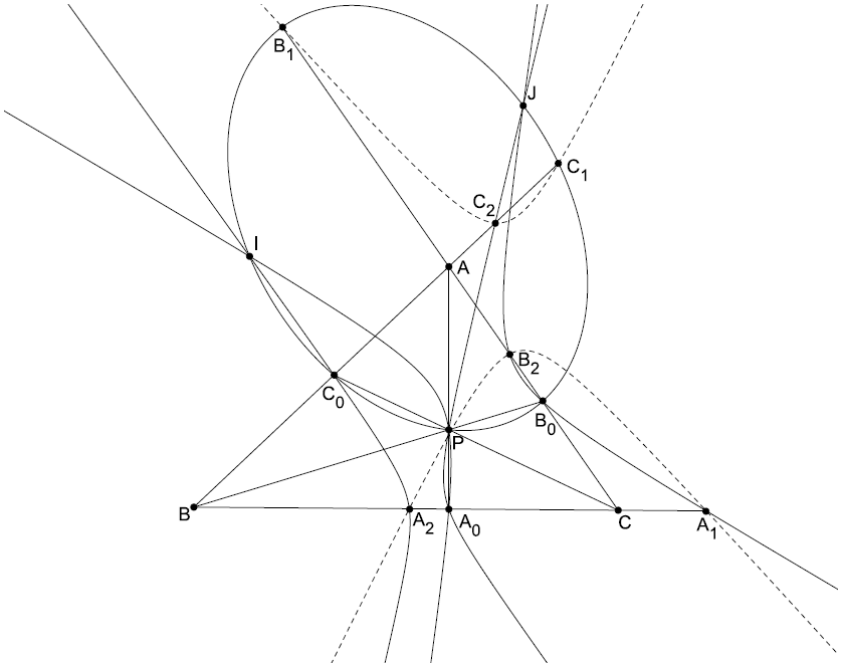
By the theorem of Carnot this proves our claim. □

Remark. It is well known that for any triangle the lines connecting the vertices to the point of contact of the incircle on the opposite sides are concurrent. (This statement can easily be proved using the theorem of Ceva, or the theorem of Brianchon, which is the dual of the theorem of Pascal.) The point of concurrency is called the *Gergonne point* of the triangle. In [3] Bradley proved, using lengthy calculations, that *the Cevian conic of the Gergonne point with respect to a triangle is a circle, whose centre is the incentre of the triangle.* We give an easy elementary proof of his result.

We use the notations of the previous proof and we suppose that P is the Gergonne point of ABC. In this case $AB_0 = AC_0$, so from (5.1) we get $AC_1 = AB_1$. So B_1C_1A is an isosceles triangle, thus the perpendicular bisector of $\overline{B_1C_1}$ is the bisector of the angle $\angle BAC$. Similarly we can prove that the perpendicular bisector of $\overline{B_2C_2}$ is the bisector of $\angle BAC$, the perpendicular bisector of $\overline{A_1C_1}$ and $\overline{A_2C_2}$ is the bisector of $\angle ABC$, and the perpendicular bisector of $\overline{A_2B_1}$ and $\overline{A_1B_2}$ is the bisector of $\angle BCA$. Thus the perpendicular bisectors of the sides of the hexagon $A_1B_2C_2A_2B_1C_1$ pass through the incentre of ABC, so the vertices of the hexagon are on a circle whose centre is the incentre of ABC.

The following result is a projective generalization of the previous theorem.

Corollary 5.2. *Let ABC be an arbitrary triangle in the real projective plane, and let P, I, J be arbitrary points not incident to any of the sides of ABC. Denote the feet of the Cevians through P by A_0, B_0 and C_0. Suppose that the conic through I, J, A_0, B_0 and P intersect $\overleftrightarrow{BC}$ in A_1 and $\overleftrightarrow{AC}$ in B_2; the conic through I, J, B_0, C_0 and P intersect $\overleftrightarrow{AB}$ in C_1 and $\overleftrightarrow{AC}$ in B_1; the conic through I, J, A_0, C_0 and P intersect $\overleftrightarrow{BC}$ in A_2 and $\overleftrightarrow{AB}$ in C_2. Then A_1, A_2, B_1, B_2, C_1, C_2 are on a conic (called the* Cevian conic of P with respect to ABC and (IJ)*).*



Proof. The real projective plane is a subplane of the complex projective plane, so we may consider our configuration in the complex projective plane. Apply a projective collineation of the complex projective plane that sends I and J to $[1, i, 0]$ and $[1, -i, 0]$ (i.e., to the *circular points at infinity*), respectively. It is well known (see e.g. [6]) that a conic of the extended euclidean plane is a circle if and only if (after embedding to the complex projective plane) it is incident with the circular points at infinity. Thus applying our collineation we get the same configuration as in our previous theorem. □

References

[1] C. Bradley: Problems requiring proof, `http://people.bath.ac.uk/masgcs/Article182.pdf`, 2011.

[2] C. Bradley: The Cevian Conic, `http://people.bath.ac.uk/masgcs/Article132.pdf`, 2011.

[3] C. Bradley: When the Cevian Conic is a circle, `http://people.bath.ac.uk/masgcs/Article134.pdf`, 2011.

[4] J. L. S. Hatton: *The Principles of Projective Geometry Applied to the Straight Line and Conic*, Cambridge, 1913.

[5] J. W. P. Hirschfeld: *Projective Geometries Over Finite Fields*, Oxford, 1998.

[6] J. G. Semple, G. T. Kneebone: *Algebraic Projective Geometry*, Oxford, 1952.

Annales Mathematicae et Informaticae
***40** (2012) pp. 145–153*
http://ami.ektf.hu

The rank of certain subfamilies of the elliptic curve $Y^2 = X^3 - X + T^2$*

Petra Tadić†

Institute of Analysis and Computational Number Theory
Technische Universität Graz, Graz, Austria
tadic@math.tugraz.at

Submitted May 24, 2012 — Accepted October 13, 2012

Abstract

Let E be the elliptic curve over $\mathbb{Q}(T)$ given by the equation

$$E : Y^2 = X^3 - X + T^2.$$

It is known that the torsion subgroup is trivial,

$$\text{rank}_{\mathbb{C}(T)}(E) = 2 \quad \text{and} \quad \text{rank}_{\mathbb{Q}(T)}(E) = 2.$$

We find a parametrization of rank ≥ 3 over the function field $\mathbb{Q}(a, i, s, n, k, l)$ where $s^2 = i^3 + a^2$. From this we get families of rank ≥ 3 and ≥ 4 over fields of rational functions in four variables and a family of rank ≥ 5 parametrized by an elliptic curve of positive rank. We also found a particular elliptic curve with rank ≥ 11.

Keywords: parametrization, elliptic surface, elliptic curve, function field, rank, family of elliptic curves, torsion

MSC: 11G05

1. Introduction

Let E be the elliptic curve over $\mathbb{Q}(T)$ given by the equation

$$Y^2 = X^3 - X + T^2.$$

*I would like to thank professor Andrej Dujella for his guidance and discussions on the topic of this paper.

†The author was supported by the Ministry of Science, Education and Sports, Republic of Croatia, grant 037-0372781-2821 and by the Austrian Science Fund (FWF): P 24574-N26.

In [2, Theorem 1], Brown and Myers proved that if $t \geq 2$ is an integer, the elliptic curve $E_t : Y^2 = X^3 - X + t^2$ has rank at least 2 over $\mathbb{Q}$, with linearly independent points $(0, t)$ and $(-1, t)$. They also prove that there are infinitely many integer values of t for which the elliptic curve E_t over $\mathbb{Q}$ has rank at least 3. In [5], Eikenberg showed that the torsion subgroup is trivial, the rank of the group $E(\mathbb{Q}(T))$ equals 2 as does the rank of $E(\mathbb{C}(T))$, both groups have as generators the points $(0, T)$ and $(1, T)$. These results follow also from the more general result by Shioda (see [14, Theorem A_2]). Eikenberg gives quadratic polynomials $T(n) \in \mathbb{Q}[n]$ for which $E_{T(n)}(\mathbb{Q}(n))$ is of rank at least 3, [5, Theorem 4.2.1.]. He also shows that there are infinitely values of t for which E_t has rank at least 5.

In this paper we find a subfamily of E for which the rank over the function field $\mathbb{Q}(a, i, s, n, k, l)$ where $s^2 = i^3 + a^2$ is ≥ 3 and three independent points are listed. From this we get families of rank ≥ 3 and ≥ 4 over fields of rational functions in four variables and a family of rank ≥ 5 over an elliptic curve of positive rank. We also found a particular elliptic curve with rank ≥ 11.

In [16], an elliptic curve $Y^2 = X^3 - T^2X + 1$ was analyzed in a similar way, and the results obtained contain some resemblances with the results of this paper.

2. Subfamilies of higher rank

We know that the elliptic curve E observed in this section and defined above, has rank 2 over $\mathbb{Q}(T)$ and $\mathbb{C}(T)$, with generators $(0, T)$ and $(-1, T)$. First we give two subfamilies which have generic rank ≥ 3 and we give the third independent point. By observing $T(n)$ which are polynomials in the variable n of degree 3 over $\mathbb{Q}$ with an additional point with first coordinate $X(n)$ which is a polynomial in the variable n of degree 2 over $\mathbb{Q}$ on the elliptic curve $Y^2 = X^3 - X + T(n)^2$ over $\mathbb{Q}(n)$ (see [13, Theorem 10.10]), we obtain the following.

Theorem 2.1.
For $T^{(1)}_{\pm}(a, i, s, n, k, l) =$

$$an^3 + (3ak + sl)n^2 + \left(3ak^2 + 2slk - al^2 \pm \frac{s}{i}\right) n - sl^3 - akl^2 + slk^2 \pm \frac{a}{i} l + ak^3 \pm \frac{s}{i} k,$$

the elliptic curve $Y^2 = X^3 - X + T^{(1)}_{\pm}(a, i, s, n, k, l)^2$ *has rank* ≥ 3 *over the function field* $\mathbb{Q}(a, i, s, n, k, l)$ *where* $s^2 = i^3 + a^2$, *with an additional independent point* $C^{(1)}_{\pm}(a, i, s, n, k, l)$ *with first coordinate*

$$X_{C^{(1)}_{\pm}}(a, i, s, n, k, l) = i(n + k)^2 - il^2.$$

Proof. For
$Y_{C^{(1)}_{\pm}}(a, i, s, n, k, l) = sn^3 + (al + 3ks)n^2 + \frac{2aikl \pm a - isl^2 + 3isk^2}{i} n + \frac{-iskl^2 \pm ak - ail^3 + isk^3 + aik^2 l \pm sl}{i}$,
we have

$$X_{C^{(1)}_{\pm}}(a, i, s, n, k, l)^3 - X_{C^{(1)}_{\pm}}(a, i, s, n, k, l) + T^{(1)}_{\pm}(a, i, s, n, k, l)^2 - Y_{C^{(1)}_{\pm}}(a, i, s, n, k, l)^2$$

$$= (-s^2 + i^3 + a^2)q_{\pm}(a, i, s, n, k, l) = 0,$$

where $q_{\pm} \in \mathbb{Q}(a, i, s, n, k, l)$. Here we work over the function field $\mathbb{Q}(a, i, s, n, k, l)$ where $s^2 = i^3 + a^2$.

For the positive case the specialization $(a, i, s, n, k, l) \mapsto (6, -3, 3, 1, 1, 1)$ gives $T_+^{(1)}(6, -3, 3, 1, 1, 1) = 41$, and on the curve $E_{T_+^{(1)}(6,-3,3,1,1,1)} : Y^2 = X^3 - X + 41^2$ there are three corresponding points $(0, 41)$, $(-1, 41)$, $(-9, 31)$ which are independent points of $E_{41}(\mathbb{Q})$. This shows that the points from the claim of the theorem are independent elements of the group

$$E_{T_+^{(1)}(a,i,s,n,k,l)}(\{\mathbb{Q}(a, i, s, n, k, l) : s^2 = i^3 + a^2\}).$$

The proof for $T_-^{(1)}$ is analogous, we used the same specialization. □

Now we will construct two subfamilies of generic rank ≥ 4 by intersecting the families we have obtained. We try to find the solution to the equation

$$T_{\pm}^{(1)}(a, i, s, n, k, l) = T_{\pm}^{(1)}\left(a, 2a\frac{a-s}{i^2}, a\frac{4a^2 - 4as + i^3}{i^3}, n, k_2, m\right),$$

where actually $(i_2, s_2) := \left(2a\frac{a-s}{i^2}, a\frac{4a^2-4as+i^3}{i^3}\right) = (i, s)+(0, a)$ on the elliptic curve $Y^2 = X^3 + a^2$. This gives a polynomial $P(n)$ in the variable n of degree two. Now we choose

$$k_2 := \frac{1}{3}\frac{-4a^3m + 4a^2ms - ami^3 + 3aki^3 + sli^3}{i^3a}$$

so that the coefficient of the polynomial $P(n)$ of the term n^2 is zero. Now that we have $P(n)$ a linear polynomial in n we can choose $n_{\pm}(a, i, s, k, l, m) := (256a^{10}m^3 - 1024a^9m^3s + (-288m^2ki^3 + 192m^3i^3 + 1536m^3s^2)a^8 + (864m^2ski^3 - 96m^2sli^3 - 1024m^3s^3 - 576m^3si^3)a^7 + (256m^3s^4 - 144m^2i^6k \mp 144i^5m - 96m^3i^6 + 288m^2s^2li^3 + 576m^3s^2i^3 - 864m^2s^2ki^3)a^6 + (288m^2i^6ks + 192m^3i^6s \pm 288i^5ms - 192m^3s^3i^3 + 288m^2s^3ki^3 - 48m^2i^6sl - 288m^2s^3li^3)a^5 + (96m^2s^4li^3 \pm 108i^8k \mp 144i^5s^2m - 32m^3i^9 \mp 54li^8 \mp 72i^8m + 54kl^2i^9 + 96m^2i^6s^2l - 144m^2s^2i^6k - 96m^3s^2i^6 - 72m^2i^9k)a^4 + (72m^2i^9ks - 54kl^2i^9s + 54sl^3i^9 \pm 72i^8sm \pm 90li^8s + 32m^3i^9s \mp 162ski^8 - 24m^2i^9sl - 48m^2s^3i^6l)a^3 + (\pm 54s^2ki^8 \pm 18i^{11}m \mp 36i^8s^2l - 54s^2l^3i^9 \pm 27i^{11}k + 18ki^9s^2l^2 + 24m^2i^9s^2l)a^2 + (2s^3l^3i^9 - 18ki^9s^3l^2 \pm 9i^{11}sl)a - 2s^4l^3i^9)/(9ai^3(32a^7m^2 - 96a^6m^2s + (16m^2i^3 + 96m^2s^2)a^5 + (-32m^2i^3s - 32m^2s^3)a^4 + (\mp 12i^5 + 16m^2s^2i^3 - 6l^2i^6 + 8m^2i^6)a^3 + (6l^2i^6s - 8m^2i^6s \pm 18i^5s)a^2 + (\mp 3i^8 \mp 6s^2i^5 - 2s^2l^2i^6)a + 2s^3l^2i^6))$ such that

$$T_{\pm}^{(1)}(a, i, s, n_{\pm}(a, i, s, k, l, m), k, l) =$$

$$= T_{\pm}^{(1)}\left(a, 2a\frac{a-s}{i^2}, a\frac{4a^2 - 4as + i^3}{i^3}, n_{\pm}(a, i, s, k, l, m), \frac{1}{3}\frac{-4a^3m + 4a^2ms - ami^3 + 3aki^3 + sli^3}{i^3a}, m\right).$$

Proposition 2.2. *Let*

$$S^{(1)}_{\pm}(a,i,s,k,l,m) := T^{(1)}_{\pm}(a,i,s,n_{\pm}(a,i,s,k,l,m),k,l),$$

where $n_{\pm}$ is given above and $T^{(1)}_{\pm}$ is as in Theorem 2.1. The elliptic curve

$$Y^2 = X^3 - X + S^{(1)}_{\pm}(a,i,s,k,l,m)^2$$

over the function field $\mathbb{Q}(a,i,s,k,l,m)$ where $s^2 = i^3 + a^2$ has rank ≥ 4 with four independent points, the two generators $(0, S^{(1)}_{\pm}(a,i,s,k,l,m))$, $(-1, S^{(1)}_{\pm}(a,i,s,k,l,m))$ mentioned in the introduction, and two additional points

$$A^{(1)}{}_{\pm}(a,i,s,k,l,m) := C^{(1)}_{\pm}(a,i,s,n_{\pm}(a,i,s,k,l,m),k,l)$$

and

$$B^{(1)}{}_{\pm}(a,i,s,k,l,m) :=$$

$$C^{(1)}_{\pm}\left(a, 2a\frac{a-s}{i^2}, a\frac{4a^2-4as+i^3}{i^3}, n_{\pm}(a,i,s,k,l,m), \frac{1}{3}\frac{-4a^3m+4a^2ms-ami^3+3aki^3+sli^3}{i^3a}, m\right)$$

(notation for $C^{(1)}_{\pm}$ from Theorem 2.1).

Proof. With the specialization $(a,i,s,k,l,m) \mapsto (6,-3,3,1,1,1)$ we prove that the above listed four points on the elliptic curve (over $\mathbb{Q}(a,i,s,k,l,m)$ where $s^2 = i^3 + a^2$) are independent, since the specialization gives the elliptic curve

$$E_{S^{(1)}_{+}(6,-3,3,1,1,1)} : Y^2 = X^3 - X + \left(-\frac{5647}{13122}\right)^2$$

with the corresponding four independent points with first coordinates 0, -1, $-\frac{805}{972}$, $\frac{7084}{729}$.

The proof for $S^{(1)}_{-}$ is analogous, by picking an adequate specialization. □

Remark 2.3. The variety (from Theorem 2.1)

$$s^2 = i^3 + a^2$$

can be observed as an elliptic curve $Y^2 = X^3 + T^2$ over the field $\mathbb{Q}(T)$. In [12, Corollary 8] it is shown that the torsion subgroup of $s^2 = i^3 + a^2$ over $\mathbb{Q}(a)$ is equal $\{O,\ (0,a),\ (0,-a)\}$. This elliptic curve has rank 0 over $\mathbb{Q}(a)$. For more details see [6, p. 112]. Points on the variety $s^2 = i^3 + a^2$ from Theorem 2.1 can easily be obtained, for example $(a,i,s) = (6,-3,3)$ is a point on the variety. For $a = 0$ we have $i = u^2$ and $s = u^3$, in this case $T^{(1)}_{\pm}(0,u^2,u^3,n,k,l)$ in Theorem 2.1 is a quadratic polynomial in n. We also have parametrizations of this variety [3, Section 14.2]:

$$\begin{cases} a(t) = 2t^3 - 1, \\ i(t) = 2t, \\ s(t) = 2t^3 + 1, \end{cases}$$

For this parametrization Theorem 2.1 and Proposition 2.2 transform into:

Corollary 2.4.

(i) Let
$T_\pm^{(2)}(t,n,k,l) := T_\pm^{(1)}(2t^3-1,2t,2t^3+1,n,k,l) = ((4t^4-2t)n^3+((4l+12k)t^4+(2l-6k)t)n^2+((-4l^2+8lk+12k^2)t^4\pm 2t^3+(4lk-6k^2+2l^2)t\pm 1)n+(-4kl^2-4l^3+4k^3+4lk^2)t^4\pm(2k+2l)t^3+(2lk^2-2l^3+2kl^2-2k^3)t\pm(k-l))/(2t)$. *The elliptic curve* $Y^2=X^3-X+T_\pm^{(2)}(t,n,k,l)^2$ *over* $\mathbb{Q}(t,n,k,l)$ *has rank* ≥ 3 *and three independent points have first coordinates* $(0,T_\pm^{(2)}(t,n,k,l))$, $(-1,T_\pm^{(2)}(t,n,k,l))$, $C_\pm^{(1)}(2t^3-1,2t,2t^3+1,n,k,l)$. *Notation for* $T_\pm^{(1)}$ *and* $C_\pm^{(1)}$ *as in Theorem 2.1.*

(ii) Let
$$S_\pm^{(2)}(t,k,l,m) := S_\pm^{(1)}(2t^3-1,2t,2t^3+1,k,l,m).$$
Then the elliptic curve $Y^2=X^3-X+S_\pm^{(2)}(t,k,l,m)^2$ *over the function field* $\mathbb{Q}(t,n,k,l)$ *is of rank* ≥ 4, *with four independent points* $(0,S_\pm^{(2)}(t,k,l,m))$, $(-1,S_\pm^{(2)}(t,k,l,m))$, $A_\pm^{(1)}(2t^3-1,2t,2t^3+1,k,l,m)$, $B_\pm^{(1)}(2t^3-1,2t,2t^3+1,k,l,m)$. *Here the notation is from Proposition 2.2.*

Proof.

(i) For the specialization $(t,n,k,l)\mapsto(1,2,1,1)$ on the curve
$$E_{T_+^{(2)}(1,2,1,1)} : Y^2 = X^3 - X + 53^2$$
the corresponding points with first coordinates 0, -1, 16 are independent, so the claim of the corollary is true. The proof for $T_-^{(2)}$ is analogous, by picking an adequate specialization.

(ii) The specialization $(t,k,l,m)\mapsto(2,1,1,1)$ gives the elliptic curve
$$E_{S_+^{(2)}(2,1,1,1)} : Y^2 = X^3 - X + \left(-\frac{49050562229}{10497600}\right)^2$$
over $\mathbb{Q}$ for which the four listed points with first coordinates 0, -1, $\frac{14863849}{72900}$, $-\frac{48719569}{311040}$ are independent. This proves that for the elliptic curve $Y^2=X^3-X+S_+^{(2)}(t,k,l,m)^2$ over the field $\mathbb{Q}(t,k,l,m)$ the corresponding four points the two generators mentioned in the introduction and the points $A_\pm^{(1)}(2t^3-1,2t,2t^3+1,k,l,m)$ and $B_\pm^{(1)}(2t^3-1,2t,2t^3+1,k,l,m)$ (from Proposition 2.2) are independent. The proof for $S_-^{(2)}$ is analogous, by picking an adequate specialization. □

3. Subfamily of generic rank ≥ 5

Remark 3.1.

- In [5, Theorem 3.5.1.] a rational function is given
$$M(m) = \frac{1017m^4 - 8487m^3 + 19298m^2 - 14145m + 2825}{(3m^2-5)^2},$$
with the property that the rank of $E_{M(m)}$ over $\mathbb{Q}(m)$ is ≥ 4.
- We have two additional points coming from [5, Theorem 3.5.1.], R_3 with first coordinate
$$-\frac{69m^2 - 414m + 295}{3m^2-5}$$
and the point R_4 with first coordinate
$$\frac{357m^2 - 410m + 95}{3m^2-5}.$$
- This rational function $M(m)$ is equal $T_+^{(1)}\left(0, 9, 27, n, -\frac{1}{3}\frac{9nm^2-20m^2+69m-15n-35}{3m^2-5}, 1\right)$ in Theorem 2.1. The third point R_3 in [5] is equal $(0, T_+^{(1)}) + (-1, T_+^{(1)}) - C_+^{(1)}$, where $C_+^{(1)}$ is the third independent point in Theorem 2.1.
- The rational function $M(m)$ is also equal
$$T_+^{(1)}(0, 25, 125, n, -\frac{1}{25}\frac{75nm^2 - 102m^2 + 205m - 125n - 175}{3m^2-5}, 1).$$
The fourth point R_4 in [5] is equal $(-1, T_+^{(1)}) - C_+^{(1)}$, where $C_+^{(1)}$ is the third independent point in Theorem 2.1.
- In [5] an elliptic surface over a curve is found for which the Mordell-Weil group has rank ≥ 5. Here we give another example of an infinite family of elliptic curves of generic rank ≥ 5.

Theorem 3.2. *The elliptic curve*
$$Y^2 = X^3 - X + \left(\frac{3723875}{729}n^2 + \frac{155}{9}n - \frac{3723875}{729}\right)^2$$
over the function field $\mathbb{Q}(m, n)$ *where* $\left((3m^2-5)\left(\frac{48050}{81}n+1\right)\right)^2 =$
$$= \frac{2257735321}{729}m^4 + 584660m^3 - \frac{25995527290}{2187}m^2 + \frac{2923300}{3}m + \frac{56443383025}{6561},$$
has rank ≥ 5 *with five independent points with first coordinates*
$$0,\ -1,\ -\frac{69m^2 - 414m + 295}{3m^2-5},\ \frac{357m^2 - 410m + 95}{3m^2-5},\ \frac{24025}{81}n^2 - \frac{24025}{81}.$$

Proof. Here we will intersect $M(m)$ with $T_+^{(1)}(0,u^2,u^3,n,k,l)$ from Theorem 2.1 to obtain a subfamily of higher rank:

$$M(m)=T_+^{(1)}(0,u^2,u^3,n,k,l)=u^3l(n+k+\frac{1}{2u^2l})^2-\frac{1}{4}\frac{(2u^2l^2-2ul+1)(2u^2l^2+2ul+1)}{ul}.$$

This gives $(2u^2l(3m^2-5)(n+k+\frac{1}{2u^2l}))^2=$

$$=(9+36(ul)^4+4068(ul))m^4-33948(ul)m^3+(-30+77192ul-120(ul)^4)m^2$$
$$-56580(ul)m+25+100(ul)^4+11300(ul).$$

So, the point $m=1$ will be the solution of the above equation if $c=ul$ is the first coordinate on

$$\square=16c^4+2032c+4.$$

The corresponding elliptic curve is of rank five and from one of the generators of the free part we get $c=ul=-\frac{155}{9}$ (chosen such that the specialization $m=1$ gives the independence of points). So we take $k=0$, $l=1$ and we look at the intersection

$$M(m)=T_+^{(1)}\left(0,\left(-\frac{155}{9}\right)^2,\left(-\frac{155}{9}\right)^3,n,0,1\right)=-\frac{3723875}{729}n^2-\frac{155}{9}n+\frac{3723875}{729},$$

and we get that (m,n) lies on

$$\left((3m^2-5)\left(\frac{48050}{81}n+1\right)\right)^2=\frac{2257735321}{729}m^4+584660m^3$$
$$-\frac{25995527290}{2187}m^2+\frac{2923300}{3}m+\frac{56443383025}{6561}. \tag{3.1}$$

So (m,n) on (3.1) gives five points from the claim of the theorem (where the third and fourth point are from [5] and the last point is from Theorem 2.1).

For the specialization $(m,n)\mapsto(1,-\frac{4753}{4805})$ we get the elliptic curve

$$E_{M_2(1)}=E_{T_+^{(1)}\left(0,\left(-\frac{155}{9}\right)^2,\left(-\frac{155}{9}\right)^3,-\frac{4753}{4805},0,1\right)}=E_{127}:Y^2=X^3-X+127^2,$$

with corresponding five independent points with first coordinates 0, -1, -25, -21, $-\frac{6136}{961}$. So the five points from the claim of the theorem are independent. $\square$

Remark 3.3. Points (m,n) in the above theorem can be obtained with the transformation

$$m=\frac{11602011740X-139896435555764171800+47449Y}{47449Y+7099196538X-80704505760225548460},$$

where (X,Y) is a point on the curve

$$Y^2=X^3-4119006235730787327 00X+32137586998783982379698901460 00.$$

The value of n can be obtained from (3.1). This curve is of positive rank by [7], so the subfamily of elliptic curves from Theorem 3.2 is infinite.

4. Specializations of high rank

The highest rank found for the elliptic curve $E_t : Y^2 = X^3 - X + t^2$ over $\mathbb{Q}$ is ≥ 11 and is obtained for $t = 1118245045$. In this case we get the elliptic curve $E_{1118245045} : Y^2 = X^3 - X + 1118245045^2$ and eleven independent points

$$(1, 1118245045), (-1, 1118245045), (-149499, 1116750055), (-187723, 1115283209)$$

$$(208403, 1122284857), (-357751, 1097581405), (-369623, 1095433091),$$

$$(-398399, 1089604235), (402083, 1146942473), (506597, 1174940551),$$

$$(919987, 1424474279).$$

This was found using the sieve method explained in [4, 8, 10]. Here we observed $t = \frac{t_1}{t_2}$ ($1 \leq t_2 \leq 10000$, $1 \leq t_1 \leq 100000$), and elliptic curves E_t with $S(523, E_t) > 23$ for which $S(1979, E_t) > 43.5$. The lower bound was found using the command `Seek1` in Apecs [1]. In addition we observed integers $1 \leq t \leq 1130000000$, and elliptic curves E_t with $S(523, E_t) > 23$ for which $S(1979, E_t) > 41.5$ for the remaining ones. Here is the list of values t which we obtained with rank ≥ 8:

rank	t
≥ 8	$\frac{1567}{3025}$, $\frac{7247}{7688}$, $\frac{23618}{9025}$, $\frac{14809}{4800}$, $\frac{32971}{9072}$, $\frac{22069}{5329}$, $\frac{23581}{3481}$, $\frac{18353}{2197}$, $\frac{4882}{529}$, $\frac{88745}{8496}$, $\frac{74227}{6859}$, $\frac{47059}{3698}$, $\frac{6913}{242}$, $\frac{17489}{343}$, $\frac{53708}{529}$, $\frac{11689}{50}$, $\frac{29689}{2}$, 78560, 2011060, 14083286, 14083286, 21717559, 35498230, 38998023, 45321449, 58235977, 67190943, 67292109, 83402041, 86010677, 96384349, 101940616, 122421035, 159056061, 171981307, 200300248, 217135540, 230684707, 266349308, 307253369, 329132909, 331903387, 342825543, 349640440, 391942721, 423787655, 436687265, 484259053, 484594343, 566328793, 586597025, 594744835, 594782908, 594869501, 598442638, 620933242, 631151494, 747946597, 781809427, 787815289, 836422595, 851738165, 919540903, 1015597721, 1029670387, 1111072411
≥ 9	$\frac{20155}{7442}$, $\frac{90719}{9248}$, $\frac{36749}{1225}$, $\frac{51691}{1089}$, $\frac{83351}{1521}$, $\frac{70313}{845}$, 423515, 829999, 1741033, 2650019, 7030799, 11180651, 53958107, 70808669, 76758473, 97399947, 101469425, 154523221, 197903551, 281137843, 300361741, 304354681, 352968853, 355308367, 599768545, 863227439, 911227325, 1040969455
≥ 10	765617, 17708315, 64232534, 77799653, 236076508, 269371865, 337557943, 450112831, 808983247
≥ 11	1118245045

The greatest rank obtained in [5] was rank 6 for $t = 337$, while the greatest rank obtained in [2] was rank 10 for $t = 765617$.

References

[1] I. Connell, *APECS*, ftp://ftp.math.mcgill.ca/pub/apecs/

[2] E. Brown, B.T. Myers, *Elliptic Curves from Mordell to Diophantus and Back*, Amer. Math. Monthly, 109, Aug-Sept 2002, 639-648.

[3] H. Cohen, *Number Theory. Volume II: Analytic and Modern Tools*, Springer Verlag, Berlin, 2007.

[4] A. Dujella, *On the Mordell-Weil groups of elliptic curves induced by Diophantine triples*, Glas. Mat. Ser. III 42 (2007), 3-18.

[5] E.V.Eikenberg, *Rational points on some families of elliptic curves*, PhD thesis, University of Maryland, 2004.

[6] A. Knapp, *Elliptic Curves*, Princeton University Press, Princeton, NJ, 1992.

[7] Computational Algebra Group, *MAGMA*, University of Sydney http://magma.maths.usyd.edu.au/magma/, 2002.

[8] J.-F. Mestre, *Construction de courbes elliptiques sur* $\mathbb{Q}$ *de rang 12*, C. R. Acad. Sci. Paris Ser. I 295 (1982) 633-644.

[9] R. Miranda, *An overview of algebraic surfaces*, in Algebraic geometry (Ankara,1995), Lecture Notes in Pure and Appl. Math. 193, Dekker, New Yore, 1997, 197-217.

[10] K. Nagao, *An example of elliptic curve over Q with rank* $\geq$ *20*, Proc. Japan Acad. Ser. A Math. Sci. 69 (1993), 291-293.

[11] C. Batut, K. Belabas, D. Bernardi, H. Cohen, M. Olivier, *The Computer Algebra System PARI - GP*, Université Bordeaux I, 1999, http://pari.math.u-bordeaux.fr

[12] N. F. Rogers, *Elliptic Curves* $x^3 + y^3 = k$ *with High Rank*, PhD thesis, Harvard University, 2004.

[13] T. Shioda, *On the Mordell-Weil lattices*, Comment. Math. Univ. St. Pauli 39 (1990), 211-240.

[14] T. Shioda, *Construction of elliptic curves with high rank via the invariants of the Weyl groups*, J. Math. Soc. Japan 43 (1991), no. 4, 673-719.

[15] J. Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves*, Graduate Texts in Mathematics 151, Springer-Verlag, Berlin - New York, 1994.

[16] P. Tadić, *On the family of elliptic curves* $Y^2 = X^3 - T^2X + 1$, Glas. Mat. Ser. III, 47 (2012), 81-93.

Annales Mathematicae et Informaticae
***40** (2012) pp. 155–160*
http://ami.ektf.hu

Measurement of visual smoothness of blending curves*

Robert Tornai

University of Debrecen, Faculty of Informatics
Department of Computer Graphics and Image Processing
tornai.robert@inf.unideb.hu

Submitted September 25, 2012 — Accepted December 11, 2012

Abstract

This paper considers the visual smoothness of interpolating curves. It will examine skinning algorithms in detail. Especially the 2D ball skinning algorithms will be covered. Slabaugh introduced an energy function [1] and Kunkli defined a process to find the touching points [2] and made an elegant skinning method with Hoffmann based on classical geometry [3]. I will try to give a simple metric for visual smoothness based on the number of direction change of the yielded interpolation curve. Minimizing this metric will give the best visual result.

Keywords: measurement, visual smoothness, interpolation, skinning, circles, spheres, avatar

MSC: MSC 2010: 68U05 Computer graphics; computational geometry

1. Introduction

Skinning algorithms are gaining more and more popularity in industry, engineering, design and art. They provide an intuitive and effective way to describe complex shapes.

These complex shapes can include the face and the body of three dimensional avatars. If an avatar's face shall be customized, for example to follow the viewer's physiognomy, then a skinning surface is needed. This surface can be considered

*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.

better or nicer, if it is smoother. I will investigate this visual smoothness by the aspect of direction changes among generated curves.

In the following, 2D ball skinning algorithms will be covered. Let's take a series of circles. Slabaugh introduced an energy function for making a skinning curve (see Figure 1).

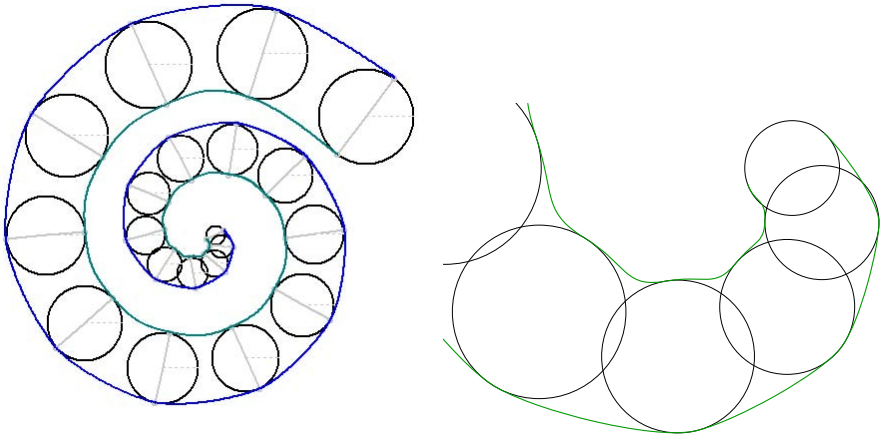


Figure 1: Slabaugh's skinning curves for the series of circles and the zoomed wavy inner part

If the inner part is enlarged, it can be noticed that the inner curve is very wavy. If we take the same enlarged inner part of the Kunkli-Hoffmann's algorithm, it can be noticed that the inner curve is following only one direction (see Figure 2). There are no inflection points.

It seems that the number of the inflection points of a curve is a good measure for the visual smoothness.

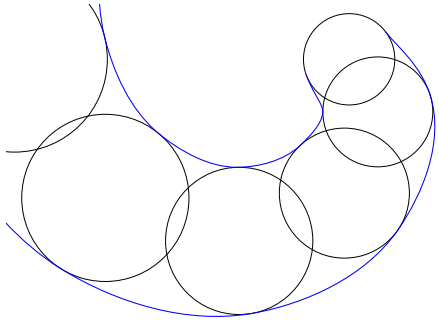


Figure 2: The result of the Kunkli-Hoffmann interpolation curve

2. Measurement for Visual Smoothness

Let's introduce a measure that will sum the inflection points of the Bézier curve parts that make up the whole curve.

"The algebraic form of a parametric cubic belongs to one of three projective types, as shown in Figure 4. Any arbitrary cubic curve can be classified as a serpentine, cusp, or loop. A very old result (Salmon 1852) on cubic curves states that all three types of cubic curves will have an algebraic representation that can be written $k^3 - lmn = 0$, where k, l, m, and n are linear functionals corresponding to lines k, l, m, and n as in Figure 3.

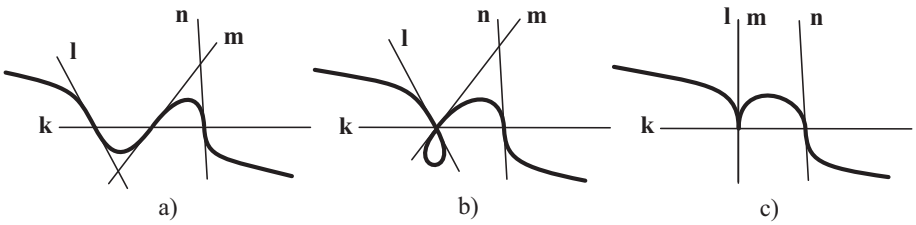


Figure 3: All parametric cubic plane curves can be classified as the parameterization of some segment of one of these three curve types.
a) Serpentine curve. This curve has three collinear inflection points (on line k) with tangent lines l, m and n at those inflections.
b) Loop curve. This curve has one inflection and one double point with k the line through them. The lines l and m are the tangents to the curve at the double point and n is the tangent at the inflection.
c) Cusp curve. This curve has one inflection point and one cusp, with k the line through them. The line $l = m$ is the tangent at the cusp and n is the tangent at the inflection. [4]

A cubic Bézier curve in homogeneous parametric form is written

$$C(s,t) = [(s-t)^3 \quad 3(s-t)^2 s \quad 3(s-t)^2 s^2 \quad s^3] \cdot \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix},$$

where the $\mathbf{b}_i$ are cubic Bézier control points.

The first step is to compute the coefficients of the function $I(s,t)$ whose roots correspond to inflection points of $C(s,t)$. An inflection point is where the curve changes its bending direction, defined mathematically as parameter values where the first and second derivatives of $C(s,t)$ are linearly dependent. The derivation of the function I is not needed for the current purposes. For integral cubic curves,

$$I(s,t) = t(3d_1 s^2 - 3d_2 st + d_3 t^2),$$

where

$$d_1 = a_1 - 2a_2 + 3a_3,$$
$$d_2 = -a_2 + 3a_3,$$
$$d_3 = 3a_3$$

and

$$a_1 = \mathbf{b}_0 \cdot (\mathbf{b}_3 \times \mathbf{b}_2),$$
$$a_2 = \mathbf{b}_1 \cdot (\mathbf{b}_0 \times \mathbf{b}_3),$$
$$a_3 = \mathbf{b}_2 \cdot (\mathbf{b}_1 \times \mathbf{b}_1).$$

The function I is a cubic with three roots, not all necessarily real. It is the number of distinct real roots of $I(s,t)$ that determines the type of the cubic curve. For integral cubic curves, $[s\ t] = [1\ 0]$ is always a root of $I(s,t)$. This means that the remaining roots of $I(s,t)$ can be found using the quadratic formula, rather than by the more general solution of a cubic – a significant simplification over the general rational curve algorithm.

The cubic curve classification reduces to knowing the sign of the discriminant of $I(s,t)$, defined as

$$discr(I) = d_1^2(3d_2^2 - 4d_1d_3).$$

If $discr(I)$ is positive, the curve is a serpentine; if negative, it is a loop; and if zero, a cusp. Although it is true that all cubic curves are one of these three types, not all configurations of four Bézier control points result in cubic curves. It is possible to represent quadratic curves, lines, or even single points in cubic Bézier form. The procedure will detect these cases, and the rendering algorithm can handle them. It is not needed to consider (or render) lines or points, because the convex hull of the Bézier control points in these cases has zero area and, therefore, no pixel coverage. The general classification of cubic Bézier curves is given by Table 1.

Serpentine	$discr(I) > 0$
Cusp	$discr(I) = 0$
Loop	$discr(I) < 0$
Quadratic	$d_1 = d_2 = 0$
Line	$d_1 = d_2 = d_3 = 0$
Point	$\mathbf{b}_0 = \mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_3$

Table 1: Cubic Curve Classification

If the Bézier control points have exact floating-point coordinates, the classification given in Table 1 can be done exactly. That is, there is no ambiguity between cases, because $discr(I)$ and all intermediate variables can be derived from exact floating representations." [5]

A serpentine has three inflection points while a cusp have one inflection point. Bézier curves of other type will have one inflection point if $\mathbf{b}_0$ and $\mathbf{b}_3$ are on the

two different sides of the line defined by $\mathbf{b}_1$ and $\mathbf{b}_2$. It can be easily determined by the scalar product of the homogeneous coordinates of the control points and the equation of the line.

To define an exact measurement for visual smoothness, the number of these inflection points shall be summarized. This measurement shall be extended by the inflection points at the joining points of the Bézier curves that make up the interpolating curves used for skinning the series of circles.

Two joining Bézier curves defined by $(\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ and $(\mathbf{b}_3, \mathbf{b}_4, \mathbf{b}_5, \mathbf{b}_6)$ will have an inflection point in the joining $\mathbf{b}_3$ control point if $\mathbf{b}_1$ and $\mathbf{b}_5$ are on the different sides of the line defined by $\mathbf{b}_2$, $\mathbf{b}_3$ and $\mathbf{b}_4$ control points.

The final measurement will be the summarized inflection points inside the Bézier curves and the inflection points at the joining end points of the curves.

Figure 4 shows that Slabaugh's algorithm has five, while the Kunkli-Hoffmann's curve has only four inflection points on the top skinning curves.

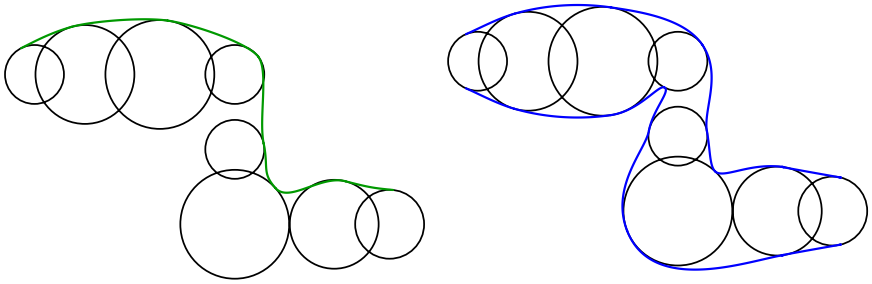


Figure 4: Second example for Slabaugh and Kunkli-Hoffmann curves for the same series of circles

3. Results

Examining the interpolating inner curves only of the last six inner circles of the series of Figure 1 and Figure 2 shows that Slabaugh's curve has five inflection point among the six circles while Kunkli-Hoffmann's curve has none. Even on the more complex Figure 4 this ratio is five to four. Thus, in both cases the second curve is more smooth.

Testing for further arrangements, Kunkli-Hoffmann's interpolation curves usually have fewer inflection points and this way they yield a more smooth interpolating curve. By using these more smooth curves and surfaces, better customized sulutions can be provided for simulations or avatars' heads and bodies.

References

[1] G. Slabaugh, G. Unal, T. Fang, J. Rossignac, B. Whited, Variational skinning of an ordered set of discrete 2D balls, *Lecture Notes on Computer Science,* Vol. 4795 (2008), 450–461.

[2] R. Kunkli, Localization of touching points for interpolation of discrete circles, *Annales Mathematicae et Informaticae*, 36 (2009), 103–110.

[3] R. Kunkli, M. Hoffmann, Skinning of circles and spheres, *Computer Aided Geometric Design*, 27 (2010), 611–621.

[4] C. Loop, J. Blinn, Resolution independent curve rendering using programmable graphics hardware, *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2005*, Vol. 24 Issue 3 (July 2005), 1000–1009.

[5] H. Nguyen, GPU Gems 3 (Chapter 25. Rendering Vector Art on the GPU), *Addison-Wesley Professional*, (2007), 543–562.

Methodological papers

Annales Mathematicae et Informaticae
***40** (2012) pp. 163–173*
http://ami.ektf.hu

A possible general approach of the Apollonius problem with the help of GeoGebra

László Budai

University of Debrecen
budai0912@gmail.com

Submitted November 2, 2012 — Accepted November 28, 2012

Abstract

The Apollonius problem is one of the circle tangency problems that the eponymous Apollonius the "greatest Greek geometer" himself dealt with first and solved in general. Since then many people in many ways dealt with the problem. The new geometry is looking for ways to find simpler solutions for the problem differing from the draw up of elementary geometry. The current abstract shortly describes the frequent possible solution methods and more particularly deals with the three dimensional geometric solution methods as well as the practicability with the use in GeoGebra.

Keywords: Apollonius-problem, GeoGebra, cyclography

MSC: M10, G10, G40

1. Introduction

Apollonius was born in Perga (Greece) around 262 BC. He conducted his mathematical studies in Alexandria in Euklides School where he also worked for a longer period of time. His works have not survived in originals; even one of them titled Konika on conic sections can be reached only in Arabic revision.

The original "Apollonius problem" is the following. Given three circles lying in the same plane, from these circles those need to be drawn up which are tangent with the three circles. If the radiuses of the given circles are reduced beyond all limits, then the circles are reduced to points or shrink into so called point circles. If

the radiuses of the given circles increase beyond all limits, then the circles become straight lines. Taking into consideration the this way generated degenerated cases we can conclude the following: from point, straight line and circle i.e. the number of combinations of the repetition of the three elements give all the possible problems. Therefore we need to examine ten different tangency problems.

So, the amendment of the text for the problem can be the following: Those circles need to be drawn up that are points, straight lines and circles, on the same horizontal plane and from these they are tangent with three. In case if a circle is shrinking into a point the tangency means the fitting on the point.

Numbers of people ever since have dealt with solving these tangency problems such as: Viéte, Descartes, Newton, Lambert, Euler, Carnot, Gauss etc. The new geometry is seeking simpler ways for solutions, differing from elementary geometry draw ups, taking the problem out from plane and placing it into three dimensions, this way the draw up is done with the help of spheres or cones to perform the draw up.

2. Some more frequent methods for solving the problem, brief

n order to perform the draw ups we can group the problems in such way that first we take the simple drawing up tasks and later from these take the hard draw ups and we can perform the more difficult draw ups or trace them back to simpler cases. We can use it for almost all solution methods which will be used.

The grouping is as follows (Figure 1.):

1.	p	p	p	2.	e	e	e	3/a.	p	p	e	4/a.	p	e	e	5/a.	p	p	k
								3/b.	p	e	k	4/b.	e	e	k	5/b.	p	k	k
								3/c.	e	k	k					5/c.	k	k	k

Figure 1: The 10 problems built on each other

2.1. Elementary draw up

The 1st and the 2nd cases are the simplest cases and the draw ups can be carried out on the basis of elementary school knowledge: the perpendicular bisector segments defined by points represent the centre of the circle and in the 2nd case the geometrical place of the searchable centres of the circles are by the bisecting straight lines at the angle bisector and these are the intersections of the straight lines.

Solving case 3/a. can be summarized as follows: the geometrical place in case of searching for the centres of the circles to be identified by the straight line's section bisector of two points specified by a straight line and the circles which is

providing solution for the problem has tangency points on the straight line and the perpendicular intersection drawn at the straight line by the point of contact on the straight line is giving the solution to the problem.

In case of 3/b. we can use a handy trick: let's consider the problem solved, this way we can note some correlations of the geometrical place of the centres of the circles which are to be searched.

3/c. case is closely linked to the draw up done in case 3/b. Sub cases are obtained if we do the draw up by decreasing or increasing the radius (in this case the radius of the straight line is shifted with the given radius to a specific direction).

In short the solution for problem 4/a.: one point can be appointed anywhere after the draw up of the angel bisector identified by two straight lines, after this using similarity we arrive to the geometrical place of the centres of the circles searched.

The 4/b. problem can be traced back to problem 4/a.: we shrank the circles to points and shift the straight lines with the appropriate directionality.

Let's consider the problem already solved during the solution of case 5/a., this way we arrive to the following: the geometrical place of the centres of the searched circles is the perpendicular intersection of the power points used as tangents to the circle and their contact points to the circle as well as the perpendicular bisector of the distance between the points.

We perform the 5/b. case draw up similarly, as the above techniques: the previous methods are used with the use of the similarity as far as the external and internal points.

The most obvious solution to case 5/c. is to trace it back to problem 5/b. The detailed elaboration with figures can be found among others in [1].

2.2. Problem solving with inversion

The inversion as a geometrical transformation is extremely suitable to solve such problems where circles, straight lines and their tangency are involved.

If the inversion is known as a transformation we can proceed as follows: the draw up of the circles contact points of the three circles can be simplified to the draw up of two given circles tangency points passing through one point of the drawn up circle (with decreasing the radius). Then when we select the specified point as a pole, the draw up is altered to the draw up of a straight line with tangency points with the circle as an inverse solution.

All possible solutions to the problem can be obtained by reducing or increasing the radius of the circles. Where straight lines are included also we need to apply the parallel shift of the straight lines instead of the increase/decrease of the radius. This way we can carry out the solutions to all the problems (similar to those seen in the elementary draw up).

2.3. Hyperbola sections

The solution method originating from Adriaan van Roomen (1596) is based on the sections of two hyperboles. He simplified the problem with looking for tangent circles for two given circles. He concluded that the two tangential circles' centres fall to that hyperbola, whose focal point coincides with the centres of the given circles. The centre of the searched circle and the difference of the distance between the centres of two given circles are independent from the radius of the tangential circles, those are permanent. Thus, the hyperbola's sections belonging to the three circles in pairs are giving the centres of the searched circles.

It should be noted that Newton (1687) has further simplified the Roomen method: he has traced it back to the geometrical place of the centre of the searched tangency circles to the intersection of a point and a circle (trilateral problem).

2.4. Algebraic solution

Of course the Apollonius problem can be approached in an algebraic way. The baseline is the equation system which is arising from the coordinate-geometry equations of the given circles. With the help of the resultant we can reach the quadratic equation system solutions, with the analysis of those the number of Apollonius problem solutions can be analysed (two real radicals, coinciding radicals, conjugated and complex solutions ...).

2.5. Solution in three dimensions

If we place the problem from plane to three dimension then we arrive to a general solution of the problem with the help of spheres and cones. We will be dealing with this solution method in more detail later in chapter 3.

2.6. Gergonne type of solution

If three circles are given, draw up their exterior and interior similarity points. Connect them with straight lines. We obtain four straight lines. Select one of them and determine the inverse of those points for each circle which are the closest to the circles. (Let's take the pole of this straight line at the polar connection point for each circle). Then draw up the mutual power points which are determined by power lines of the three circles. This is the centre point of that circle which is perpendicularly intersecting all three circles. Let's connect this with the previously given poles. These three lines are intersecting the three circles in two-two points these are going to be the tangency points. The three points out of the six needs to be chosen in a way as the lines have separated the circles. The point with a different characteristic needs to be chosen on that particular circle which is on the side of the straight line, then the other two with similar characteristics since they are positioned on the other side of the chosen straight line. The centre of the tangency circle can be drawn up from this point. Since there are four straight

lines, and we always arrive two tangency circles using this method we can always have all eight tangency circles.

2.7. The use of GeoGebra

The inclusion of a DGS into the geometry education is almost natural nowadays. In case of the just examined problem it is especially true that it makes our job easier especially in the field of discussion. On a sheet of paper, we may choose any draw up methods; it is going to be difficult to follow the actual steps due to the lot of auxiliary lines. Not to mention if we would like to draw up all solutions on the paper and how we may want to add the 3 objects, so all the solutions may be visible on the paper.

The following GeoGebra worksheet contains solutions for all 10 cases as a built in tool. (Figure 2):

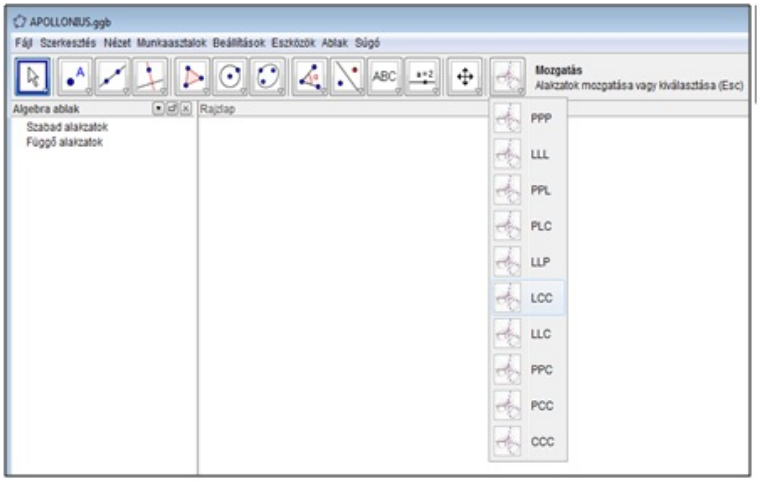


Figure 2: Solutions of the Apollonius problems as a built in tool

Here you simply must select the particular case, after giving the input parameters the tangency circles are going to be drawn. Parallel to the dynamic change of the given object it will dynamically change the placement of the tangent circles, giving an opportunity for a rapid discussion (Figure 3).

The before mentioned GeoGebra worksheet achieves the solutions for the first 9 problems with elementary draw ups. In case of the 10^{th} problem the solution is achieved by the help of hyperbola sections. Each macro tool can be opened from the menu, so with the function of replay the detailed draw up step sequence and the execution can be viewed.

Draw up executed by the way of inversion can be carried out similarly; the inversion has been built in as a tool into the GeoGebra worksheet for easier handling.

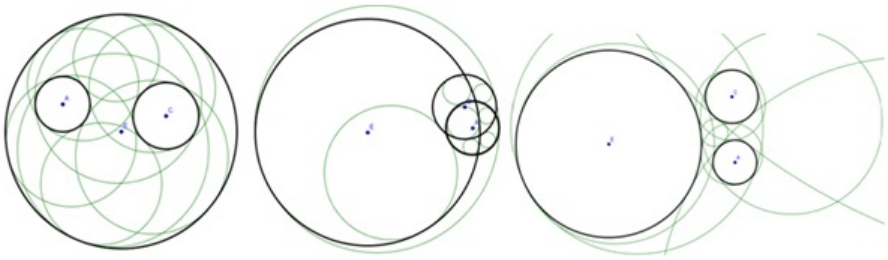


Figure 3: Given tangency circles in a given arrangement

3. Theoretical background of the three dimensional geometry solutions

One reasonable and advantageous problem solving mean to tackle plane geometrical draw ups is the index number portrayal (lative projection), which is one of the tool of cyclography. Cyclography was born with the help of the Apollonius type problem related search as simple and unified theoretical problem solution.

With the help of cyclographyc mapping lots of plane geometrical problems can be solved which would be very difficult to solve in the plane. So the problems are solve with the help of three dimensional objects, then the solutions are transformed back to the plane.

Let's review briefly the cyclographyc mapping. Cyclography is not a type of linear mapping, where mutual definite relation is established between the points of three dimension and the plane directed circles. In cyclography in the plane of the drawing corresponds with the point of any of its circle which is in the centre of the circle on the perpendicular line of the plane of the drawing and is away from the centre of the circle with the distance of the radius. This way each circle corresponds with two points on the opposite side of the plane of the drawing. For clarity, we provide direction for the circle and it will be positive (counter clockwise) or negative (clockwise) in accordance with whether it is placed above the drawing of the plane of the corresponding point or below. The directed circles are called cycles after Laguerre. The image plane's directed lines are called spreads. This can be interpreted as a cycle with an infinite radius. We need to handle the point as a cycle with zero radius. For every cycle and its corresponding point in three dimensions a 45° angular aperture size rotating cone can be fitted. We call this cone C cone (in case of a straight line it has a 45° angular aperture to the image plane, can be treated also as a plane which fits to a given straight line).

Solving the Apollonius problem is getting simplified if we direct the circles and work with cycles this way. Then we search for cycles which are tangent to the three cycles. We know that the solution is given by the cyclographyc image of the mutual points of C cone belonging to the three cycles. The problem is thus lead back to the determination of the three C cone interpenetration effect.

Let us examine briefly the location of two different parallel rotating cones, their interpenetration effects as well as their common points. The interpenetration effect of the second-order cone is usually a fourth-order three dimensional curve. We also know that if two cones have two mutual tangential planes then the interpenetration effect will create two cone sections. In any case the parallel rotating cone has two mutual tangential planes which of course fit to the line which is connecting the vertex points of the cones. Therefore the interpenetration effect is always two cone sections. The question is where can we find these cone sections?

From depicting geometrical knowledge we can conclude the following: the three parallel rotating cones power plane defined in pairs intersect each other in one line in the power line of the three cones and the power line also fits to the plane of the interpenetration affect of the cones. It is obvious that it can intersect the cones only on the scone sections (up to two points if there are not indefinitely many points in common). Therefore the two thrust points are common points of all three cones. The daw up of the common points of the cones, which follows from the foregoing, can be simplified with the draw up of an intersection of a line and a cone (Figure 4.).

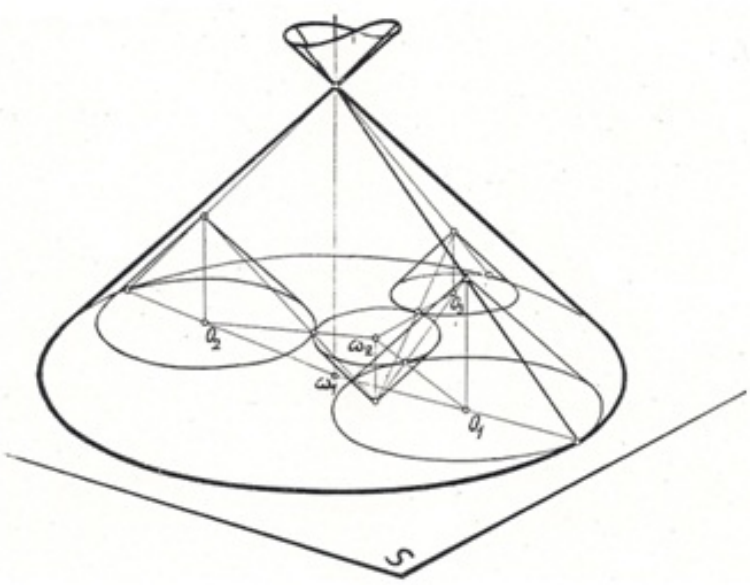


Figure 4: Common points of three cones

After this the general procedure for the actual draw up is the following:

1. We define the axis of the lineal cyclic congruence formed by the three cycles.
2. We draw up the power points of the three cycles (this will be the trace point of the straight multiplier).

3. We define the conjugate of the axis for some of the surface and with this we arrive to the three plane interpenetration affect of the intersection line.

4. With the given line status we draw a parallel through the power point, this way we arrive to the intersection line.

5. We thrust any of the cones with the intersection line. The cyclographyc images of the thrust points give the solution to the problem in case of directing the circles one way.

The original Apollonius problem has 8 solutions. By the way of cyclographic solutions it is as follows: We provide directions to the given three circles which will define by cycle one-one C cone. On these three cones we execute the prior draw up. We may give directions to the three circles in total eight different ways, which may give in total 16 solutions. But considering the circles of solutions 8 solutions are to be considered. For example if the directing K1, K2, K3 circles by turns are +, +, - and -, -, + then the three dimensional elements used in the draw up (planes, cones etc.) are symmetrical to the plane of the drawing. The solution cycles are directed in an opposite way, but give identical circles. In (2) we can see detailed guidelines for all 10 cases.

4. The realization of three dimensional geometric solutions with GeoGebra

Why exactly GeoGebra?

On the internet there are two worksheets published, which solved the problem with cote projection mentioned in the previous chapter.

One is named Geometers sketch pad drawn up in DGS which is a paid service therefore a private citizen or an educational institution cannot access it free. The other was prepared in Carbi which can be viewed in a time and tool limited trial version. It only contains on case the problem of the three circles. In today's Hungarian education systems it is a very important point of view (if not the most important) that the programs which are used would be free for the institution, teachers and students alike, due to the limited (financial) possibilities.

Taking into consideration the above mentioned shortcomings GeoGebra worksheets have been prepared (Figure 5) which can be accessed at

`http://geogebratube.com/student/m18934`

URL and it is available and can be downloaded by anybody for free of charge.

Before discussing the possibilities of the program let's have few words about the execution of the draw up. Since the steps of the draw up were rather general in the previous chapter, regarding the specific draw up executions in GeoGebra, for example looking at the three circles case we can simplify the draw ups by the following specifications. Let's take the three given circles and consider it to one-one

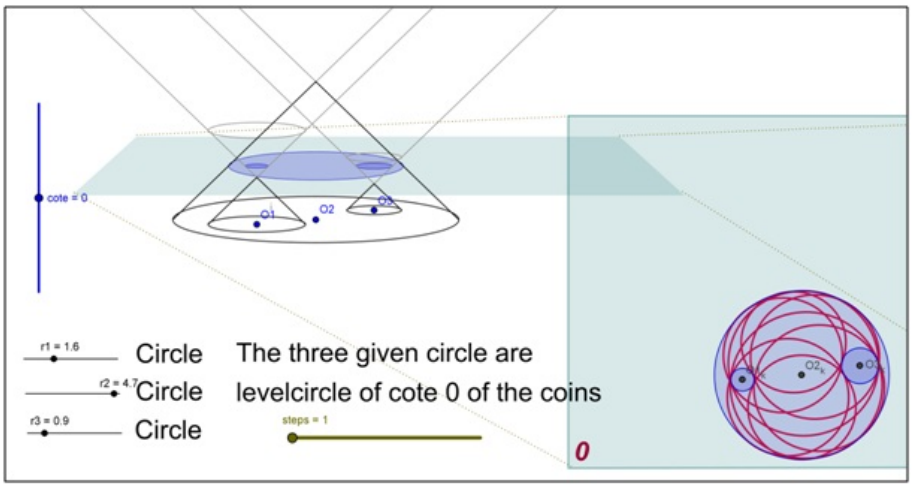


Figure 5: Apollonius problem with cote projection in GeoGebra

cone's level circle with the cote being 0. Let's increase the radius of all three circles with the same interval; in this case in order to simplify the draw up let's have the O2 cantered circle's radius be the benchmark. Care must be taken during the draw up so the level circles would intersect each other in two-two points (in case we plan that the circles should be increasingly drifting away from each other, then it is better to take higher cote). With this interval we draw the level circles of the cones with 1-2 cotes, then we draw up the mutual power points of 1-2 level circles, which give us the 1-2 cote points of the power line's of the three cones.

The further assignment is to draw up of the power line's thrust point with some of the cones. In our case this is the cone which is fitting on the O2 level circle in order to draw simpler and less lines as just indicated before. The cote of the cone's vertex is 1, so we just mete the divisor section of the line parallel to the power line. The level line going through the 0 level point of the two straight lines out sects the base points of the two cone elements, the element also out sect from the searched power lines the thrust points, which are the geometrical places for the centres of the tangency circles.

Two cones belong to each circle and two planes fit on each straight line, which form symmetric pairs to the plane of the drawing.

The program contains two different views: one is the cones and their intersecting plane (that is parallel to the plane of the base circles); the other is perpendicularly showing from the top the actual status of the objects.

The plane can be moved with the slide called cote, this way it out sects level circles with different cote from the cones. The r1, r2 radiuses of the cones base circle are fitting to the base plane and can be set with a slide between 0 and 5.

The degenerated point circles arising from the level circles can be given in two different ways: If we set the radius to 0 or if the image plane is fitting exactly on

the vertex of the cone. We can arrive to a straight line if we move the image plane below the base plane. In theory the plane would need to move on the symmetrical oppositely directed cone, but since the radius cannot be increased infinitely (as in information technology infinite does not exist) therefore we simulate the circle with infinite radius this way. By the radius slides we can see the actual status of the objects in text display form.

The other view is to show the actual solutions; here we can see in an appropriate way the actual cotes of the level circles in accordance with the objects (point, circle, line). Here it should be mentioned that we have a very difficult job to do as far as drawing up techniques, if we want to arrive to a completely general and dynamic solution. Since it "costs a lot" that the objects can be located anywhere in relation to each other and the user can move them any way he/she pleases. The draw ups needed to be executed in several situations (for example, whether two pints are located inside or outside of the circle) or rather a relatively complex criteria system needed to be programmed to solve visibility issues.

Since to show the solutions for each scenario is macro programmed (since GeoGebra could not not have handled safely the lots of objects and calculations on the worksheet) this way the built in function which plays back the draw up steps one by one cannot be used. Thus with the help of a slide called steps (if the conditions are right) we can follow the actual draw up steps (Figure 6.) shown with text descriptions and also with animation.

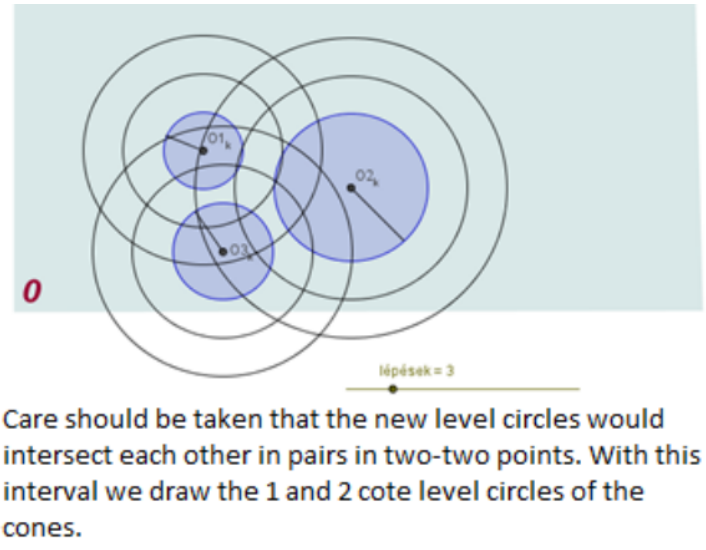


Figure 6: Presentation of the draw up steps

In my opinion the worksheet is suitable to present the Apollonius problem and to illustrate its general and elegant solutions as well as raise the demand for interest with reference to the topic.

References

[1] Makláři, J., Érintő körök szerkesztése sík-, és térmértani megoldással, illetve geometriai transzformációval, *Tankönyvkiadó*, 1963.

[2] Bácsó, S., Papp, I., Ciklográfiai példatár, *Debreceni Egyetem*, (jegyzet), 2006.

[3] Cranz, H., Das Apollonische berührungsproblem in stereographischer projektion, 1907.

[4] Hunyadi, J., Apollonius feladata a gömbfelületen, *Bp.*, 1877.

[5] `http://matematika.belvarbcs.hu/apollonius/index.htm` (2012-10-31).

[6] Ripco Sipos, E., A geometria tanítása számítógép alkalmazásával, *doktori (PhD) értekezés*, 2011.

[7] `http://salat.web.elte.hu/VIM/modules/apol_mod/apol_mod.htm` (2012-10-31).

[8] `http://mathworld.wolfram.com/ApolloniusProblem.html` (2012-10-31).

Annales Mathematicae et Informaticae
***40** (2012) pp. 175–186*
http://ami.ektf.hu

Rotation about an arbitrary axis and reflection through an arbitrary plane

Emőd Kovács

Department of Information Technology
Eszterházy Károly College
emod@ektf.hu

Submitted April 22, 2012 — Accepted November 7, 2012

Abstract

The aim of this paper is to give a new deduction of Rodrigues' rotation formula. An other benefit of the this deduction is to give a transformation matrix of reflection through an arbitrary plane with the same deduction method. In our opinion this deduction method is better for students, who are learning computer graphics.

Keywords: Point transformation, Transformation Matrix, Rotation, Reflection, Rodrigues' rotation formula,

MSC: Primary 68U05, Secondary 65D18

1. Introduction

In the theory of three-dimensional (3D) rotation Rodrigues' rotation formula (see [7]) is an efficient matrix for rotating an object around arbitrary axis. In this paper we will deduct the matrix form in a different way from the well known method which is published in Rodrigues' paper [7], cited in Johan's paper [6] and also described in Wolfram Mathworld site (see [1]). First we give a short introduction of linear point transformation, then we inroduce a new deduction of reflection about an arbitrary axis. Next, we will prove, that our matrix is analogous to the original Rodrigues' formula. In section three, we describe a matrix of reflection through an arbitrary plane, which is a consequence of our deduction.

1.1. Linear Point Transformation

Three dimensional point transformation is one of the well known computer graphics methods, when we manipulate the points of objects, like rotate, translate and scale. Based on the advantages of homogeneous coordinates, 3D transformations can be represented by 4×4 matrices (see [2] and [3]). Generally the following matrix equation describes the point transformation.

$$\begin{aligned} \mathbf{p}' &= \mathbf{M} \cdot \mathbf{p}, \\ \begin{bmatrix} x_1' \\ x_2' \\ x_3' \\ x_4' \end{bmatrix} &= \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}. \end{aligned} \tag{1.1}$$

When we use more 3D transformations after each other, it is constructed of matrix multiplications (see [8]), therefore composition of 3D transformation can be represented by the multiplication of transformation matrices. The order of multiplication depends on the original form of matrix equation, since the matrix multiplication was noncommutative operation. If we multiply the point from left with the transformation matrix in Eq. (1.1), then we must multiply the transformation matrices in reverse order. Lets $\mathbf{M}_1$ the first, $\mathbf{M}_2$ the second transformation matrix, then

$$\mathbf{p}' = \mathbf{M}_1 \cdot \mathbf{p}, \quad \mathbf{p}'' = \mathbf{M}_2 \cdot \mathbf{p}'.$$

Using the associative property it becomes

$$\mathbf{p}'' = \mathbf{M}_2 \cdot (\mathbf{M}_1 \cdot \mathbf{p}) = (\mathbf{M}_2 \cdot \mathbf{M}_1) \cdot \mathbf{p}.$$

Therefore we multiply the matrices in reverse order $\mathbf{M}_3 = \mathbf{M}_2 \cdot \mathbf{M}_1$, from that

$$\mathbf{p}'' = \mathbf{M}_3 \cdot \mathbf{p}.$$

If the point is multiplied by the transformation matrix from the right, then it means the equivalent system. Some graphics library, e.g. DirectX use the latter method, in this case these system use the transposed matrices.

In this paper we deal with the general case of rotation about an arbitrary axis in space. It frequently occurs e. g. in robotics, animation and simulation.

2. Rotation about an arbitrary axis

If we want to construct rotation about an arbitrary axis, then we have a good solution namely Rodrigues' rotation formula, see [1] on Wolfram MathWorld site. Lots of literatures and internet sources give this method. The problem is that the mathematical deduction is not suited for the previous section from methodical aspect. Lots of students could not understand the mathematical deduction of Rodrigues' formula, which is presented on the Wolfram MathWorld site, and therefore some

of them could not use it. When we teach basic point transformations and we try to extend it towards the composition of 3D transformations, then it could be a good example about the rotation about an arbitrary axis. It would be better if we can give the Rodrigues' rotation matrix with the composition of basic linear point transformations, and apply multiplication of transformation matrices. In this paper we deduce the rotation matrix and prove the computed matrix is an equivalent of the Rodrigues' formula. Anyone can find the deduction in Rogers's textbook [8], but now we continue the computation.

The basic idea is to make the arbitrary rotation axis coincide with one of the coordinate axis. Assume an arbitrary axis in space passing through the point $P_0\,(x_0, y_0, z_0)$ and $P_1\,(x_1, y_1, z_1)$.

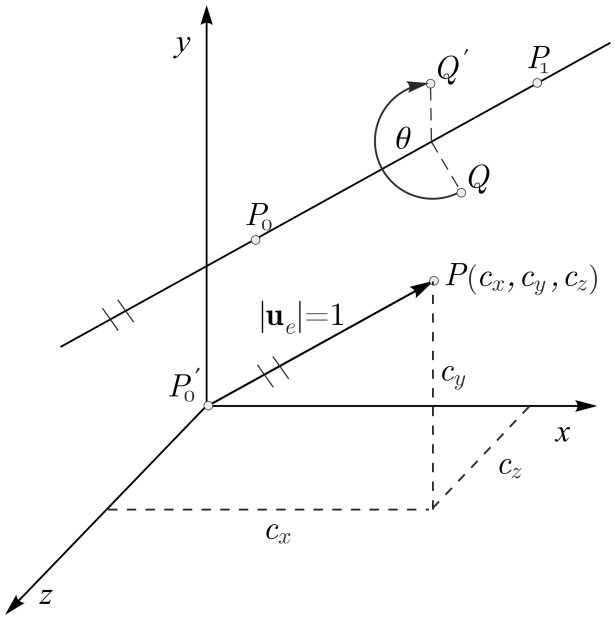


Figure 1: Rotation about an arbitrary axis

In this case rotation about this axis by some angle θ is accomplished using the following procedure:

1. Translate the $P_0\,(x_0, y_0, z_0)$ axis point to the origin of the coordinate system.
2. Perform appropriate rotations to make the axis of rotation coincident with z-coordinate axis.
3. Rotate about the z-axis by the angle θ.
4. Perform the inverse of the combined rotation transformation.
5. Perform the inverse of the translation.

For the simplicity we compute the $\mathbf{u} = P_1 - P_0$ vector, which after the normalization can give us the direction cosines of axis:

$$\mathbf{u_e} := \frac{\mathbf{u}}{|\mathbf{u}|} = (c_x, c_y, c_z)\,.$$

In Fig. 2 the direction cosines are satisfied the following equation:

$$c_x^2 + c_y^2 + c_z^2 = 1,$$
$$\cos\phi_x = c_x, \quad \cos\phi_y = c_y, \quad \cos\phi_z = c_z.$$

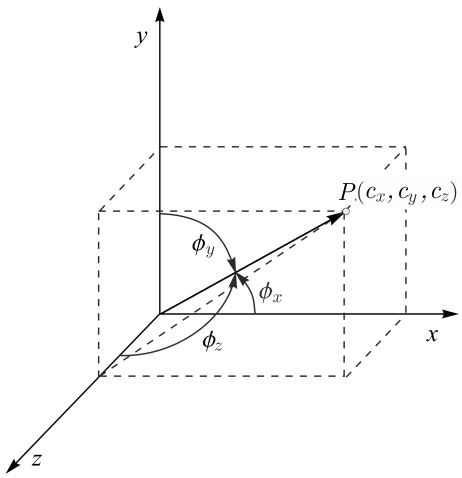


Figure 2: Direction cosines

The required translation matrix is

$$\mathbf{T}(-\mathbf{p}_0) = \begin{bmatrix} 1 & 0 & 0 & -x_0 \\ 0 & 1 & 0 & -y_0 \\ 0 & 0 & 1 & -z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In the next step the procedure requires two successive rotation about the x-axis by the angle θ_x and y-axis by the angle θ_y. After the rotation around the the x-axis the original rotation axis will be in the $[x, z]$ coordinate pane. (See Fig. 3).

From the Fig. 3 comes $d = \sqrt{c_y^2 + c_z^2}$, and we do not calculate explicitly the angle θ_x, because we only use its sin and cosine values in the rotation matrix:

$$\sin\theta_x = \frac{c_y}{d}, \quad \cos\theta_x = \frac{c_z}{d}.$$

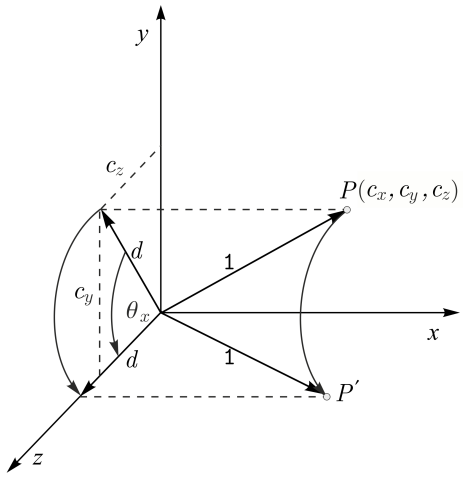


Figure 3: Rotation around x-axis

The rotation matrix is

$$\mathbf{R}_x(\theta_x) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c_z/d & -c_y/d & 0 \\ 0 & c_y/d & c_z/d & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.1}$$

We can get the second rotation matrix in a similar way, where we rotate around the y-axis by angle θ_y.

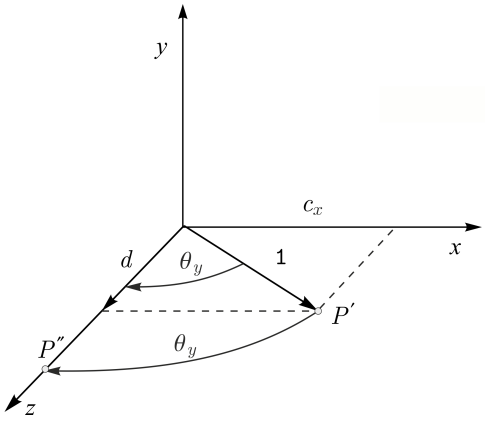


Figure 4: Rotation around y-axis

From the Fig. 4 comes

$$\sin\theta_y = d, \quad \cos\theta_y = d.$$

The rotation matrix is with negative direction

$$\mathbf{R}_y(-\theta_y) = \begin{bmatrix} d & 0 & -c_x & 0 \\ 0 & 1 & 0 & 0 \\ c_x & 0 & d & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.2}$$

The complete transformation is

$$\mathbf{M} = \mathbf{T}^{-1}(-\mathbf{p}_0)\mathbf{R}_x^{-1}(\theta_x)\mathbf{R}_y^{-1}(-\theta_y)\mathbf{R}_z(\theta)\mathbf{R}_y(-\theta_y)\mathbf{R}_x(\theta_x)\mathbf{T}(-\mathbf{p}_0), \tag{2.3}$$

where the upper index -1 means the inverse transformation, so

$$\mathbf{M} = \mathbf{T}(\mathbf{p}_0)\mathbf{R}_x(-\theta_x)\mathbf{R}_y(\theta_y)\mathbf{R}_z(\theta)\mathbf{R}_y(-\theta_y)\mathbf{R}_x(\theta_x)\mathbf{T}(-\mathbf{p}_0), \tag{2.4}$$

where we used the reverse multiplication order as we mentioned in the previous section. The computation is finished at this point in Rogers's textbook [8].

Now we are giving one of our new results, and in the section 2.1 we are proving the formulas with Maple computer algebra system.

The formula in (2.3) can be enough, if someone only use the basic transformation matrices in the matrix class of the graphics engine. But methodically for the better understandability and based on our students searching practice in internet literature, we must continue the calculation.

Let multiply the inside five matrices

$$\mathbf{R} = \mathbf{R}_x(-\theta_x)\mathbf{R}_y(\theta_y)\mathbf{R}_z(\theta)\mathbf{R}_y(-\theta_y)\mathbf{R}_x(\theta_x), \tag{2.5}$$

and

$$\mathbf{M} = \mathbf{T}(\mathbf{p}_0)\mathbf{R}\mathbf{T}(-\mathbf{p}_0).$$

Consider that the inverse of rotation matrix equals with the transposed matrix, we get

$$\mathbf{R} = \mathbf{R}_x^T(\theta_x)\mathbf{R}_y^T(-\theta_y)\mathbf{R}_z(\theta)\mathbf{R}_y(-\theta_y)\mathbf{R}_x(\theta_x). \tag{2.6}$$

In the next section we are going to prove that if we expand the matrix multiplication in Eq. (2.5), then we get the general Rodrigues' form. In [9] or in [1] we can find the totally different deduction of the Rodrigues' form, but as we mentioned we are not satisfied the authors deduction way, therefor we give a new solution. The Maple CAS is very robust and efficient tool for calculating multiplication of transformation matrices.

2.1. Proof with Maple

The main problem of the proof is that multiplication of five matrices in Eq. (2.6). In order to correct calculation we used Maple computer algebra system (CAS). In [4] and [5] the author explains why Maple is a useful tool for teaching computer graphics in higher education. In Eszterházy Károly College we use CAS software in teaching undergraduate students studiing Software Information Technology bachelor course.

We can use the power of the linalg package of Maple, to easily multiply the five matrices.

The Maple command is

> rod:=simplify(Transpose(RX).Transpose(RY).RZ.RY.RX);

where we used that the transposed rotation matrix equals the inverse of the rotation matrix, and the "rod" means the Rodrigues' form. After we used the built-in *simplify* function we got the output in Fig. 5.

$$
\begin{aligned}
rod := \Big[\Big[& cy^2\cos(\theta) + cz^2\cos(\theta) + cx^2, -\sin(\theta)\,cz - cy\,cx\cos(\theta) \\
& + cy\,cx, \sin(\theta)\,cy - cz\,cx\cos(\theta) + cz\,cx, 0\Big], \\
& \Big[-cy\,cx\cos(\theta) + \sin(\theta)\,cz + cy\,cx, \\
& \frac{cz^2\cos(\theta) + cy^2\,cx^2\cos(\theta) + cy^4 + cz^2\,cy^2}{cy^2 + cz^2}, \\
& -\frac{1}{cy^2 + cz^2}\big(cx\,cy^2\sin(\theta) + cy\,cz\cos(\theta) - cz\,cy\,cx^2\cos(\theta) \\
& + cz^2\,cx\sin(\theta) - cz\,cy^3 - cz^3\,cy\big), 0\Big], \\
& \Big[-cz\,cx\cos(\theta) - \sin(\theta)\,cy + cz\,cx, \\
& \frac{1}{cy^2 + cz^2}\big(cz^2\,cx\sin(\theta) - cy\,cz\cos(\theta) + cz\,cy\,cx^2\cos(\theta) \\
& + cx\,cy^2\sin(\theta) + cz\,cy^3 + cz^3\,cy\big), \\
& \frac{cy^2\cos(\theta) + cz^2\,cx^2\cos(\theta) + cz^2\,cy^2 + cz^4}{cy^2 + cz^2}, 0\Big], \\
& \Big[0, 0, 0, 1\Big]\Big]
\end{aligned}
$$

Figure 5: First result in Maple

The computed formula is extremely complicated. So we must look for other

simplification possibilities. We can use the combination of *simplify* and *substitution* functions repeatedly:

```
for i from 2 to 3 do
 for j from 2 to 3 do
  rod[i,j]:=simplify(subs({cx^2=1-(cy^2+cz^2)},rod[i,j]));
 od;
od;
```

We can see the output in Fig. 6. The *collect* function was used many times, which

$$\begin{aligned}&\left[\left[cy^2\cos(\theta)+cz^2\cos(\theta)+cx^2,-\sin(\theta)\,cz-cy\,cx\cos(\theta)\right.\right.\\&\quad\left.+cy\,cx,\sin(\theta)\,cy-cz\,cx\cos(\theta)+cz\,cx,0\right],\\&\quad\left[-cy\,cx\cos(\theta)+\sin(\theta)\,cz+cy\,cx,-cy^2\cos(\theta)+cy^2\right.\\&\quad\left.+\cos(\theta),cz\,cy-cy\,cz\cos(\theta)-cx\sin(\theta),0\right],\\&\quad\left[-cz\,cx\cos(\theta)-\sin(\theta)\,cy+cz\,cx,cz\,cy-cy\,cz\cos(\theta)\right.\\&\quad\left.+cx\sin(\theta),\cos(\theta)-cz^2\cos(\theta)+cz^2,0\right],\\&\quad\left.[0,0,0,1]\right]\end{aligned}$$

Figure 6: After the simplification

collected coefficients. One of them is the following:

```
rod[1,1]:=collect(rod[1,1],cos(theta));
```

After we use the $cy^2+cz^2=1-cx^2$ equation we get better form. In this paper we do not give the total Maple worksheet. The reader can download it from the following link:
http://aries.ektf.hu/~emod/mapleporoof.html

Finally we got the following result:

$$\begin{bmatrix}\cos\theta+c_x^2(1-\cos\theta) & c_xc_y(1-\cos\theta)-c_z\sin\theta & c_xc_z(1-\cos\theta)+c_y\sin\theta & 0\\ c_yc_x(1-\cos\theta)+c_z\sin\theta & \cos\theta+c_y^2(1-\cos\theta) & c_yc_z(1-\cos\theta)-c_x\sin\theta & 0\\ c_zc_x(1-\cos\theta)-c_y\sin\theta & c_zc_y(1-\cos\theta)+c_x\sin\theta & \cos\theta+c_z^2(1-\cos\theta) & 0\\ 0 & 0 & 0 & 1\end{bmatrix}.$$

Obviously the result is analogous with the Rodrigues' formula in the MathWorld sites. (http://mathworld.wolfram.com/RodriguesRotationFormula.html)

Over 90% of the built-in commands in maple are programmed in Maple's own Pascal-like programming language. Beside this Maple also give exporting facilities to other programming languages. For example the C command translates the Maple pretty output to ANSI C code. The result was converted to C code in optimized form. When we develop new application, then with the help of "copy paste method"

we can put the code easily into the our C, C++, C# or Java program code. Maple command:

```
C(rod,optimize);
```

Maple output in C:

```
t1  = cos(theta);     t2 = -t1 + 0.1e1;   t3 = cx * cx;
t7  = t2 * cy * cx;   t8 = sin(theta);    t9 = t8 * cz;
t11 = t2 * cz;        t12 = t11 * cx;     t13 = t8 * cy;
t16 = cy * cy;        t19 = t11 * cy;     t20 = cx * t8;
t24 = cz * cz;

cg0[0][0] = t2 * t3 + t1;   cg0[0][1] = t7 - t9;
cg0[0][2] = t12 + t13;      cg0[0][3] = 0.0e0;

cg0[1][0] = t7 + t9;        cg0[1][1] = t2 * t16 + t1;
cg0[1][2] = t19 - t20;      cg0[1][3] = 0.0e0;

cg0[2][0] = t12 - t13;      cg0[2][1] = t19 + t20;
cg0[2][2] = t2 * t24 + t1;  cg0[2][3] = 0.0e0;

cg0[3][0] = 0.0e0;          cg0[3][1] = 0.0e0;
cg0[3][2] = 0.0e0;          cg0[3][3] = 0.1e1;
```

3. Reflection through an arbitrary plane

It is often necessary to reflect an object through an arbitrary plane other than one of the coordinate planes like $x = 0$, $y = 0$ and $z = 0$. We can deduct the transformation matrix similar to what was described in the previous section. The basic idea is to make the arbitrary reflection plane coincide with one of the coordinate planes. Assuming an arbitrary plane in space is given by three points $P_0\,(x_0, y_0, z_0)$, $P_1\,(x_1, y_1, z_1)$ and $P_2\,(x_2, y_2, z_2)$, these points are noncollinear.

According to the previous section one possible procedure is:

1. Translate the reflection plane to the origin of the coordinate system with the help of known $P_0\,(x_0, y_0, z_0)$ point.
2. Perform appropriate rotations to make the normal vector of the reflection plane at the origin until it coincides with the $+z$-axis (see Eqs. (2.1) and (2.2)); this makes the reflection plane the $z = 0$ coordinate plane.
3. After that reflect the object through the $z = 0$ coordinate plane.
4. Perform the inverse of the combined rotation transformation in step 2.
5. Perform the inverse of the translation in step 1.

From the three points of the reflection plane is easy to calculate the normal vector by the crossproduct of $\mathbf{a} = P_1 - P_0$ and $\mathbf{b} = P_2 - P_0$ vectors:

$$\mathbf{n} = \mathbf{a} \times \mathbf{b}.$$

For simplicity we normalize the normal vector, which can give us the direction cosines similar to the previous section:

$$\mathbf{n_e} := \frac{\mathbf{n}}{|\mathbf{n}|} = (c_x, c_y, c_z)\,.$$

In step 2 the rotation matrices will be the same that were used during the rotation about an arbitrary axis. Finally the seven transformation matrices were multiplied in reverse order:

$$\mathbf{M} = \mathbf{T}(\mathbf{p}_0)\mathbf{R}_x(-\theta_x)\mathbf{R}_y(\theta_y)\mathbf{R}_{eflect}^{(x,y)}\mathbf{R}_y(-\theta_y)\mathbf{R}_x(\theta_x)\mathbf{T}(-\mathbf{p}_0).$$

Similar to the previous section we multiply the five inside matrices:

$$\mathbf{R}_{eflect} = \mathbf{R}_x(-\theta_x)\mathbf{R}_y(\theta_y)\mathbf{R}_{eflect}^{(x,y)}\mathbf{R}_y(-\theta_y)\mathbf{R}_x(\theta_x), \tag{3.1}$$

then

$$\mathbf{M} = \mathbf{T}(\mathbf{p}_0)\mathbf{R}_{eflect}\mathbf{T}(-\mathbf{p}_0). \tag{3.2}$$

The general reflection matrix in Eq. (3.1) has no special name in the textbooks. Similarly to the previous section we use the Maple computer algebraic system to give a simplified form of the transformation matrix:

$$\mathbf{R}_{eflect} = \begin{bmatrix} 1-2c_xc_x & -2c_yc_x & -2c_zc_x & 0 \\ -2c_yc_x & 1-2c_yc_y & -2c_yc_z & 0 \\ -2c_zc_x & -2c_yc_z & 1-2c_zc_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

since the inverse of reflection is the same with itself $\mathbf{R}_{eflect}^{-1} = \mathbf{R}_{eflect}$. Good choice to perform the multiplications in Eq. (3.2), because the result is not complicated:

$$\mathbf{R}_{eflect=} \begin{bmatrix} 1-2c_x^2 & -2c_xc_y & -2c_xc_z & -2c_xd \\ -2c_xc_y & 1-2c_y^2 & -2c_yc_z & -2c_yd \\ -2c_xc_z & -2c_yc_z & 1-2c_z^2 & -2c_zd \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3.3}$$

where the $d = -c_xx_0 - c_yy_0 - c_zz_0$. The previous matrix is a good result from the methodological aspect, because the student also can find the similar version in the D3DXMatrixReflect function of DirectX:

$$P = normalize(Plane);$$

$$\begin{bmatrix} -2*P.a*P.a+1 & -2*P.b*P.a & -2*P.c*P.a & 0 \\ -2*P.a*P.b & -2*P.b*P.b+1 & -2*P.c*P.b & 0 \\ -2*P.a*P.c & -2*P.b*P.c & -2*P.c*P.c+1 & 0 \\ -2*P.a*P.d & -2*P.b*P.d & -2*P.c*P.d & 1 \end{bmatrix},$$

where the matrix equals the matrix in Eq. (3.3) after transposition. The $normalize(Plane)$ function normalizes the normal vector of the plane. $P.a, P.b$ and $P.c$ are the coefficients in the plane equation, and obviously the coordinates of normalized normal vector as well (see [10]):

$$P.a * x + P.b * y + P.c * z + P.d = 0,$$

and $P.d$ contains the one point $P(x_0, y_0, z_0)$ of the plane as we mentioned above:

$$P.d = -P.a * x_0 - P.b * y_0 - P.c * z_0,$$

moreover $P.d$ comes from the dot product of the normal vector and P (see [10]). If our student use the deduction in the section 3, then the D3DXMatrixReflect DirectX function becomes understandable without difficulties.

4. Conclusion

In this paper we proposed a better deduction of Rodrigues' rotation formula than you can find in a lots of literature (e.g see [1]). In other text book you can find the similar deduction to our method (see [8]), but in this paper we continued the deduction and demonstrate the equality between the result and the Rodrigues' form. From the aspect of teaching computer graphics in higher education, our method is better understandable for the students. Moreover, our deduction comes naturally from the composition of 3D transformations. An additional benefit is that our method is a good solution to deduct transformation matrix of reflection through an arbitrary plane.

References

[1] Belongie, S., Rodrigues' Rotation Formula, From MathWorld –A Wolfram Web Resource, created by Eric W. Weisstein.
http://mathworld.wolfram.com/RodriguesRotationFormula.html

[2] Foley, J., van Dam, A.,Feiner, S., Hughes, J., Computer Graphics Principles and Practice, Addison-Wesley, 1996.

[3] Juhász, I., Számítógépi grafika és geometria, Miskolci Egyetemi Kiadó, 1993.

[4] Kovács, E., Using some mathematical program in computer graphics teaching, 7th ICECGDG Cracow. International Conference on Engineering Computer Graphics and Descriptive Geometry July 18–22. 1996. Conference Proceedings Volume 2 p. 546–549.

[5] Kovács, E., Using Maple in teaching of computer graphics, International Conference on Applied Informatics, Eger 1995. Conference Proceedings, p. 83–92.

[6] Mebius, J. E., Derivation of the Euler-Rodrigues formula for three-dimensional rotations from the general formula for four-dimensional rotations., arXiv General Mathematics 2007. http://arxiv.org/abs/math/0701759

[7] Rodrigues, O., Des lois géométriques qui régissent les déplacements d'un systéme solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. Journal de Mathématiques 5, 1840, 380–440.

[8] Rogers, D. F., Adams, J. A., Mathematical elements for computer graphics, Second Edition, McGraw-Hill publishing Company, 1990.

[9] Watt, A., Policarpo, F., 3D Games: Real-Time Rendering and Software Technology, New York : ACM Press, 2001.

[10] Weisstein, Eric W., From MathWorld – A Wolfram Web Resource `http://mathworld.wolfram.com/Plane.html`

Annales Mathematicae et Informaticae
***40** (2012) pp. 187–201*
http://ami.ektf.hu

Engineering students' spatial abilities in Budapest and Debrecen

Rita Nagy-Kondor[a], Csilla Sörös[b]

[a]University of Debrecen, Debrecen, Hungary
rita@eng.unideb.hu

[b]Szent István University, Budapest, Hungary
soros.csilla@ybl.szie.hu

Submitted February 11, 2012 — Accepted October 13, 2012

Abstract

The goal of this paper is to provide the level of first-year engineering students' spatial abilities. We made our comparative survey at the Szent István University Budapest and University of Debrecen, Faculty of Engineering among first-year students of architecture. We made the survey among those students, who were successful in Descriptive Geometry I and II. We were looking for the answer for the question whether which part of the first-year architecture students' spatial ability and spatial geometrical knowledge is incomplete, and whether the students of two universities have sufficient differences between their spatial ability. The test results have been statistically evaluated and conclusions formulated.

Keywords: Spatial ability, Descriptive Geometry education

MSC: 51N05

1. Introduction

Spatial ability is very important for engineering students; it is decisive for their career. This ability is not determined genetically, but rather a result of a long learning process. [12] The definition of spatial ability according to Séra and his colleagues is [13] "the ability of solving spatial problems by using the perception of two and three dimensional shapes and the understanding of the perceived information and relations" - relying on the ideas of Haanstra [3] and others. McGee [7]

defines spatial ability as "the ability to mentally manipulate, rotate, twist or invert pictorally presented stimuli" and classifies five components of spatial skills as

- Spatial perception
- Spatial visualization
- Mental rotations
- Mental relations
- Spatial orientation

Maier [6] distinguishes five branches of spatial intelligence too:

- Spatial perception: the vertical and horizontal fixation of direction regardless of troublesome information;
- 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.

Vásárhelyi's [17] definition of geometrical spatial ability: the mathematically controlled complex unity of abilities and skills that allows: the exact conception of the shape, the size and the position of the spatial configurations; the unequivocal illustration of seen or imaginary configurations based on the rules of geometry; the appropriate reconstruction of unequivocally illustrated configurations; the constructive solution of different spatial (mathematical, technological) problems, and the imagery and linguistic composition of this solution.

In the classification of the exercises we followed the recommendation of Séra and his colleagues [13] who approach the spatial problems from the side of the activity. The types of exercises:

- Projection illustration and projection reading: establishing and drawing two dimensional projection pictures of three dimensional configurations (Task 3, Task 4, Task 5);
- Reconstruction: creating the axonometric image of an object based on projection images (Task 6);
- Transparency of the structure: developing the inner expressive image through visualizing relations and proportions;

- Two-dimensional visual spatial conception: the imaginary cutting up and piecing together of two-dimensional figures;
- Recognition and visualization of a spatial figure: the identification and visualization of the object and its position based on incomplete visual information;
- Recognition and combination of the cohesive parts of three-dimensional figures: the recognition and combination of the cohesive parts of simple spatial figures that were cut into two or more pieces with the help of their axonometric drawings;
- Imaginary rotation of a three-dimensional figure: the identification of the figure with the help of its images depicted from two different viewpoints by the manipulation of mental representations (Task 7);
- Imaginary manipulation of an object: the imaginary following of the phases of the objective activity (Task 1 and Task 2);
- Spatial constructional ability: the interpretation of the position of three-dimensional configurations correlated to each other based on the manipulation of the spatial representations;
- Dynamic vision: the imaginary following of the motion of the sections of spatial configuration.

The measurement of spatial abilities is standardized by international tests, among which the Mental Rotation Test (MRT) and the Mental Cutting Test (MCT) are of greatest importance. MRT is introduced by Vanderberg and Kuse [16]. MCT is widely used for testing the spatial ability at any level [14]. Németh and her colleagues [9, 10, 11] presented an analysis of MCT results of first-year engineering students, with emphasis on gender differences and attempted to find possible reasons of gender difference, concluding, that typical mistakes play central role in it.

In the second section we report about the circumstances of the survey. The third section contains the results of the survey and then we examine the most frequent mistakes. The last section is the summary of the article and our experiences.

2. The comparative survey

At the Faculty of Engineering at the University of Debrecen, the architecture students selected for the engineering program acquire the basics of the Descriptive Geometry - the elements of the Monge projection, axonometric representation, perspectivity - for a year, with two lectures and two seminars per week, which they use later in their professional subjects. The lecturer made two tests and four technical drawings for the students in all semesters. The Descriptive Geometry I, II end with exam mark.

At the Szent István University, Faculty of Engineering students of architecture need to do 3 semesters of Descriptive Geometry as a basic and compulsory subject. Geometry I and II are introduced in the first two semesters with one lecture and two seminars per week and thematically structured in the classical method which is familiar with the structure of the University of Debrecen. The lecturer made two tests and ten technical drawings for the students in Descriptive Geometry I; and one test, four technical drawings and two models in Descriptive Geometry II. The Descriptive Geometry I, II end with exam mark. Descriptive Geometry III is introduced in one semester with one seminar per week. We made the survey among those students who were successful in Descriptive Geometry I and II.

The short syllabus of the Descriptive Geometry I and II at our universities:
`http://www.eng.unideb.hu/userdir/mat/hallgatok/tantargy.html`
`http://asz.tanszek.ymmf.hu`

We made our comparative survey at the Szent István University Budapest and University of Debrecen, Faculty of Engineering among first-year students of architecture. At the university in Budapest 111 students, at the university in Debrecen 87 students took the test. The test took place on the last week to check the students' spatial abilities. The students had 60 minutes to complete the task sheet.

We prepared the test in a way that it contained the important components of spatial ability. Following the theory of Séra and his colleagues [13] we made the task sheet from the more important types of tasks.

The survey:

1. We have marked a cube on three sides with three different signs: /, V, X, leaving the other three sides empty. These cubes are the ones with a star next to them. Circle the cube from among the rest of the cubes that could be the same one in a different position. (Figure 1)

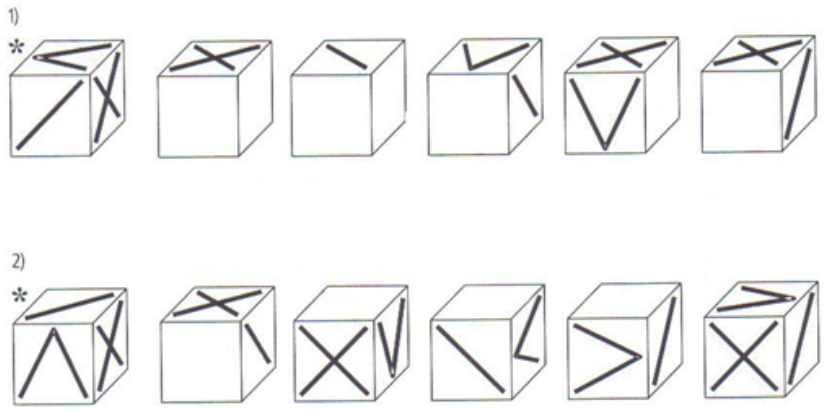


Figure 1

2. Next to the illustrations marked with a star you can see 5 objects, assembled of identical cubes, showing all the edges of the all the cubes, including the normally

hidden ones. Circle the object that could be rotated to fit through the hole in the one marked with a star. (Figure 2)

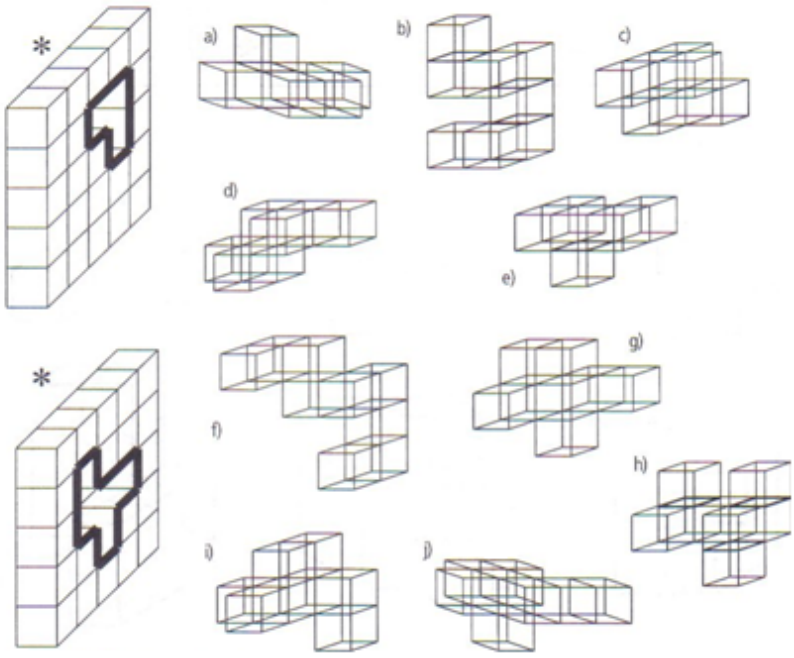


Figure 2

3. Draw the projections of the objects below, which have been cut out of cubes. Frontal view (E), Side view (O), Top view (F), based on the axonometric pictures of the objects. (Figure 3)

4. The exercises below each show two or three different perspectives of the same cable twisted into a certain shape. If we consider the first drawing as the front-view, then which view is the right side view, the left side view, back view or top view? (SZEMBŐL: Front) (Figure 4)

5. There is an axonometric picture of a wire framework built inside of a cube. The vertices of the figure are the same as the vertices of the cube or the midpoints of the sides of the cube. How can this figure be shown from the front, top, back, left and the right side? (Figure 5) (F: Top, E: Front, H: Back, J: Right, B: Left)

6. Reconstruct the solids by drawing the visible picture of it! Draw only the visible edges! (Figure 6) (F: Top, E: Front, O: Left)

Mental Rotation Test:

Of the four objects to the right which ones are identical to the original (to the left), rotated into another position? In each case there are two correct solutions. (Figure 7, Figure 8)

The first and second tasks focus on the imaginary manipulation of the solid. The task is to follow the phases of the objective activity that consist of the complex

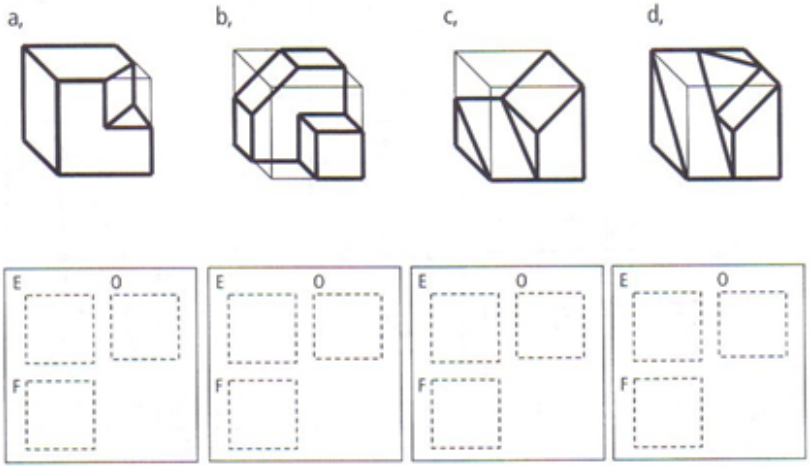


Figure 3

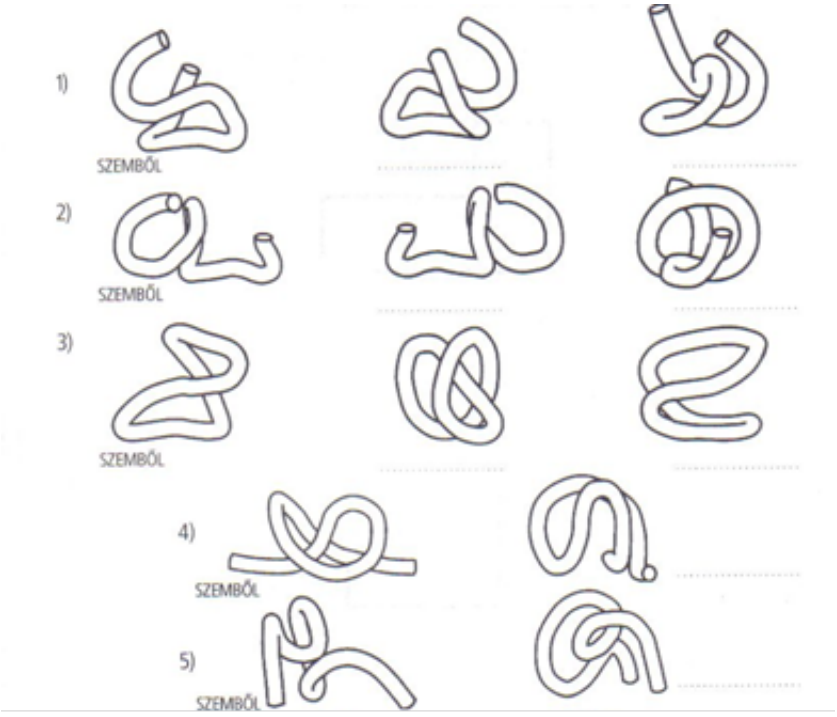


Figure 4

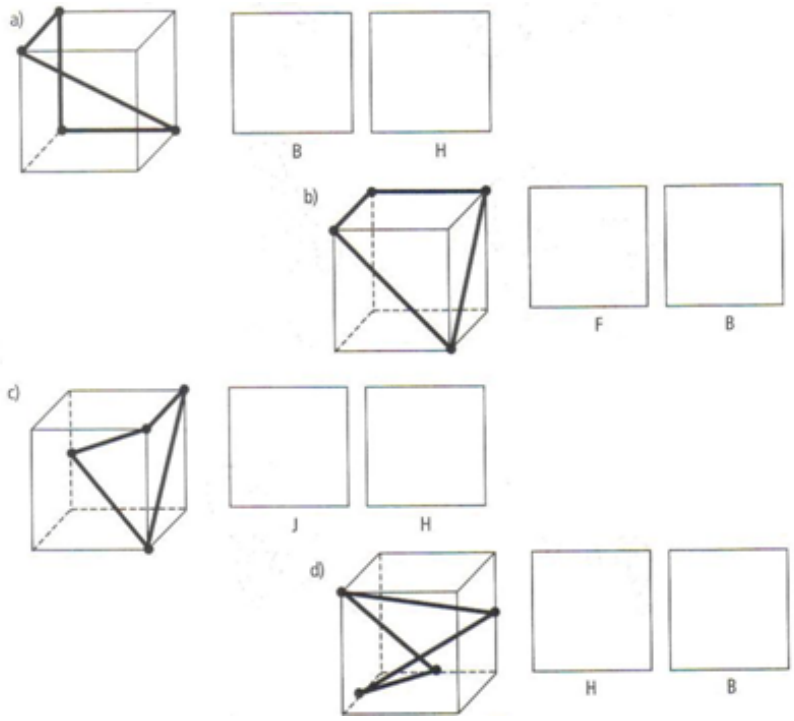


Figure 5

spatial transformation of the solid. The first task is the identification of the figure, and the second task is the manipulation of mental representations.

The third, fourth and fifth tasks belong to the types of tasks that deal with representation and reading of the projection. Mobilizing the experience of the motion, changing the inner viewpoint, imaginary rotation, manipulation of mental representations, and the task is to produce and draw the two-dimensional projection picture of a three-dimensional solid. This type of task is characterized by analytical operations from concrete to abstract.

The sixth task is a task of reconstruction. We have to create the axonometric picture of the solid based on the projected pictures. During the reconstruction the student synthesizes the visual information by studying the projected pictures. The map will be constructed by the series of changing the inner viewpoint by harmonizing three channels.

The last task is a sample of the MRT problem. Each problem is composed of a criterion figure, two correct alternatives and two incorrect alternatives. Correct alternatives are structurally identical to the criterion, but shown in a rotated position. The subjects are asked to find the two correct alternatives. The last task contains 5 MRT problems. Two points are given for a problem. The best possible score in the MRT is 10.

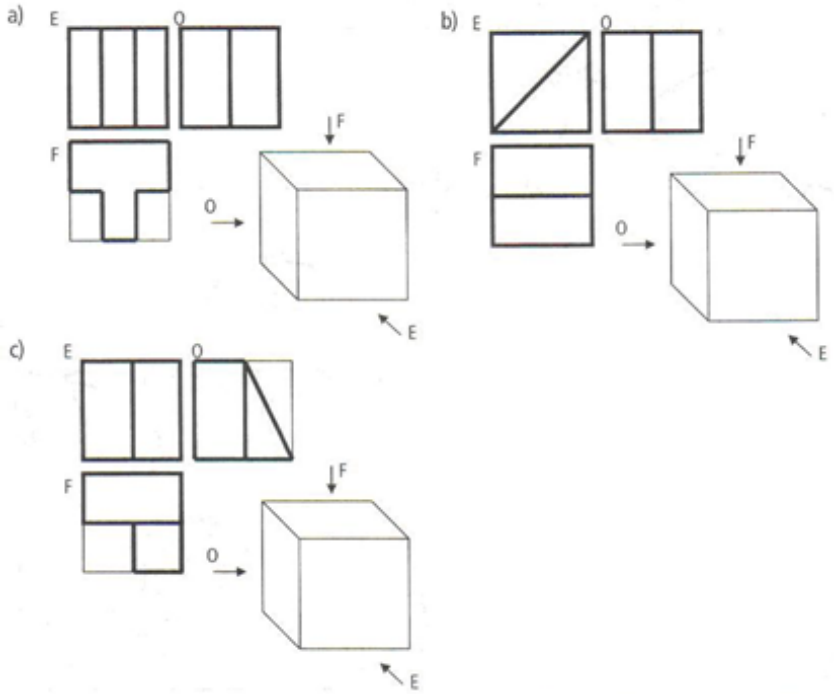


Figure 6

3. Results

The students of the Szent István University were 5% (Task 1) and 7% (Task 2) better on the tasks of manipulating the imaginary solid. At the Szent István University there were 2% more students, who knew the correct solution in the reconstructural task (Task 6).

In the exercises of the representation of projections the students of the University of Debrecen scored 1% (Task 4) and 5% (Task 5) better than the other one.

We examined furthermore the differences between the genders in solving these exercises. The list of 7 tests contained altogether 18 exercises. We looked at what percentage of the students gave correct answers for each exercise and then we examined them by gender: what percentage of the female students and what percentage of the male students succeeded. Of the 111 students at the Szent István University 57 were males (51%) and 54 females (49%). At the University of Debrecen of the 87 students 45 were males (52%) and 48 females (48%). At the Szent István University the male students performed better in 16 of the 18 exercises and in only 2 exercises did the female students give more right answers (1/1 and 6/b). At the University of Debrecen the male students did better in 15 exercises of the 18, the

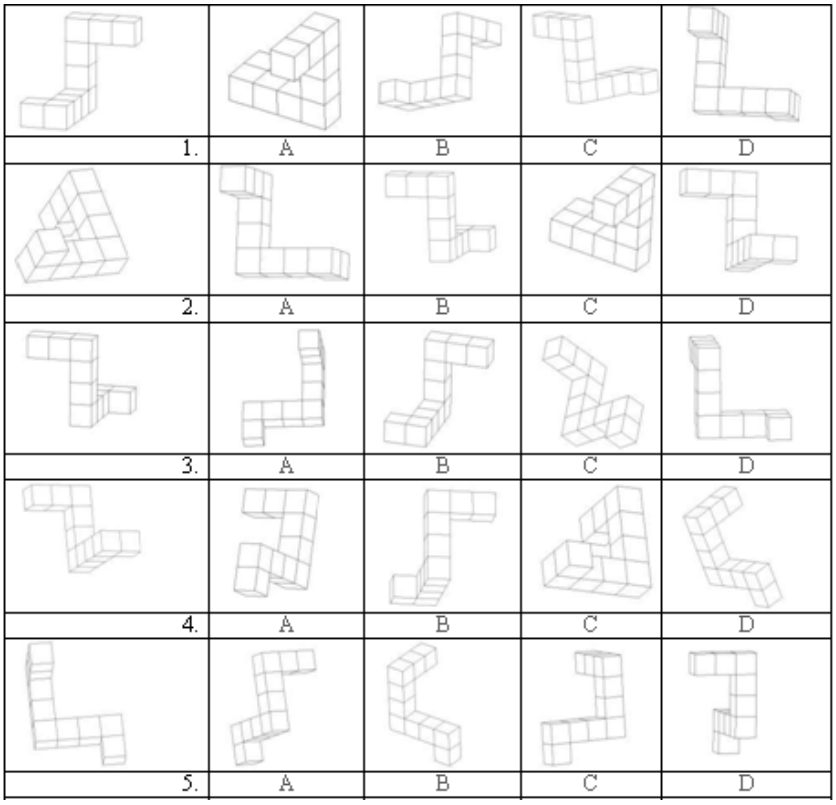


Figure 7

female students in 3.

The students of the Szent István University did better than the students of the University of Debrecen in all the exercises that concentrate on mental rotation. They were better by 6% at 1/1, 3% at 1/2, 9% at 2/1, 6% at 2/2. At the Szent István University the female students did better by 1% at exercise 1/1, while at the University of Debrecen they did 6% worse than the male students. At both universities the males performed better in all the rest of the exercises of both tests.

Task 2, in which both genders performed less well, is composed of 2 exercises. Both groups managed to solve the Task 2 the worst of all, and this exercise was the biggest difference between the two groups. In the diagram we can tell that the second part proved to be the more difficult one. Both the male students and the female students made the most mistakes in this part. This exercise can be solved with mental rotation and requires excellent spatial abilities. The MRT shows that mental rotation was not the difficulty because this is where they actually did the best. This exercise can be linked to number 4 where we could rephrase the question and ask whether there are any perspectives of the given objects that can be fit

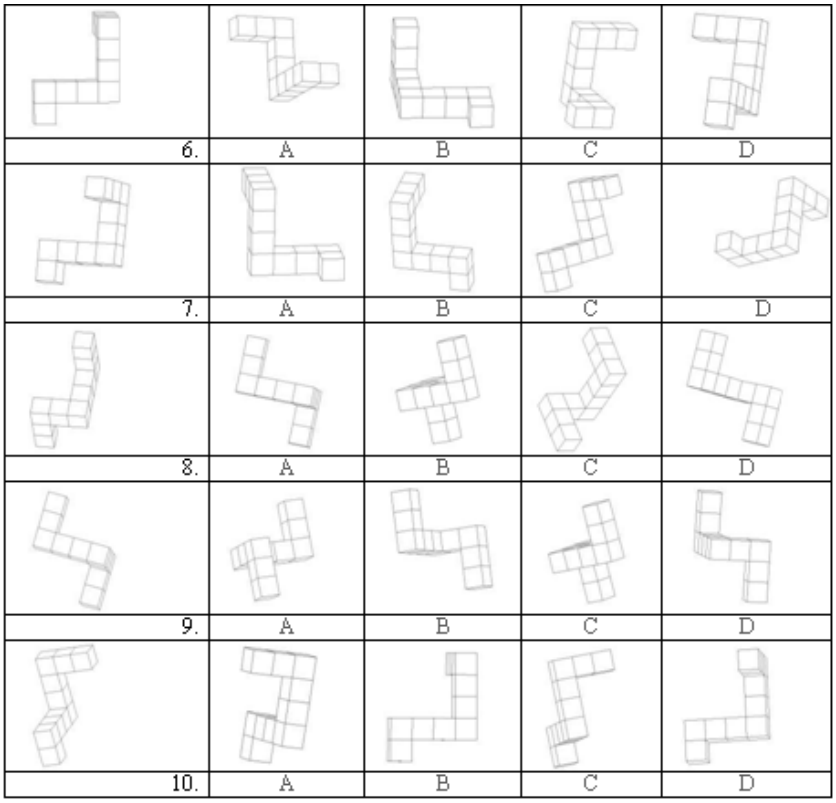


Figure 8

through the holes. So for successfully solving this task they need to be able to deal with perspectives and rotation. The hidden edges being shown further complicated the task.

The students of the University of Debrecen performed better at most of the tasks concentrating on representation and reading projection. 9% better at 3/b, 1% at 3/c, 1% at 4, 6% at 5/a, 4% at 5/b, 12% at 5/c. At the University of Debrecen the female students were better by 4% at 3/a, by 8% at 3/b and by 2% at 5/b. In all the part of tasks 3, 4 and 5 the male students did better at both universities, while the most difficult for all of them was 5/d.

One of the typical errors the students made was not to do with their spatial abilities but rather their consistency. In exercise 3, where they had to draw perspectives of objects, the problem was that they couldn't prepare the drawing. The ability to make these drawings is expected of students of architecture that had taken two semesters of Descriptive Geometry, drawing and other basic studies. The most common mistake was inconsistency in their drawings. In some cases they included in their drawing the frame that had been prepared for them ahead, while in other

cases they didn't. The students of the Szent István University 21%, the University of Debrecen 22% made the mistake of not being consistent with using the frame. The students of the Szent István University 38%, while the University of Debrecen 37% gave a perfect solution for the task.

We can see a great difference between the male students and female students of the Szent István University at exercise 4, where 70% of the female students and 80% of the male students gave the correct solution, while at the University of Debrecen 75% of the female students and 77% of the male students solved the exercise well. This is a perspective task where the object has to be left unrotated and the student has to decide which perspective the given object matches. Although this exercise was not where they performed worst, after completing the test they all agreed that this had been the most difficult for them to solve.

At task 5 some students couldn't picture mentally the perpendicular projection of some parts of the twisted cable. They either didn't draw any of the projections or left out parts of the object. In the three parts of this task (5/a, 5/b, 5/c) the students of the University of Debrecen performed better.

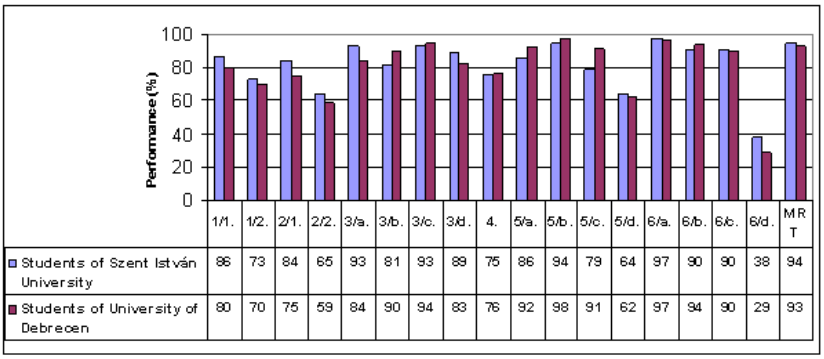
Of the reconstruction tasks in 6/a both groups performed at 97%, this is the part of task both groups did better at. The students of the University of Debrecen did better by 4% at 6/b, which is one where the female students at the Szent István University did better by 5% than the male students. At all the part of task 6 the male students performed better at both universities. At 6/a and 6/c the students performed with the same results at both universities. In 6/d Szent István University did better by 9%. 6/a was the one where both universities did best and 6/d where they did worst. 6/d has proved to be by far the most difficult one of task 6 for all the students. Here most of the students made were the reconstructions of object either incomplete or wrong. This is where we can observe the largest difference between the male students and female students: at the Szent István University the males performed better by 27%, while at the University of Debrecen by 11%.

Based on these findings we can conclude that there is no significant difference between the performances of the students of these two universities.

The students of the Szent István University were better in the tasks of manipulating the imaginary solid and in reconstructural task. Based on the comparison of the curricula, tests, technical drawings and the results of our test we can conclude that the students of the Szent István University were better at imaginary manipulation of the object, since they have more technical drawings and models creating.

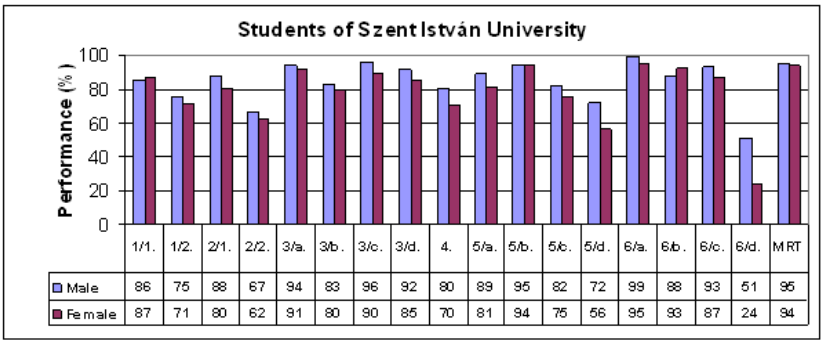
But in the exercises of the representation of projections the students of the University of Debrecen scored better than the other one, maybe because and they spent more time with descriptive geometry in two lectures per week, so they can see the connections better and have more practice in description and reading of projection tasks solving.

Figure 9, Figure 10 and Figure 11 show the performance of the students on the test.



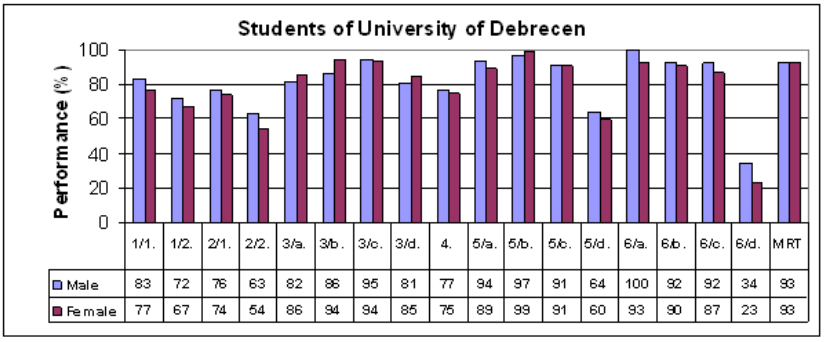
	1/1.	1/2.	2/1.	2/2.	3/a.	3/b.	3/c.	3/d.	4.	5/a.	5/b.	5/c.	5/d.	6/a.	6/b.	6/c.	6/d.	MRT
Students of Szent István University	86	73	84	65	93	81	93	89	75	86	94	79	64	97	90	90	38	94
Students of University of Debrecen	80	70	75	59	84	90	94	83	76	92	98	91	62	97	94	90	29	93

Figure 9: Students' performance



	1/1.	1/2.	2/1.	2/2.	3/a.	3/b.	3/c.	3/d.	4.	5/a.	5/b.	5/c.	5/d.	6/a.	6/b.	6/c.	6/d.	MRT
Male	86	75	88	67	94	83	96	92	80	89	95	82	72	99	88	93	51	95
Female	87	71	80	62	91	80	90	85	70	81	94	75	56	95	93	87	24	94

Figure 10: Students' performance



	1/1.	1/2.	2/1.	2/2.	3/a.	3/b.	3/c.	3/d.	4.	5/a.	5/b.	5/c.	5/d.	6/a.	6/b.	6/c.	6/d.	MRT
Male	83	72	76	63	82	86	95	81	77	94	97	91	64	100	92	92	34	93
Female	77	67	74	54	86	94	94	85	75	89	99	91	60	93	90	87	23	93

Figure 11: Students' performance

4. Conclusion and further research

Spatial abilities of first-year engineering students have been studied in this paper. Their abilities were tested by several tests (MRT, imaginary manipulation, repre-

sentation and reading of the projection, reconstructural task). The results of our survey prove that the reconstruction and representation of the projection cause a problem for many students, as well as to imagine a spatial figure.

Stachel and his colleagues' [15] results were as follows in an experience: There were statistically significant differences between males and females in almost all groups except one. In accordance with the international experiences [2] [4] we observed improvement after two semesters of descriptive geometry courses. [8] [11] Several research papers reported gender differences in the results of the spatial abilities test, and we have observed in our research gender differences too. Female students may not have the same spatial ability skills as male students, which can partly explain gender differences in spatial ability test; also female students choose typical mistake in some tasks more frequently, than male students.

The results of the survey prove that it causes a problem for many students to imagine and manipulate a spatial figure. It would be very useful for the university courses some review-systematization classes should be devoted for the summary of spatial ability, spatial geometry and solving spatial geometrical tasks. The effectiveness of teaching spatial geometry can be influenced to a great extent by using several different models, more technical drawings, manipulation activities with spatial models, especially dynamic, in the demonstration of relations between spatial models and operations with them. Vásárhelyi [17] calls the attention to the use of computers besides the traditional models. The three-dimensional models can be a great help in the teaching and learning of geometry. It is much easier to imagine and represent different views of a solid when we can see the formal characteristics. The proper use and frequent study of spatial visual aids can result in such an inner spatial vision that makes the individual imagination of the spatial relations possible.

Lord [5] applied a 30 minutes practice on a 14 weeks course with first/second year students where they had tasks in which they had to cut three-dimensional solids in their mind and then they had to draw the surface of the two-dimensional planes they got. In the post-test the spatial awareness and efficiency became better. Field [1] describes work conducted at Monash University aimed at measuring spatial skills, improving the sensitivity of visualization tests, and developing the skill for some engineering undergraduates. The testing of undergraduate students at Monash University has indicated the following factors:

- First level engineering students are to possess specially higher spatial skills than the general population.
- Spatial skills are not measurably developed by a conventional mechanical engineering undergraduate course.
- A special course with about 50 contact hours appears to have been successful in developing visualization skill in first level engineers. (There is some evidence that freehand drawing of three dimensional objects, in orthogonal, isometric and perspective views makes a major contribution to the development of spatial skill.)

We all agree in that the development of the spatial ability is a very important task because we have to understand and develop the geometry knowledge of the students in the unity of the theoretical knowledge and the spatial abilities. Every skill, like the spatial ability as well can be developed at the right age with the suitable teaching strategy. The results of the survey prove that it causes a problem for many students to imagine a spatial figure and this way it affects the solving of spatial geometrical tasks as well. Therefore it would be very useful to start the teaching of spatial geometry with spending more time with the models of spatial solids, and should be devoted for the summary of spatial ability, spatial geometry and solving spatial geometrical tasks. The effectiveness of teaching spatial geometry can be influenced to a great extent by using several different models, manipulation activities with spatial models, especially dynamic, in the demonstration of relations between spatial models and operations with them. Future work will be focused on task based student interviews to reveal the problems and their results.

References

[1] B. W. Field, A Course in Spatial Visualisation, *Journal for Geometry and Graphics*, 3/2 (1999) 201–209.

[2] G. Gittler, J. Glück, Differential Transfer of Learning: Effects of Instruction in Descriptive Geometry on Spatial Test Performance, *Journal for Geometry and Graphics*, 2/1 (1998) 71–84.

[3] F. H. Haanstra, Effects of art education on visual-spatial and aesthetic perception: two meta-analysis, *Rijksuniversiteit Groningen, Groningen*, (1994).

[4] C. Leopold, R. A. Górska, S. A. Sorby, International Experiences in Developing the Spatial Visualization Abilities of Engineering Students, *Journal for Geometry and Graphics*, 5/1 (2001) 81–91.

[5] T. R. Lord, Enhancing the visuo-spatial aptitude of students, *Journal of Research in Science Teaching*, 22/5 (1985) 395–405.

[6] P. H. Maier, 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.

[7] M. G. McGee, Human Spatial Abilities: Psychometric studies and environmental, genetic, hormonal and neurological influences, *Psychological Bulletin*, 86 (1979) 899–918.

[8] R. Nagy-Kondor, Spatial ability of engineering students, *Annales Mathematicae et Informaticae*, 34 (2007) 113–122.

[9] B. Németh, Measurement of the development of spatial ability by Mental Cutting Test, *Annales Mathematicae et Informaticae*, 34 (2007) 123–128.

[10] B. Németh, M. Hoffmann, Gender differences in spatial visualization among engineering students, *Annales Mathematicae et Informaticae*, 33 (2006) 169–174.

[11] B. Németh, C. Sörös, M. Hoffmann, Typical mistakes in Mental Cutting Test and their consequences in gender differences, *Teaching Mathematics and Computer Science*,(2007) 1–8.

[12] J. Piaget, B. Inhelder, A gyermek logikájától az ifjú logikájáig, *IEEE Comp. Graph. and Appl.*, 13 (1993) 43–49.

[13] L. Séra, A. Kárpáti, J. Gulyás, A térszemlélet, *Comenius Kiadó, Pécs*, (2002).

[14] K. Shiina, D. R. Short, C. L. Miller, K. Suzuki, Development of Software to Record Solving Process of a Mental Rotations Test, *Journal for Geometry and Graphics*, 5/2 (2001) 193–202.

[15] E. Tsutsumi, H.-P. Schroecker, H. Stachel, G. Weiss, Evaluation of Students' Spatial Abilities in Austria and Germany, *Journal for Geometry and Graphics*, 9/1 (2005) 107–117.

[16] S. G. Vanderberg, A. R. Kuse, Mental Rotations, a group test of three dimensional spatial visualization, *Perceptual and Motor Skills*, 47 (1978) 599–604.

[17] É. Vásárhelyi, A vizuális reprezentáció fontossága a matematikaoktatásban, `http://ikon.inf.elte.hu/~kid/ELEMIMAT/BLOKK2003/vizualis/VIZUALIS.HTML`

Annales Mathematicae et Informaticae
***40** (2012) pp. 203–216*
http://ami.ektf.hu

Spatial ability of students of mathematics education in Croatia evaluated by the Mental Cutting Test

Željka Milin Šipuš[a], Aleksandra Čižmešija[b]

[a]Department of Mathematics
Faculty of Science, University of Zagreb, Croatia
zeljka.milin-sipus@math.hr

[b]Department of Mathematics
Faculty of Science, University of Zagreb, Croatia
aleksandra.cizmesija@math.hr

Submitted December 22, 2011 — Accepted April 11, 2012

Abstract

Spatial ability of students of mathematics education at the Department of Mathematics, University of Zagreb, Croatia, has been evaluated by the Mental Cutting Test (MCT) and the analysis of the results is presented in this paper. Furthermore, the obtained results are compared with the results of engineering students at University of Zagreb. Gender analysis of the results is also presented.

Keywords: Mental Cutting Test, spatial ability, geometry education

MSC: 51N05, 97G80, 97D70

1. Introduction

The aim of this paper is to analyze the results of a survey of spatial ability of prospective mathematics teachers at the Department of Mathematics, University of Zagreb, Croatia. With one hundred enrolled students a year, it is the largest institution of higher education in Croatia offering 3-years Bachelor programme and 2-years Master programme in mathematics education. The students have been

evaluated by the classical Mental Cutting Test and it is important to emphasize that this is the first time such a survey is carried out in Croatia.

Development of spatial ability is declared as one of the key goals of mathematics education in primary and secondary school all around the world. It is also of central importance for further (higher) education of engineering professions. Therefore, special attention should be paid to improvement and better structuring of spatial ability of prospective mathematics teachers for the compulsory and secondary level education.

Spatial ability can be defined as an intuition about shapes and the relationships among shapes, that is, as the ability to generate, retain, retrieve, and transform well-structured visual (mental) images ([10]). Individuals with well-developed spatial ability have a feeling for the geometric aspects of their surroundings and shapes formed by objects in the environment. The most widely recognized model for the development of geometric thinking is described by the so-called van Hiele theory ([21]) and especially can be applied to development of spatial skills. The theory identifies five hierarchical developmental levels of individuals' understanding of spatial ideas. These are visualization, analysis, informal deduction, deduction and rigor. It is expected for students of mathematics education to reach the level of rigor by the end of their study while the compulsory education should provide plenty of opportunities for each student to reach the level of informal deduction.

In Croatia, the new National Framework Curriculum for Early Education, General Compulsory and Secondary Education ([5]) has been adopted in August 2010, but has not been fully implemented so far. Anyhow, it is crucial to stress out that this document pays utmost attention to the continuous development of spatial ability, which can be particularly seen in the curriculum areas of Mathematics (strand Shape and space) and Arts (strand Visual arts and design) throughout all four curriculum stages. The new curriculum remedies the serious weakness of the present one, in which development of spatial ability has been to certain extent neglected. For example, there was an obvious discontinuity in learning spatial geometry since 3D shapes have been studied only in the first, fourth and eight grade, mainly by considering their metrical properties (lengths, angles, areas and volumes). Our intention is therefore to analyze spatial skills of students before implementing the new curriculum. We also summarize the expected learning outcomes in spatial geometry now taught in primary and secondary school as well as in pre-service mathematics teacher education. Considering the fact that majority of mathematics education students are female and that males typically outperform females at various spatial ability tests, a gender analysis of the results is presented as well.

2. Theoretical background

Different spatial skills (e.g. mental rotations and mental cutting) can be measured by various tests and for various purposes. Special attention is usually paid to investigation of spatial skills of engineering students (e.g. [7, 13, 15, 16, 17, 18, 19]). The Mental Cutting Test (MCT) is one of the tests widely used for evaluation

of spatial skills although it was originally developed as an entrance exam to US higher institutions (a sub-set of CEEB Special Aptitude Test in Spatial Relations, 1939). It is the standardized test containing 25 multiple choice items with 5 given alternative answers among which one is correct. Solving time of the test is 20 minutes. Each item is a perspective drawing of a solid body cut by a plane. There are two different types of problems – pattern recognition problems and quantity problems. Pattern recognition problems are problems in which a student should recognize the pattern of the cross section among strongly different alternatives. In quantity problems a student should determine the right relative quantities (ratios of lengths, angles). An example of pattern recognition problems is the item 8 (Fig. 1) and examples of quantity problems are the items 13 and 25 (Fig. 2), see [20].

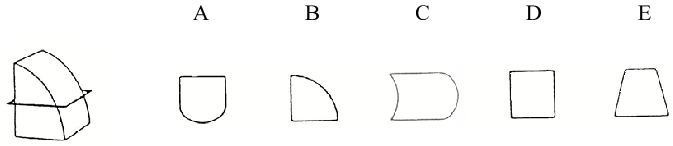


Figure 1: A pattern recognition problem - item 8

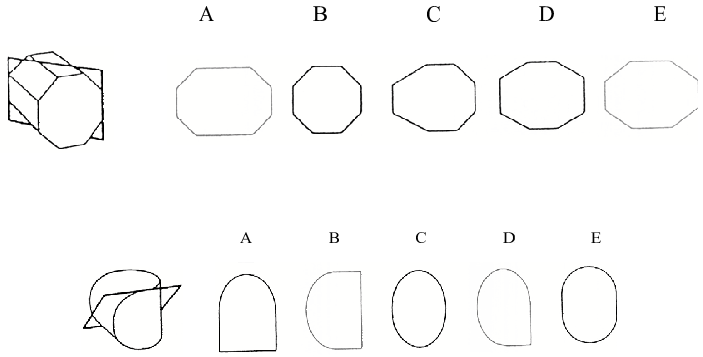


Figure 2: Two quantity problems - items 13 and 25

As it was stated in [20], in order to solve the MCT problems, a student passes through three phases: recognizing the solid from the perspective drawing, cutting the solid by the assumed plane and judging the characteristic quantity of the section, if necessary. Failures are mostly based on the fact that students do not recognize spatial forms of the objects. In recent times, besides the classical MCT test, some other tests for evaluating spatial abilities of students were developed in order to meet specific educators needs (e.g. [6, 8]).

Here we also stress out some results from the literature which report on gender differences in spatial skills. Decades of intense research on that subject have confirmed that men typically outperform women in spatial ability tests (see e.g. exhaustive psychological meta-analyses presented in [9, 22] and the recent overview with focus on engineering in [13]). Those differences are most evident in mental rotation tasks, even from early ages ([14]). However, it is found that previous spatial activity participation significantly influences performance on spatial ability tests for both sexes ([2]) and that spatial ability can be improved through instruction if appropriate materials and activities are provided ([1, 3, 4, 7, 11, 12, 13, 18]). This includes thematic instruction on spatial geometry topics (such as formal academic workshop and courses described and assessed in [7] and [18]), technical education, as well as non-academic activities (e.g. playing 3D video games, having musical experiences, creating artwork and playing with construction toys) and participation in other activities relying on hand-eye coordination (e.g. playing sports), as described in [13] and in references cited therein. The studies have confirmed the gain for male and female participants, despite initial differences in their achievement. Improvements have been noticed both for pupils ([3]) and older populations ([1, 4, 7, 11, 12, 13, 18]), with persisted retention of effects ([3]).

3. Method

The main focus of the present survey is to analyze spatial skills of prospective teachers of mathematics at the Department of Mathematics, University of Zagreb, Croatia, by means of the Mental Cutting Test. The testing was carried out at the beginning and at the end of the fall semester 2009/2010 for the students of the first year of the Bachelor programme in Mathematics education, ME1 (98 students), and the students of the second year of the Master programme in Mathematics education, ME5 (32 students). The latter students are studying the fifth of five years of the mathematics education programme at the Department of Mathematics, University of Zagreb. We will present the results of the MCT as an one-time test, that is, as the screening instrument of spatial skills of students at the beginning and at near the end of the course of their study. The MCT is also very often used to evaluate the effect of study courses on spatial geometry (e.g. Descriptive geometry) – then it is performed twice, as a pre-test and a post-test, which was not our intention. In order to avoid that the item position in the test influences the scoring result (when solving last items students can be less motivated), the MCT test was distributed to students in four different permutations of items.

Besides this main focus of the survey, our intention was also to compare spatial skills of the two above mentioned groups of students with skills of students of engineering faculties at the University of Zagreb, Croatia. Therefore, the MCT was administered to the first-year engineering students of the Faculty of Civil Engineering, Faculty of Architecture, Faculty of Geodesy and Faculty of Mining, Geology and Petroleum Engineering. These faculties train students in solving spatial problems and therefore require well-developed spatial skills, which is especially notable

for one of the faculties which has a test of spatial abilities as a part of the entrance exam.

For better understanding of the context of the survey and its results, we present a brief review of the expected learning outcomes in spatial geometry in Croatian primary and secondary education. According to the present curriculum, upon completion of their primary education (first eight years of schooling), students are expected to be able to recognize basic solids and their nets, reason about simple spatial relationships of points, lines and planes in space, apply the Pythagorean theorem to determine lengths in solids, and calculate surface areas and volumes of right prisms (such as boxes and prisms with an equilateral triangle and a regular hexagon as their bases), spheres, right cones and cylinders.

With respect to primary education, spatial geometry of the common core of majority of secondary mathematics programmes extends only to cover basic properties of skew prisms, pyramids, cones and cylinders.

However, students of the first year entering the higher education in engineering, mathematics and science programmes usually take the stronger track in secondary mathematics education. But still, they do not get any special education in spatial geometry – neither in e.g. analytic space geometry, nor in descriptive geometry during their secondary education. As stated in the Introduction, the major accent in the present curriculum is put on applying trigonometry to solids in order to measure lengths, angles, areas and volumes. Regarding the higher education, ME5 students have met contents in spatial geometry during their university study, namely in courses on analytic geometry, calculus of multivariable real functions (obligatory) and classical differential geometry (elective). The Descriptive geometry is an one-semester obligatory course given at the fifth year (the second year of the Master programme in Mathematics education). The majority of ME5 students have enrolled in the course of Descriptive geometry during the semester when the MCT test was administered while spatial skills of the first-year students regard only their pre-university education. This applies to all first-year students tested, both to ME1 and to all engineering (Eng) groups.

4. Results

In Table 1 the classical descriptive statistics of the results of students involved in the analysis is given.

Furthermore, in Fig. 3 and Fig. 4 overall results of ME1 and ME5 students are presented, where the achieved points of students are grouped into groups from 1-5 points, 6-10 points, 11-15 points, 16-20 points and 21-25 points. In Fig. 5, in the same diagram, results of ME1, ME5 and students of the best performing engineering faculty Eng1 are presented. Already from Table 1, it can be seen that ME5 students, as well as students of engineering faculties (except Eng4), achieve higher results in MCT then ME1 students. These results are confirmed by applying statistical t-test which shows that there are statistically significant differences between results of ME1 and ME5 students, at the level of significance of

	ME1	ME5	Eng1	Eng2	Eng3	Eng4
N	98	32	116	102	204	192
Mean	9.97	14.66	15.83	13.06	11.87	8.71
SD	4.21	4.96	3.86	4.59	5.33	4.06
Median	10	14.5	16	12	11	11.8
Mode	10	18	15	10	6	7
Max	25	23	24	23	25	23
Min	2	2	6	3	2	2

Table 1: Results of the first MCT test for ME1 and Eng groups and of the second MCT test for ME5 group

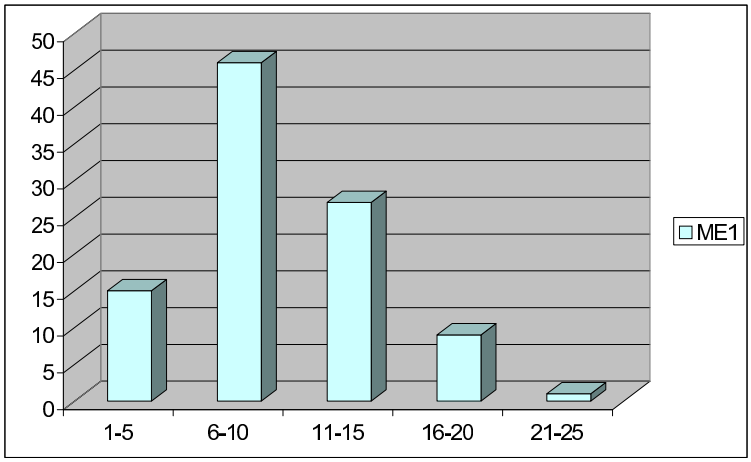


Figure 3: Percentages of ME1 students who correctly solved certain number of items

$p = 0.01$. Similarly, there are statistically significant differences between results of ME1 and Eng1, Eng2 and Eng3 students, at the level of significance of $p = 0.01$. A possible reason for an evident difference between results of ME1 and ME5 students could be that ME5 is a smaller, more homogeneous group of students motivated in obtaining Master degree, having achieved and improved their spatial skills from various mathematics courses.

Similar comparisons of different groups of engineering students in Austria and Germany can be found in [19].

4.1. Gender analysis of the results

With respect to gender, statistical results are presented in Table 2, Table 3, whereas diagrams in Fig. 6, Fig. 7 present overall scores. Already from these data, it can

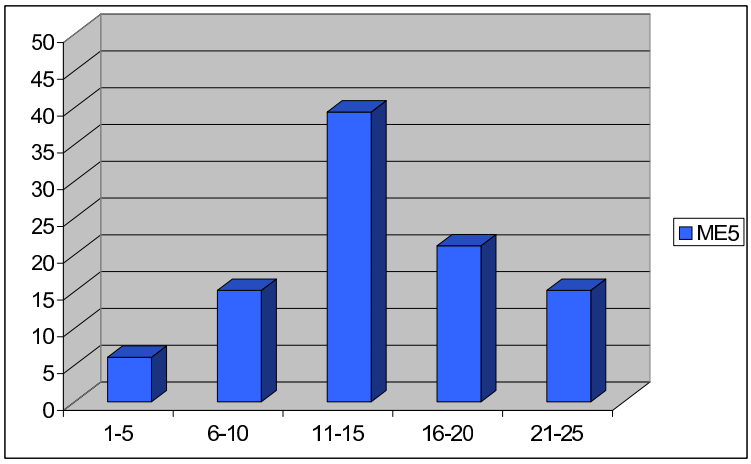


Figure 4: Percentages of ME5 students who correctly solved certain number of items

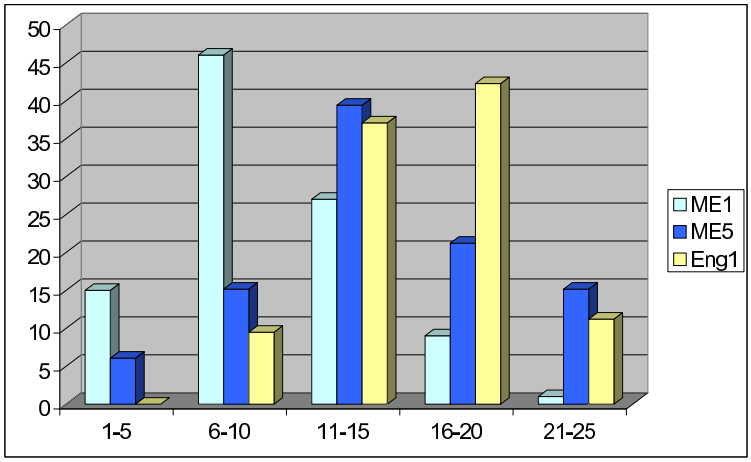


Figure 5: Comparison between ME1 students, ME5 students and Eng1 students in number of correctly solved items

	male	female
N	29	69
Mean	12.03	9.10
SD	4.85	3.60

Table 2: Statistics of results with respect to gender of ME1 students

	male	female
N	6	26
Mean	19.00	13.65
SD	3.46	4.75

Table 3: Statistics of results with respect to gender of ME5 students

be seen that male students are performing better in the MCT than their female colleagues which was confirmed by applying statistical t-test. It is shown for both ME1 and ME5 group of mathematics education students that, at the level of significance of 0.01, male students perform statistically significantly better then female students. This result agrees with the result obtained in [6, 15, 19]. It is of great importance to be aware of this result since mathematics education female students at the Department of Mathematics in Zagreb outnumber their colleague male students (in the observed year 2009/2010, percentage of female students was 70% at the first year and 81% at the fifth year).

In Table 4, gender differences of students of the best performing engineering faculty Eng1 are presented. Comparing ME1 and ME5 female students with female students from the group Eng1 we can conclude again that female students of the best performing engineering faculty Eng1 are performing statistically significantly better at the level of significance of 0.01.

	male	female
N	50	66
Mean	17.08	14.8
SD	3.37	3.95

Table 4: Statistics of results with respect to gender of Eng1 students

Some further evidence on gender differences is presented in the next section.

5. Items analysis

In this section, the analysis of the items 8, 13 and 25 for ME1, ME5 and Eng1 students is given. Some of these items are included in the analysis in [17, 19, 20]. The analysis involves the analysis of distractors (distribution of answers) with spacial attention to gender differences.

Distribution of answers for the item 8 is given in Fig. 8. Already from the diagram, it can be seen that the students from the group Eng1 are very confident with the correct answer, whereas the first-year mathematics education students make the typical mistake by marking A as the correct answer more often than D.

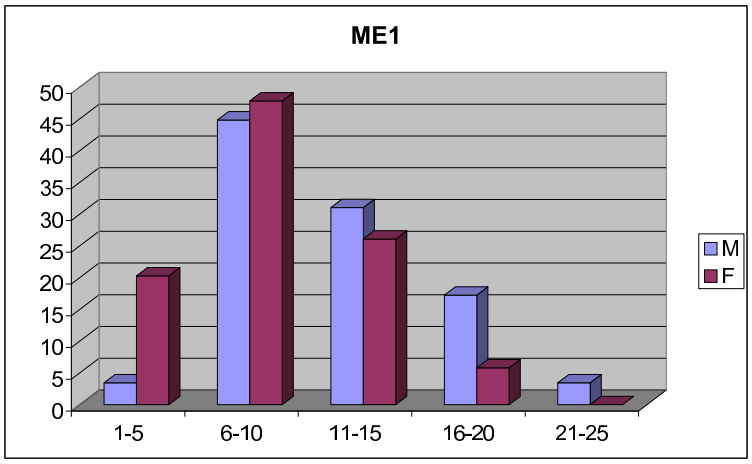


Figure 6: Gender comparison – number of correctly solved items by female and male students of the group ME1

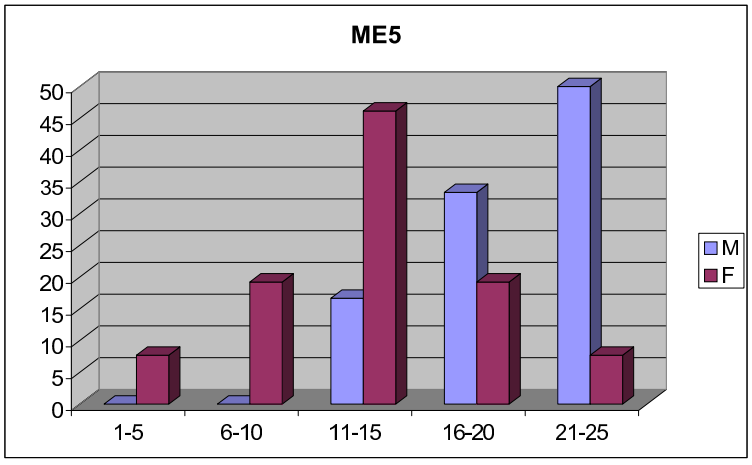


Figure 7: Gender comparison – number of correctly solved items by female and male students of the group ME5

As it is stated in [17], the natural explanation for this choice is that a curved object usually yields curved cuts.

Distribution of answers for the item 13 is given in Fig. 9. Being a quantity problem, item 13 turns out to be difficult. But again, students from the group Eng1 are confident with the correct answer, although the typical mistake A appears as well (as in [20]). The ME1 students and the ME5 students in much higher percentage mark incorrect answers A as correct, whereas ME1 students mark as correct also the answer C.

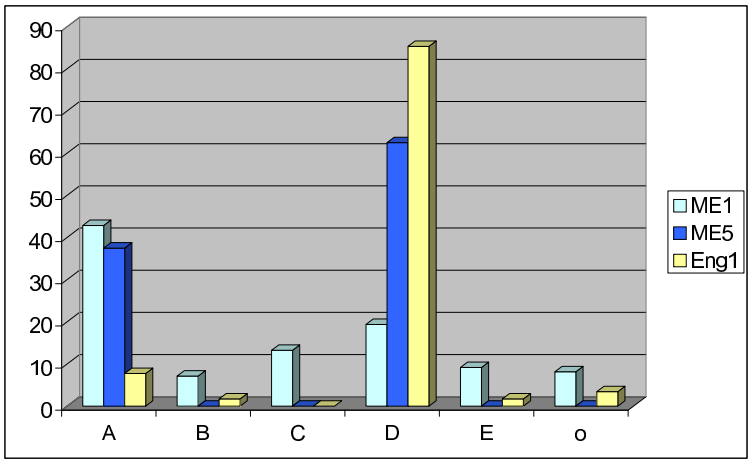


Figure 8: Item 8 – distribution of answers in percentages of ME1, ME5 and Eng1 students, correct answer is D

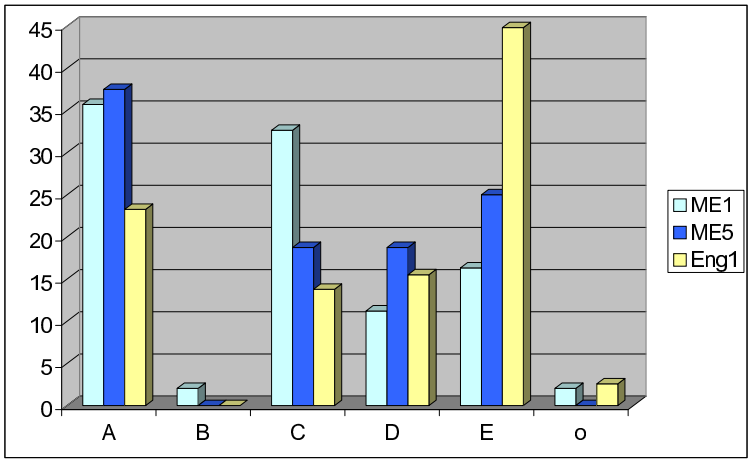


Figure 9: Item 13 – distribution of answers in percentages of ME1, ME5 and Eng1 students, correct answer is E

Distribution of answers for the item 25 is given in Fig. 10. It can be noticed that more students in all groups, compared to other problems, did not try to answer the problem. Item 25 can be considered as a pattern recognition and quantity problem at the same time since many students failed in recognizing the pattern of the cross section. There is no strongly dominating answer and even three incorrect answers (B, D, E) are chosen more often than the correct answer C. In answers B and D one should decide on the pattern of the section, where the answer B offers the front view of the object and the answer D the figure which is not symmetrical.

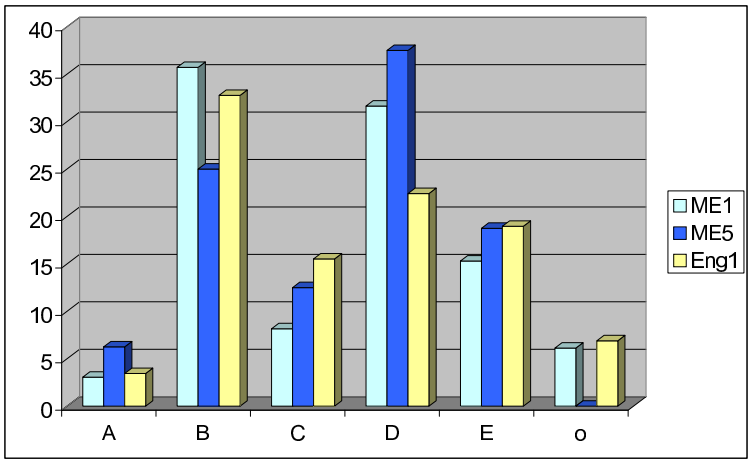


Figure 10: Item 25 – distribution of answers in percentages of ME1, ME5 and Eng1 students, correct answer is C

Furthermore, even the students who decided correctly on the pattern, chose more often incorrect answer E than C. Low results of the groups ME1 and Eng1 can be explained as in [19], that a student needs not only the intuitive spatial ability, but some geometrical consideration as well, but this explanation cannot be applied to the group ME5. The ME5 students have acquired formal knowledge sufficient for solving this item, so possible explanation of their failure on this problem lies in their lacking abilities to analyze and make connections to required knowledge, which was even more stressed in testing situation with little time available (less than one minute per item on average).

Analysis with respect to gender can be found in Fig. 11 and Fig. 12. In these figures, results on choosing the typical mistakes by male and female students in percentages of students who failed the problem are given. It can be seen that female students more often chose the typical mistakes. Similar results of the first-year engineering students at Faculty of Engineering, Szent István University in Hungary can be found in [17].

6. Conclusion

The Mental Cutting Test was performed as a screening test of spatial abilities of students of mathematics education at the Department of Mathematics, University of Zagreb. The survey was motivated by the changes planned in mathematics curriculum for the general compulsory and secondary education. Throughout the present mathematics curriculum, educational goals in spatial geometry are neglected, which has been now changed in the new curriculum. Therefore, special

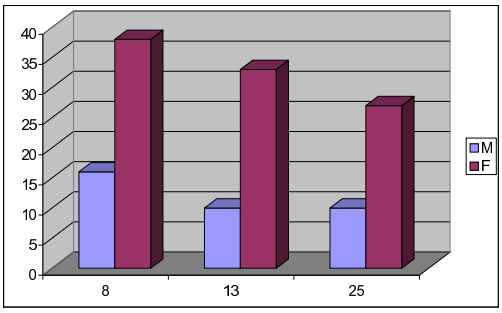


Figure 11: Typical mistakes by male and female students of ME1 in percentages of students who failed problems 8, 13 and 25

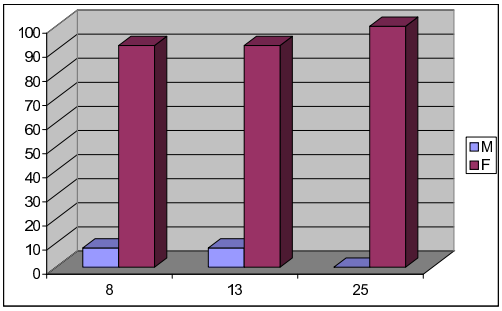


Figure 12: Typical mistakes by male and female students of ME5 in percentages of students who failed the problem 8, 13 and 25

attention should be paid to pre-service education of prospective teachers of mathematics, since, hopefully, they will have more instructional time to develop spatial abilities of their students.

As a screening test of the present situation, evaluation of spatial abilities by MCT gave the following results:

- Significant difference in results between the first-year ME1 students and the results of peer-students at engineering faculties at University of Zagreb.
- Significant difference in results of female and male students of mathematics education ME1 and ME5.
- Significant difference in results between the first-year ME1 students and the fifth-year ME5 students.
- Significant difference in results of female students of mathematics education ME1, ME5 and female students of the best performing engineering faculty Eng1.

Since the survey was administered to the students at very beginning of their university study (excluding ME5 group), the results obtained primarily reflect their spatial abilities gained through previous education. They clearly show that changes proposed in the geometry strand of the new curriculum are more than welcome. In order to use their potential and to offer formal support to foster the acquisition of spatial skills, teaching of topics in spatial geometry should provide opportunities to students to connect the underlying concepts, reason about them and to develop spatial skills continuously during all educational stages. Topics in spatial geometry are suggested to be firmly connected to real life objects and situations and to other curriculum areas as well. During all educational stages, teaching strategies should be designed to involve many student hands-on group and individual activities with manipulatives (e.g. handheld models) and technology (e.g. 2D and 3D models in Dynamic Geometry Software) that encourage students to create, explore, transform and relate 2D and 3D geometric shapes. This particularly goes to female students who are of special concern regarding their shown average lower achievements.

Acknowledgements. The authors thank the professors and assistants of the course Descriptive geometry at the Faculty of Civil Engineering, Faculty of Architecture, Faculty of Geodesy and Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, for motivating their students and helping to carry out the testing at these faculties. Furthermore, they wish to express their gratitude to professor Miljenko Huzak, Department of Mathematics, Faculty of Science, University of Zagreb, in helping them to carry out the statistical analysis of the results. They also thank the anonymous referee for valuable suggestions that improved the final version of the paper.

References

[1] Alias, M., Black, T. R., Gray, D. E., *Effect of Instruction on Spatial Visualisation Ability in Civil Engineering Students*, International Education Journal, 3(1), (2002), http://iej.cjb.net

[2] Baenninger, M., Newcombe, N., *The Role of Experience in Spatial Test Performance: A Meta-analysis*, Sex Roles, 20 (1989), 5–6, 327–344.

[3] Ben-Chaim, D., Lappan, G., Houang, R. T., *The Effect of Instruction on Spatial Visualization Skills of Middle School Boys and Girls*, American Educational Research Journal, 25 (1988), 51–71.

[4] Burnet, S. A., Lane, D. M., *Effects of Academic Instruction on Spatial Visualization*, Intelligence 4 (1980), 233–242.

[5] *Nacionalni okvirni kurikulum za predškolski odgoj i obrazovanje te opće obvezno i srednjoškolsko obrazovanje* [National Framework Curriculum for Early Education, General Compulsory and Secondary Education], http://public.mzos.hr/Default.aspx?sec=2685.

[6] Gorska, R., Juščakova, Z., *A Pilot Study of a New Testing Method for Spatial Abilities Evaluation*, Jour. Geom. Graph., **7** (2003), 237–246.

[7] Hsi, S., Linn, M. C., Bell, J. E., *The Role of Spatial Reasoning in Engineering and the Design of Spatial Instruction*, Journal of Engineering Education, **86**(2) (1997), 151–158.

[8] Juščakova, Z., Gorska, R., *TPS Test Development and Application into Research on Spatial Abilities*, Jour. Geom. Graph., **11** (2007), 223–236.

[9] Linn, M. C., Petersen, A. C., *Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-analysis*, Child Development, 56(6) (1985), 1479–1498.

[10] Lohman, D. F., *Spatial Ability and G.*, in I. Dennis, P. Tapsfield (Eds.), *Human Abilities: Their Nature and Assessment*, Hillsdale, NJ: Erlbaum, 97–116, 1996.

[11] Lord, T. R., *Enhancing the Visuo-spatial Aptitude of Students*, Journal of Research in Science Teaching 22 (1985), 395–405.

[12] Merrill, C., Devine, K. L., Brown, J. W., and Brown, R. A., *Improving Geometric and Trigonometric Knowledge and Skill for High School Mathematics Teachers: A Professional Development Partnership*, The Journal of Technology Studies, 36(2), (2010), 20–30.

[13] Metz, S. S., Donohue, S., and Moore, C., *Spatial Skills: A Focus on Gender and Engineering*, In B. Bogue, E. Cady (Eds.) *Apply Research to Practice (ARP) Resources*, 2012, http://www.engr.psu.edu/AWE/ARPResources.aspx

[14] Moore, D. S., Johnson, S. P., *Mental Rotation in Human Infants: A Sex Difference*, Psychological Science, 19(11) (2008), 1063–1066.

[15] Németh, B., Hoffmann, M., *Gender Differences in Spatial Visualization among Engineering Students*, Ann. Math. Inform., Vol. **33** (2006), 169–174.

[16] Németh, B., *Measurement of the Development of Spatial Ability by Mental Cutting Test*, Ann. Math. Inform., Vol. **34** (2007), 123–128.

[17] Németh, B., Sörös, Cs., Hoffmann, M., *Typical Mistakes in Mental Cutting Test and Their Consequences in Gender Differences*, Teaching Mathematics and Computer Science, **5**(2) (2007), 385–392.

[18] Sorby, S., *Educational Research in Developing 3-D Spatial Skills for Engineering Students*, International Journal of Science Education, **31**(3) (2009), 459–480.

[19] Tsutsumi E., Schröcker H.-P., Stachel H., Weiss G., *Evaluation of Students' Spatial Abilities in Austria and Germany*, Jour. Geom. Graph., **9** (2005), 107–117.

[20] Tsutsumi E., Shiina K., Suzaki A., Yamanouchi K., Saito T., Suzuki K., *A Mental Cutting Test on Female Students Using a Stereographic System*, Jour. Geom. Graph., **3** (1999), 111–119.

[21] van Hiele, P. M., *De problematick van het inzicht gedmonstreed van het inzicht van schodkindren in meetkundeleerstof* [The problem of Insight in Connection with School Children's Insight into the Subject Matter of Geometry], Doctoral dissertation, University of Utrecht, 1957.

[22] Voyer, D., Voyer, S., Bryden, M. P., *Magnitude of Sex Differences in Spatial Abilities: A Meta-analysis and Consideration of Critical Variables*, Psychological Bulletin, 117 (1995), 250–270.