No. 20. (2016)
ISSN 1331-1611

KoG

SCIENTIFIC - PROFESSIONAL JOURNAL
OF CROATIAN SOCIETY FOR GEOMETRY AND GRAPHICS

ISSN 1331-1611
9 771331 161005

ISSN 1331–1611

KoG

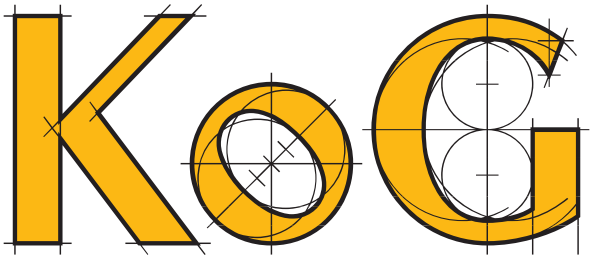
No. 20
Zagreb, 2016

SCIENTIFIC AND PROFESSIONAL JOURNAL OF
CROATIAN SOCIETY FOR GEOMETRY AND GRAPHICS

CONTENTS

ISSN 1331–1611

BROJ 20
Zagreb, 2016



ZNANSTVENO-STRUČNI ČASOPIS
HRVATSKOG DRUŠTVA ZA GEOMETRIJU I GRAFIKU

SADRŽAJ

IN MEMORIAM

ORIGINALNI ZNANSTVENI RADOVI

STRUČNI RADOVI

JELENA BEBAN-BRKIĆ, VLASTA SZIROVICZA

Vlasta Ščurić-Čudovan

(1931-2016)

U utorak 15. studenog 2016. napustila nas je sveučilišna profesorica u miru dr. sc. Vlasta Ščurić-Čudovan, dugogodišnja nastavnica Geodetskog fakulteta Sveučilišta u Zagrebu, znanstvenica, kolegica i prijateljica, te autorica *Uvodnika* za prvi broj časopisa KoG. Kao vrsna i cijenjena predavačica ostavila je prepoznatljiv trag generacijama studenata. Njena predavanja i vježbe odlikovali su se jasnoćom i sistematičnošću, a nije štedjela vremena i energije u želji da studenti u potpunosti razumiju materiju, da ju mogu objasniti, nacrtati i dati prostorno rješenje.

Vlasta Ščurić-Čudovan rođena je u Koprivnici 9. 5. 1931. gdje je pohađala osnovnu školu i gimnaziju. Godine 1958. diplomirala je na tadašnjem Matematičko-fizičkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu, smjer teorijska matematika.

Na poslijediplomskom studiju PMF-a u Zagrebu je 1966. stekla stupanj magistra matematičkih znanosti, obranivši rad iz područja projektivne geometrije pod naslovom *Pramenovi polarnih prostora i njima određeni kompleksi* koji je izradila pod mentorstvom akademika, profesora Vilka Ničea. Profesor Niče bio joj je mentor i na doktorskoj disertaciji pod naslovom *Orijentirani kompleks određen pramenom ploha 2. stupnja* koju je 1972. godine obranila te na Sveučilištu u Zagrebu bila promovirana za doktora matematičkih znanosti.

U radni odnos stupila je 1958. kao nastavnica matematike na Građevinskoj tehničkoj školi u Zagrebu, gdje je radila do kraja školske godine 1961. Za potrebe tog posla položila je stručni ispit za profesora srednje škole.

Rad nastavlja na geodetskom odjelu AGG fakulteta, kasnije Geodetskom fakultetu, kao asistentica za predmet Nacrtna geometrija. Po stjecanju stupnja magistra matematičkih znanosti, dvije akademske godine predaje Nacrtnu geometriju i na Akademiji likovnih umjetnosti u Zagrebu. Nakon stjecanja doktorata matematičkih znanosti i održanog habilitacijskog predavanja izabrana je 1973. u zvanje docentice za predmet Numerički račun (kasnije nazvanog Praktična matematika i konačno Geodetsko računanje), a uz to i dalje drži vježbe iz predmeta Nacrtna geometrija. 1978. unaprijeđena je u zvanje izvanredne, a 1984. u znanstveno-nastavno zvanje redovite profesorice za znanstveno područje matematika za predmete Nacrtna geometrija i Geodetsko računanje na studiju VII/1 i Nacrtna geometrija na studiju VI/1. Osim za redovne studente, održavala je i nastavu na Studiju uz rad, i to u Zagrebu, Splitu i Osijeku. Na svim tim studijima, predmetima, sveučilištima, Vlasta je nastojala postići angažman studenata već od početka studija, privikavati ih na kontinuirani rad kao i sažeto i pravilno izražavanje.

Na geometrijskim je predmetima bilo važno naučiti studente kako kroz prostorne odnose među zadanim elementima predložiti tijek rješenja zadatka i u konačnici traženo ispravno prikazati u ravnini. Sadržaj predmeta Geodetsko računanje bio je goniometrija, trigonometrija u ravnini i sferna trigonometrija, dakle klasični sadržaj jedne od najstarijih matematičkih disciplina primijenjenih u praksi. Kako sama Vlasta kaže u opisu predmeta...*"obrada tog sadržaja mijenjala se tokom vremena, od izrade logaritamskih tablica pa sve do za geodetsku praksu upotrebljivih džepnih kalkulatora odnosno kompjutora. Postavi li se, naime, problem i program nekritički, kompjutor može izbaciti na stotine, za geodetsku praksu posve neupotrebljivih podataka. Važno mi je da studenti, osim što će savladati nužne pojmove, odnose i relacije, već u početku studija shvate i nauče da formula nije nešto "u što se uvrsti i dobije rezultat" već da ona nosi mnoge poruke o zavisnosti njenih elemenata, te da numerički rezultat valja uvijek podvrgnuti analizi."*

Možemo samo pretpostaviti koliko je truda i vremena utrošila na sastavljanje zadataka iz navedenih predmeta, njihovu kontrolu i razgovor sa studentima, kako bi postigla svoje ciljeve kao nastavnik. Uz i dalje puno žara i angažmana i nimalo promijenjen odnos prema studentima održava nastavu i u akademskoj godini 1997./1998. nakon koje odlazi u mirovinu.

Znanstveno okruženje u kojem je V. Ščurić-Čudovan započela svoje djelovanje bilo je iznimno povoljno. Svakodnevni rad uz akademika Vilka Ničea, vrhunskog znanstvenika u području sintetičke projektivne geometrije, omogućio joj je čestu verifikaciju znanstvenih rezultata. Koliko je uz njega bila vezana možda najbolje govori tekst *Sjećanje na akademika prof. Vilka Ničea* koji je kao dio Spomenice preminulom članu, na godišnjicu smrti, izdala tadašnja JAZU. Ovdje prenosimo dio teksta.

"Prošlo je već 13 mjeseci od smrti našeg profesora i prijatelja, akademika Vilka Ničea. Zapao me je častan ali i vrlo odgovoran zadatak da pomognem evociranju sjećanja sviju nas na njegov život, znanstveni i nastavni rad, a posebno na njega kao čovjeka. Oprostite mi što ću u tome pokatkad biti subjektivna. Za to postoje mnogi razlozi, a osnovni je taj što je profesor Niče neposredno utjecao na tok čitavog mog života: od diplomskog rada pod njegovim vodstvom, poziva na rad na fakultetu, uvođenja u znanstveni rad, mentorstva magistarskog rada i doktorske disertacije do daljnjeg poticanja na znanstveni rad. To šturo nabrajanje krije u sebi mnogo, mnogo više. U prvom redu beskrajnu zahvalnost i poštovanje prema dr. Vilku Ničeu kao čovjeku i učitelju. Bilo bi prelijepo kad bi svatko imao sreće da ima svog voditelja u svim bitnim momentima života, posebno znanstvenog rada..."

Uz njega je imala veliku podršku profesorice Ljerke Dočkal Krsnik, najprije joj nadređene, a kasnije kolegice i prijateljice, s kojom je sudjelovala u radu brojnih kongresa i simpozija, domaćih i stranih.

Do devedesetih godina prošlog stoljeća njeno je znanstveno istraživanje rezultiralo nizom opširnih i iscrpnih radova iz područja sintetičke pravčaste geometrije realnog trodimenzionalnog prostora. Ako imamo na umu da su te tvorevine zvane kompleksi izučavane isključivo sintetičkom metodom, jasno je koliko je truda, dubokog poznavanja sintetičke geometrije i snažnog prostornog zora bilo potrebno da bi se u tom području došlo do novih rezultata.

Prenijet ćemo prikaz nekoliko radova iz izvještaja koji potpisuje profesor dr. Dominik Palman prilikom njenog izbora za redovitu profesoricu.

- *Der orientierte Niče-sche Strahlkomplex eines Flächenbüschels 2. Grades*, Rad JAZU 370 (1975), 57-91.

U ovom radu V. Ščurić-Čudovan istražuje singularne točke Ničeovog kompleksa i konstatira da takve točke leže na krivulji središta k^3 3. reda i na beskonačno dalekoj krivulji μ koja je također 3. reda, a poznata je i kao Jacobijeva krivulja. Nadalje definira involutorne zrake (VN) kompleksa na kojima su izlazna I i zalazna Z točka involutorno povezane i dokazuje da one čine kongruenciju 15. reda i 11. razreda. Osobitu pažnju posvećuje involutornim zrakama koje su pridružene točkama krivulja k^3 i μ. Ovisno o načinu pridruživanja točaka tih krivulja dobiva plohe P_1 6., P_2 12. i P_3 9. stupnja na kojima istražuje krivulje I i Z točaka.

- *Das (F_k^2) Flächenbüschel und eine Möglichkeit des Eintauchens des (MK) in den (VN) Komplex*, Rad JAZU 374 (1977), 57-91.

U radu Pramen ploha F_k^2 i jedna mogućnost uranjanja (MK) u (YN) kompleks *promatra se pramen ploha koji sadrži kuglu. Ta činjenica znatno utječe na osobine (VN) kompleksa. Dr. Vlasta Ščurić-Čudovan dokazuje da ovdje preuzima cijela beskonačno daleka ravnina ulogu krivulje μ 3. reda. U ovom se slučaju Ničeov kompleks raspada u dva kompleksa 3. i 5. stupnja, a Majcenov se kompleks podudara sa spomenutim kompleksom 3. stupnja. Kompleksni stožac se raspada u stošce 5. i 3. stupnja, a analogni je i raspad kompleksnih krivulja. Ispitane su i $I-$ kao i $Z-$krivulje na njima. U tom kompleksu su istražene i neke istaknute plohe i kongruencije.*

- *Die Kongruenzen der Involutorstrahlen eines durch das (F_k^2) Flächenbüschel bestimmten (VN) Komplexes*, Rad JAZU 382 (1978), 65-90.

Ovaj rad je nastavak istraživanja prethodnog rada. Utjecaj kugle na tvorevine involutornih zraka očituje se u tome da plohe P_1, P_2 i P_3 prelaze u odgovarajuće kongruencije. Involutorno pridružene $T-Z$ točke tvore u tim kongruencijama plohe 6., 9. i 3. reda. ...

- *Einige Probleme die durch die Einteilung eines Bündels der Flächen 2. Grades in ∞^1 Büschel solcher Flächen entstanden sind, I Teil*, Rad JAZU 403 (1983), 33-55.

U ovom radu se promatra svežanj ploha kao skup pramenova ploha (MF^2) sa zajedničkim stošcem. U uvodu se podsjeća na osnovna svojstva prostorne krivulje 6. reda koja sadrži vrhove svih stožaca svežnja. Istaknuvši na toj krivulji po volji odabranu točku M, pridruženu trisekantu m i konjugirani pravac m_k Vlasta Ščurić-Čudovan istražuje zrake kompleksa koje su pridružene točki M, točkama pravaca m i m_k te za svaku zraku kompleksa utvrđuje kojim je pramenom iz skupa (MF^2) određena i kojoj je točki pridružena. Na taj način promatrane su tvorevine zraka tetraedarskih i Majcenovih kompleksa pridruženih skupu pramenova (MF^2).

Iz opisa navedenih radova izlazi da dr. sc. Vlasta Ščurić-Čudovan vrlo temeljito poznaje i razrađuje opće i specijalne pramenove ploha 2. stupnja. Dala je vrijedne doprinose obradi tetraedarskog, Majcenovog i Ničeovog kompleksa koji su pridruženi pramenovima ploha 2. stupnja. Iz posljednjeg (gore navedenog, op. a.) rada vidljivo je da je područje istraživanja proširila na svežanj ploha 2. stupnja. Razloživši taj svežanj ploha na niz pramenova sa zajedničkim stošcem, došla je do vrijednih rezultata, te se može očekivati na tom području i dalji uspješan rad. U svojim radovima dr. sc. Vlasta Ščurić-Čudovan služi se sintetičkom metodom koja danas nije u centru pažnje, no ona je pokazala da se tom metodom još uvijek može doći do vrijednih rezultata. O svojim radovima referirala je na domaćim i austrijskim kongresima koji imaju karakter kongresa njemačkog govornog područja, gdje je izazvala živ interes istaknutih stručnjaka te problematike."

Devedesetih godina prošlog stoljeća smatrala je da se treba okrenuti istraživanjima u nekom njoj novom području sintetičke geometrije, jer je prethodno bilo zaokruženo. Stjecajem okolnosti našla je u vrhunskom austrijskom geometričaru, dr. Hansu Sachsu, profesoru sa Zavoda za primijenjenu matematiku i geometriju - Montanuniversität Leoben, Austrija, odgovarajućeg suradnika, te s njime otvorila nove vidike iz područja izotropne geometrije, koju se dotada smatralo prilično oskudnom.

Njezina je ideja bila istraživati pramenove krivulja 2. stupnja u izotropnoj ravnini, budući da u njoj postoji, za razliku od euklidske, sedam vrsta neraspadnutih krivulja 2. stupnja. Njihovi su zajednički radovi najprije dali klasifikaciju tih pramenova na temelju koje su izučavane pojedine vrste s obzirom na realnost i položaj temeljnih točaka tih pramenova. S obzirom na opsežnost područja, V. Ščurić-Čudovan je tu uključila svoje mlađe kolegice, Vlastu Szirovicza i Jelenu Beban Brkić. U ovom je trenutku teško nabrojati radove, te disertacije koje su u posljednjih dvadesetak godina proizašle iz te suradnje.

Iako se odlično služila njemačkim jezikom, naglasimo da nije bilo lagano održati predavanje iz geometrije prostora na njemačkom jeziku bez pomoći računala. Upravo je to intenzivno radila V. Ščurić-Čudovan i bila izvrsno prihvaćana na brojnim znanstvenim kongresima, savjetovanjima i drugim skupovima: Austrijski kongres matematike (Linz 1968., Beč 1973., Salzburg 1977., Insbruck 1982., Graz 1985.), Balkanski kongres matematičara (Beograd 1974.), Jugoslavensko-austrijski seminar za geometriju (Seggauberg 1986., Plitvice 1988.), Internacionalni simpozij za geometriju (Seggauberg 1987., 1988., 1989.), Simpozij za geometriju i diferencijalnu geometriju (Karlsruhe 1989.), Austrijsko-jugoslavenski geometrijski simpozij (Seggauberg 1990.), Kolokvij za konstruktivnu geometriju u spomen univ. red. prof. dr. H. Brauner (Seggauberg 1991.), Međunarodni geometrijski seminar (Seggauberg 1992.).

Tu treba dodati prisustvovanje V. Ščurić-Čudovan, u razdoblju od 1963. do 1990., svim Jugoslavenskim savjetovanjima nastavnika i asistenata Nacrtne geometrije koji su se u pravilu svake druge godine održavali u nekom drugom gradu nekadašnje Jugoslavije, te njeno prisustvovanje Kongresima matematičara, fizičara i astronoma Jugoslavije.

V. Ščurić-Čudovan bila je članica Društva matematičara, fizičara i astronoma SRH, Austrijskog društva matematičara, Jugoslavenskog udruženja za nacrtnu geometriju i inženjersku grafiku, Hrvatskog društva matematičara te Hrvatskog društva za geometriju i grafiku.

Kad je 1990. godine pokrenut novi međunarodni znanstveni časopis Mathematica Pannonica, čiji su osnivači bili akademik Gy. Maurer (Miskolc, Mađarska) i prof. emeritus dr. Hans Sachs (Leoben, Austria), u rad upravnog odbora, pored predstavnika iz Austrije, Češke, Italije, Mađarske, Slovačke i Poljske, bila je uključena i V. Ščurić-Čudovan kao predstavnica iz Hrvatske.

Tijekom niza godina V. Ščurić-Čudovan bila je aktivni sudionik u realizaciji pet znanstvenih projekata, od kojih su najznačajniji "Matematičke strukture, modeli i primjene" (1968. - 1975., 1976. - 1978.) i "Matematički modeli i strukture u geodeziji" (1981. - 1985., 1986. - 1990.).

Također, ne možemo se ne osvrnuti na brojne aktivnosti i funkcije koje je obnašala u stručnim i znanstvenim tijelima i organizacijama tijekom rada na Geodetskom fakultetu. Navedimo ovdje neke od njih: pored članstva u raznim komisijama i odborima, u nekoliko je mandata obnašala dužnost voditeljice Odjela za matematiku te predstojnice Zavoda za višu geodeziju, bila je tajnica Sindikalne podružnice Geodetskog fakulteta, i ono na čemu su joj svi bili niz godina izuzetno zahvalni, bila je odlična satničarka.

Osvrnimo se sada na 1994. godinu. Za sve geometričare s tehničkih fakulteta hrvatskih sveučilišta to je bila posebna i izuzetno važna godina. Naime, 16. lipnja u Zagrebu je održana Osnivačka skupština *Hrvatskog društva za konstruktivnu geometriju i kompjutorsku grafiku* (HDKGIKG), kasnije preimenovanog u *Hrvatsko društvo za geometriju i grafiku* (HDGG). Na istoj je sjednici Vlasta izabrana za prvu predsjednicu.

Teško je uopće zamisliti koliko je truda uloženo u sastavljanje svih potrebnih dokumenata kako bi Društvo zaživjelo i započelo s radom. Na sreću, Vlasta je imala nekoliko izvrsnih suradnika, na prvom mjestu profesora B. Kučinića koji je obnašao funkciju prvog potpredsjednika i docenticu Ivanku Babić kao prvu i dugogodišnju tajnicu.

Zašto nam je svima Društvo toliko važno? Kao prvo, zato što nam je dana prilika da se okupimo kao geometričari s tehničkih fakulteta. To nam je pomoglo pri rješavanju kadrovske problematike, razmjeni iskustava pri uvođenju i primjeni Bolonjskog procesa, modernizaciji nastave, implementaciji e-učenja u nastavu geometrije, znanstvenoj suradnji, organiziranju studijskih boravaka i pozvanih predavanja, prijavljivanju znanstvenih i razvojnih projekata, DAAD projekata, izradi zajedničkih repozitorija edukacijskog materijala. Osim toga Društvo organizira znanstveno-stručne skupove i izdaje znanstveno-stručni časopis KoG.

V. Ščurić-Čudovan je bila prva predsjednica HDGG-a i tu je funkciju obnašala do 2000. godine. Navedeni ciljevi i zadaće HDGG-a su ostali nepromijenjeni i kad je odstupila s mjesta predsjednice.

Da je Vlasta bila veliki znalac i zaljubljenik u branje gljiva, znali su gotovo svi u njenom okruženju. No, imala je ona i skrivenih sklonosti koje ćemo si ovdje dozvoliti iznijeti, u mladosti je učila svirati klavir i citru te je bila članica Akademskog zbora "Ivan Goran Kovačić".

Kako se na kraju zahvaliti i oprostiti od profesorice Ščurić-Čudovan osim da citiramo dijelove teksta koji je ona posvetila profesoru Ničeu:

... Zapao nas je častan ali i vrlo odgovoran zadatak da pomognemo evociranju sjećanja sviju nas na život, znanstveni i nastavni rad Vlaste Ščurić-Čudovan. Oprostite nam ako smo u tome pokatkad bile subjektivne. Za to postoje mnogi razlozi, a osnovni je taj što je profesorica Ščurić-Čudovan utjecala na tijek života mnogih geometričara: od poziva na rad na fakultetu, uvođenja i poticanja na znanstveni rad, mentorstva doktorske disertacije, otvaranjem novih znanstvenih područja, do uloge koju je imala pri osnivanju Hrvatskog društva za geometriju i grafiku i predsjedavanjem njime u puna tri mandata.

Ovo šturo nabrajanje krije u sebi mnogo, mnogo više. U prvom redu beskrajnu zahvalnost i poštovanje prema dr. sc. Vlasti Ščurić-Čudovan kao čovjeku i učitelju.

Popis radova

[1] V. ŠČURIĆ-ČUDOVAN, *Singularitäten des Majcenscchen Strahlenkomplexes*, Glasnik mat. fiz. i astr. 3(23) (1968), 117–139.

[2] V. ŠČURIĆ-ČUDOVAN, *Über die Rotationsflächen in einem Flächenbüschel 2. Grades und über ein Rotationsflächenbüschel*, Glasnik mat. fiz. i astr. 3(23) (1968), 275-286.

[3] V. ŠČURIĆ-ČUDOVAN, *Der orientierte Niče-sche Strahlkomplex eines Flächenbüschels 2. Grades*, I Teil, Rad JAZU 367 (1974), 151–205.

[4] V. ŠČURIĆ-ČUDOVAN, *Der orientierte Niče-sche Strahlkomplex eines Flächenbüschels 2. Grades*, II Teil, Rad JAZU 370 (1975), 57–91.

[5] V. ŠČURIĆ-ČUDOVAN, *Das* (F_k^2) *Flächenbüschel und eine Möglichkeit des Eintauchens des* (MK) *in den* (VN) *Komplex*, Rad JAZU 374 (1977), 57–91.

[6] V. ŠČURIĆ-ČUDOVAN, *Die Kongruenzen der Involutorstrahlen eines durch das* (F_k^3) *Flächenbüschel bestimmten* (VN) *Komplexes*, Rad JAZU 382 (1978), 65–90.

[7] V. ŠČURIĆ-ČUDOVAN, *Ergänzende Untersuchungen eines Büschels der homothetischen Flächen 2. Grades und einiger Komplexe, die durch dieses Büschel bestimmt werden*, Rad JAZU 386 (1980), 5–34.

[8] V. ŠČURIĆ-ČUDOVAN, *Einige Eigenschaften des* (VN) *Komplexes eines* (F_0^2) *Büschels*, Rad JAZU 396 (1982), 47–70.

[9] V. ŠČURIĆ-ČUDOVAN, *Einige Probleme die durch die Einteilung eines Bündels der Flächen 2. Grades in Büschel solcher Flächen entstanden sind, I Teil*, Rad JAZU 403 (1983), 33–55.

[10] V. ŠČURIĆ-ČUDOVAN, *Einige Probleme die durch die Einteilung eines Bündels der Flächen 2. Grades in* ∞^1 *Büschel solcher Flächen entstanden sind, II Teil*, Rad JAZU 421 (1986), 135–163.

[11] V. ŠČURIĆ-ČUDOVAN, *Weitere Untersuchungen in der Gesamtheit* (MF^2)*, I Teil, Komplex* (TK) *und Komplex* (MK), Rad JAZU 450 (1990), 9–21.

[12] V. ŠČURIĆ-ČUDOVAN, *Zur Klassifikationstheorie der Kegelschnittbüschel der isotropen Ebene, I Teil*, Rad JAZU 450 (1990), 41–51.

[13] V. ŠČURIĆ-ČUDOVAN, *Weitere Untersuchungen in der Gesamtheit* (MF^2)*, II Teil, Komplex* (VN), Rad HAZU 456 (1991), 39–57.

[14] H. SACHS, V. ŠČURIĆ-ČUDOVAN, *Zur Theorie der Flächen 2. Ordnung im Flaggenraum*, Rad HAZU 456 (1991), 197–216.

[15] V. ŠČURIĆ-ČUDOVAN, *Eine Kennzeichnung der speziellen Hyperbel der isotropen Ebene*, Österreichische Akad. der Wiss. Wien 201 (1992), 111–115.

[16] V. ŠČURIĆ-ČUDOVAN, H. SACHS, *Klassifikationstheorie der Kegelschnittbüschel vom Typ IV der isotropen Ebene, I*, Rad HAZU 470 (1995), 119–137.

[17] V. ŠČURIĆ-ČUDOVAN, H. SACHS, *Klassifikationstheorie der Kegelschnittbüschel vom Typ VI der isotropen Ebene, I*, Mathematica Pannonica 7/1 (1996), 47–67.

[18] V. ŠČURIĆ-ČUDOVAN, H. SACHS, *Klassifikationstheorie der Kegelschnittbüschel vom Typ IV der isotropen Ebene, II*, Rad HAZU 472 (1997), 27–53

Original scientific paper
Accepted 27. 10. 2016.

HELENA HALAS
EMA JURKIN

3rd Class Circular Curves in Quasi-Hyperbolic Plane Obtained by Projective Mapping

3rd Class Circular Curves in Quasi-Hyperbolic Plane Obtained by Projective Mapping

ABSTRACT

The metric in the quasi-hyperbolic plane is induced by an absolute figure $\mathcal{F}_{\mathbb{QH}} = \{F, \mathbf{f_1}, \mathbf{f_2}\}$, consisting of two real lines $\mathbf{f_1}$ and $\mathbf{f_2}$ incident with the real point F. A curve of class n is circular in the quasi-hyperbolic plane if it contains at least one absolute line.
The curves of the 3rd class can be obtained by projective mapping, i.e. obtained by projectively linked pencil of curves of the 2nd class and range of points. In this article we show that the circular curves of the 3rd class of all types, depending on their position to the absolute figure, can be constructed with projective mapping.

Key words: projectivity, circular curve of the 3rd class, quasi-hyperbolic plane

MSC2010: 51M15, 51N25

Cirkularne krivulje 3. razreda u kvazihiperboličnoj ravnini dobivene projektivnim preslikavanjem

SAŽETAK

U kvazihiperboličnoj ravnini metrika je inducirana s apsolutnom figurom $\mathcal{F}_{\mathbb{QH}} = \{F, \mathbf{f_1}, \mathbf{f_2}\}$ koja se sastoji od dva realna pravca $\mathbf{f_1}$ i $\mathbf{f_2}$ sa sjecištem u realnoj točki F. Za krivulju razreda n kažemo da je cirkularna u kvazihiperboličnoj ravnini ako sadrži barem jedan apsolutni pravac.
Krivulje 3. razreda se mogu dobiti projektivnim pridruživanjem između pramena krivulja 2. razreda i niza točaka. U ovom ćemo članku pokazati kako se svi tipovi cirkularnih krivulja 3. razreda mogu konstruirati projektivnim preslikavanjem.

Ključne riječi: projektivitet, cirkularna krivulja 3. razreda, kvazihiperbolična ravnina

1 Introduction

In the 19th century F. Klein founded the basis of the modern approach to geometry by defining it as the study of the properties of a space which are invariant under a given group of transformations. Later on this was know as *Erlangen program* according to the fact that Klein gave his first lecture on this subject at the University of Erlangen, [5]. There exist nine plane geometries with projective metric on a line and on a pencil of lines which can be parabolic, hyperbolic or elliptic. Due to Cayley's influence on Klein the geometries are denoted as Cayley-Klein projective metrics. Furthermore, each of these projective metrics can be embedded in the projective plane $\mathcal{P}_2 = \{\mathcal{P}, \mathcal{L}, \mathbf{I}\}$ where then an absolute figure, given as a proper or singular conic, induces the metric in the plane, [6, 7, 13] (for n-dimension see [12]).

The *quasi-hyperbolic plane*, denoted as $\mathbb{QH}_2$, is a projective plane where the metric is induced by the absolute figure $\mathcal{F}_{\mathbb{QH}} = \{F, \mathbf{f_1}, \mathbf{f_2}\}$ consisting of a pair of real lines $\mathbf{f_1}$, $\mathbf{f_2}$ intersecting at a real point F, [8, 10, 13]. The point F is called the *absolute point* and lines $\mathbf{f_1}$, $\mathbf{f_2}$ are called the *absolute lines*. In the Cayley-Klein model of the quasi-hyperbolic plane only the geometric objects inside of one projective angle between absolute lines are observed, while the points, lines and line segments inside the other angle are omitted. We observe the projectively extended quasi-hyperbolic plane where all points and lines of the projective plane are included as in [10].

In the sense of the Erlangen program, for the fundamental group of transformations in $\mathbb{QH}_2$ we use the *4-parameter general quasi-hyperbolic group of similarities* $\mathfrak{G}_4$, [8]. Transformations are of the form

$$[u_0, u_1, u_2] \mapsto [\alpha_0 u_0, \alpha_1 u_0 + \alpha_2 u_1 + \alpha_3 u_2, \alpha_4 u_0 \pm \alpha_3 u_1 \pm \alpha_2 u_2],$$
$$\alpha_i \in \mathbb{R}, \quad i = \{0 \dots 4\}, \quad \pm\alpha_2^2 \pm \alpha_3^2 \neq 0,$$

whereby the absolute figure $\mathcal{F}_{\mathbb{QH}}$ is determined by

$$F = (1,0,0),\ \mathbf{f_1} = [0,1,1],\ \mathbf{f_2} = [0,-1,1].$$

Definition 1 *A line passing through the absolute point F is called isotropic line and a point incident with the absolute line* $\mathbf{f_1}$ *or* $\mathbf{f_2}$ *is called isotropic point.*

For some further results on the basic notions in $\mathbb{QH}_2$ see [10].

Definition 2 *If the intersection of the curve* ζ *of the class n and the pencil* (F)*, in* $\mathbb{QH}_2$*, is the absolute line* $\mathbf{f_1}$ *with the intersection multiplicity t and the absolute line* $\mathbf{f_2}$ *with the intersection multiplicity r, than* ζ *is said to be a* $(t+r)$*-circular curve or circular curve of type* (t,r)*.* $t+r$ *is the degree of circularity, and if* $t+r=n$ *then the curve* ζ *is entirely circular.*

In further classification we will not distinguish circular curve of the type (t,r) from the one of the type (r,t) since the possibility of constructing one of them implies the possibility of constructing the other.

In accordance to the group $\mathfrak{G}_4$, proper curves of the 2nd class in $\mathbb{QH}_2$ are classified into nine types, see [1, 10]. They can also be classified in accordance to its degree and type of circularity as following:

i) non-circular curves of the 2nd class: ellipses (e), hyperbolas (h_1,h_2,h_3), parabolas (p);

ii) 1-circular curves of the 2nd class: special hyperbolas $(h_{s1},h_{s2}$, type of circularity $(1,0))$;

iii) 2-circular curves of the 2nd class: circles $(c$, type of circularity $(1,1))$, special parabolas $(p_s$, type of circularity $(2,0))$.

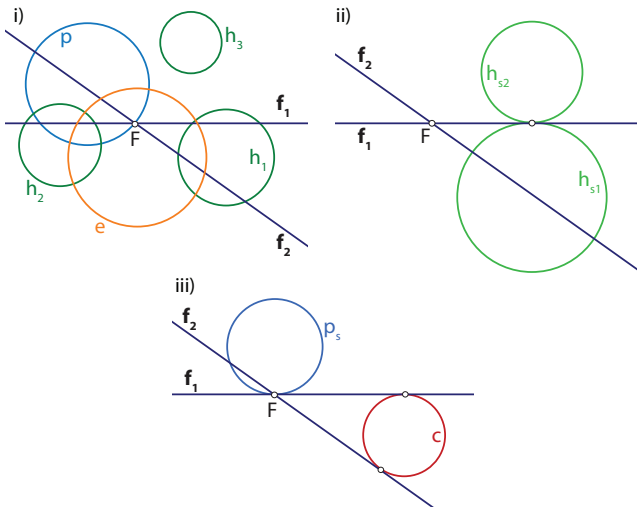


Figure 1: *Classification of the curves of the 2nd class in* $\mathbb{QH}_2$ *according to their degree of circularity*

Remark 1 *In all figures of the article the class curves are drawn as point objects as we are used to, although they are line envelopes in the quasi-hyperbolic plane.*

The circular curves of the 3rd class can be classified, according to their position with respect to $\mathcal{F}_{\mathbb{QH}}$, into the following types and subtypes:

- 1-circular curves of the 3rd class
  - type of circularity (1,0)
    a) the curve contains the absolute line $\mathbf{f_1}$ and two isotropic lines that are conjugate imaginary;
    b) the curve contains the absolute line $\mathbf{f_1}$ and two isotropic lines that are real and distinct;
    c) the curve contains the absolute line $\mathbf{f_1}$ and two isotropic lines that coincide;
    d) the curve contains the absolute line $\mathbf{f_1}$ and an isotropic double line with two conjugate imaginary tangent points (isolated double line);
    e) the curve contains the absolute line $\mathbf{f_1}$ and an isotropic double line with two real and distinct tangent points (double tangent line);
    f) the curve contains the absolute line $\mathbf{f_1}$ and an isotropic double line with two tangent points that coincide (inflection line);

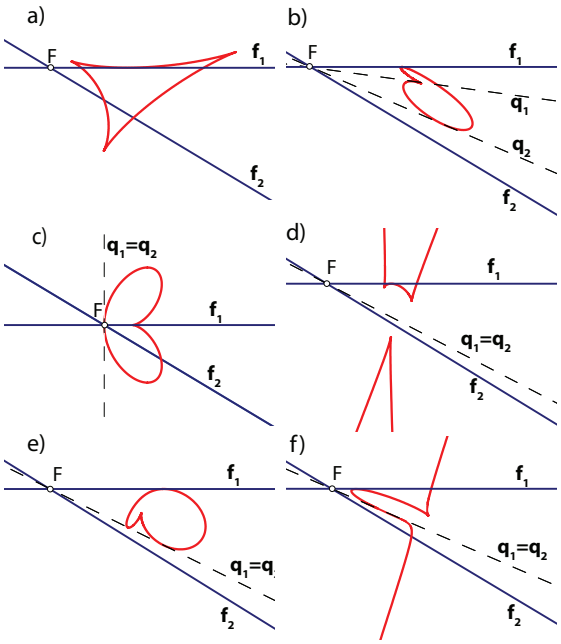


Figure 2: *Classification of the 1-circular curves of the 3rd class in* $\mathbb{QH}_2$

- 2-circular of the 3rd class
  - type of circularity (1,1)
    a) the curve contains both absolute lines $\mathbf{f_1}$ and $\mathbf{f_2}$;
  - type of circularity (2,0)
    b) the curve contains the absolute line $\mathbf{f_1}$ where the absolute point F is the tangent point;
    c) the absolute line $\mathbf{f_1}$ is an isolated double line of the curve;
    d) the absolute line $\mathbf{f_1}$ is a double tangent line of the curve;
    e) the absolute line $\mathbf{f_1}$ is an inflection line of the curve;

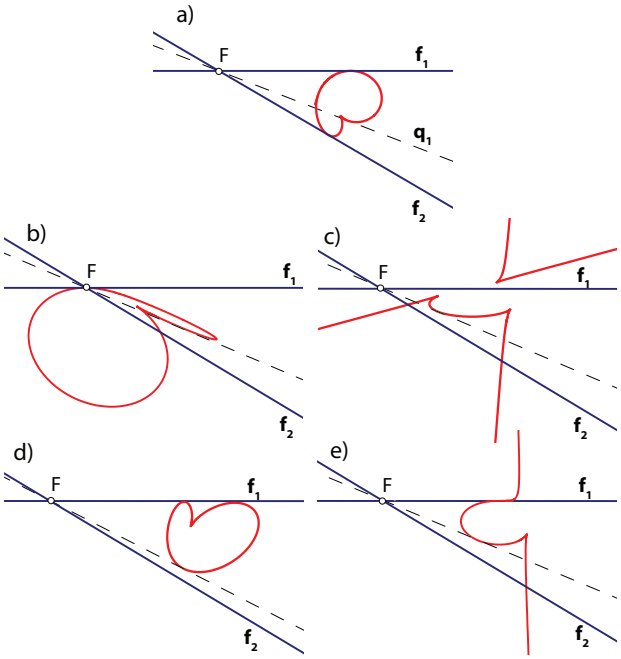


Figure 3: *Classification of the 2-circular curves of the 3rd class in* $\mathbb{QH}_2$

- 3-circular curves of the 3rd class
  - type of circularity (2,1)
    a) the curve contains both absolute lines $\mathbf{f_1}$, $\mathbf{f_2}$ and the absolute point F is the tangent point of the line $\mathbf{f_1}$;
    b) the curve contains both absolute lines $\mathbf{f_1}$, $\mathbf{f_2}$ such that $\mathbf{f_1}$ is an isolated double line;
    c) the curve contains both absolute lines $\mathbf{f_1}$, $\mathbf{f_2}$ such that $\mathbf{f_1}$ is a double tangent line
    d) the curve contains both absolute lines $\mathbf{f_1}$, $\mathbf{f_2}$ such that $\mathbf{f_1}$ is an inflection line;
  - type of circularity (3,0)
    e) the absolute line $\mathbf{f_1}$ is a double tangent with one tangent point at the absolute point F;
    f) the absolute line $\mathbf{f_1}$ is an inflection line with the tangent point at the absolute point F;
    g) the curve contains the absolute line $\mathbf{f_1}$ and has a cusp at the absolute point F.

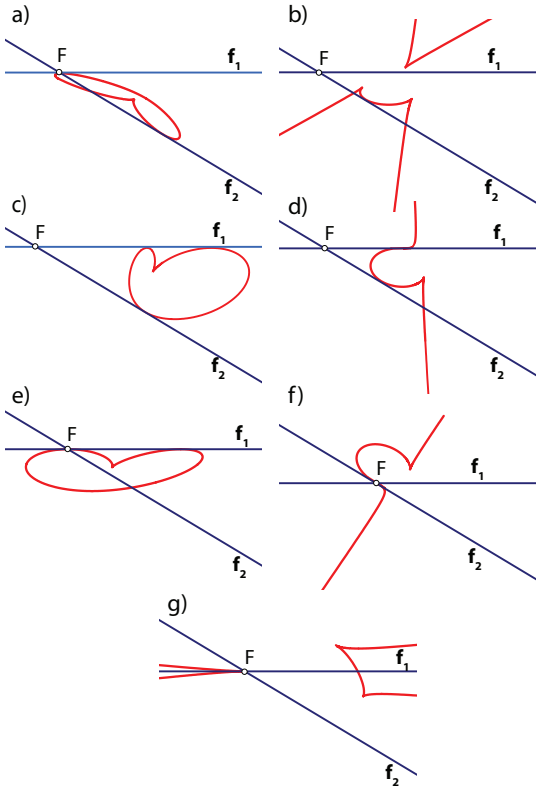


Figure 4: *Classification of the 3-circular or entirely circular curves of the 3rd class in* $\mathbb{QH}_2$

The aim of this article is to construct every type of circular curves of the 3rd class in the quasi-hyperbolic plane by using projective mapping. The classification of circular curves, according to their position with respect to the absolute figure, obtained by projective mapping in some other Cayley-Klein projective metrics can be found in [2, 3, 4, 11].

2 Projective mapping

Let points P_1, P_2 and curves of the 2nd class ζ_1, ζ_2 be given. The associated symmetric bilinear form for the 2nd class curves is given with

$$\begin{aligned}\zeta_1 \ldots \quad & f_{\zeta_1}(\mathbf{u},\mathbf{v}) := \mathbf{u}^T C_1 \mathbf{v} = 0,\\ \zeta_2 \ldots \quad & f_{\zeta_2}(\mathbf{u},\mathbf{v}) := \mathbf{u}^T C_2 \mathbf{v} = 0,\end{aligned}$$

and in the following the the curves ζ_1 and ζ_2 will be identified with its corresponding matrix representation C_1 and C_2. The result of a projective mapping

$$\begin{aligned}\pi : [C_1, C_2] &\mapsto [P_1, P_2],\\ \pi(C_1 + \lambda C_2) &= P_1 + \lambda P_2, \quad \forall \lambda \in \mathbb{R} \cup \infty,\end{aligned}$$

between the pencil of the 2nd class curves $[C_1, C_2]$ and the range of points $[P_1, P_2]$ is a curve of the 3rd class k_π^3 given by the equation

$$k_\pi^3 \ldots \; F(\mathbf{u}) \equiv \mathbf{u}^T C_1 \mathbf{u} \cdot P_2^T \mathbf{u} - \mathbf{u}^T C_2 \mathbf{u} \cdot P_1^T \mathbf{u} = 0. \qquad (1)$$

The curve k_π^3 contains the following nine lines: four basic lines of the pencil $[C_1, C_2]$, basic line of the range $[P_1, P_2]$, two intersection lines of ζ_1 and (P_1), two intersection lines of ζ_2 and (P_2). It is known that the number of lines required for determination of a curve of the 3rd class is nine, but nine lines do not determine a single curve of the 3rd class in every case, [9]. For defining the projectivity we need three pairs of elements (ζ_1, P_1), (ζ_2, P_2) and (ζ_3, P_3). Furthermore, we should point out that although the proportional matrices C_1, C_2, P_1, P_2 and αC_1, βC_2, γP_1, δP_2 represent the same two curves of the 2nd class and two points, the corresponding curves of the 3rd class are different, but they properties of circularity stay the same.

Let us observe a line $\mathbf{v} \in k_\pi^3$, such that the curve k_π^3 is obtained by a projective mapping π and without loss of generality we can assume $\mathbf{v} \in C_1$, $P_1 \in \mathbf{v}$ thus

$$\mathbf{v}^T C_1 \mathbf{v} = 0, \; P_1^T \mathbf{v} = 0$$

is valid. The behaviour of the line $\mathbf{v}$ can be studied by observing the intersection lines of curve k_π^3 and a pencil (X) such that $X \in \mathbf{v}$. Therefore an arbitrary point X on the line $\mathbf{v}$ can be given as

$$X \ldots \quad \mathbf{v} + t\mathbf{w}, \; t \in \mathbb{R} \cup \infty,$$

hence intersection lines of k_π^3 and (X) are determined by the roots of the following polynomial

$$F(\mathbf{v} + t\mathbf{w}) = F_1(\mathbf{v},\mathbf{w}) + t^2 F_2(\mathbf{v},\mathbf{w}) + t^3 F_3(\mathbf{v},\mathbf{w}), \qquad (2)$$

where

$$\begin{aligned}F_1(\mathbf{v},\mathbf{w}) &= 2P_2^T \mathbf{v} \cdot \mathbf{v}^T C_1 \mathbf{w} - P_1^T \mathbf{w} \cdot \mathbf{v}^T C_2 \mathbf{v},\\ F_2(\mathbf{v},\mathbf{w}) &= P_2^T \mathbf{v} \cdot \mathbf{w}^T C_1 \mathbf{w} + 2P_2^T \mathbf{w} \cdot \mathbf{v}^T C_1 \mathbf{w} - 2P_1^T \mathbf{w} \cdot \mathbf{v}^T C_2 \mathbf{w},\\ F_3(\mathbf{v},\mathbf{w}) &= P_2^T \mathbf{w} \cdot \mathbf{w}^T C_1 \mathbf{w} - P_1^T \mathbf{w} \cdot \mathbf{w}^T C_2 \mathbf{w}.\end{aligned}$$

From (2) we can conclude the following statements:

- tangent point on the regular line $\mathbf{v}$ of the curve k_π^3 is given by the equation
$$F_1(\mathbf{v},\mathbf{w}) = 0; \qquad (3)$$
- necessary condition to gain $\mathbf{v}$ as a double line of the curve k_π^3 is
$$F_1(\mathbf{v},\mathbf{w}) = 0, \quad \forall \mathbf{w}; \qquad (4)$$
- tangent points on a double line $\mathbf{v}$ of the curve k_π^3 are given by the equation
$$F_2(\mathbf{v},\mathbf{w}) = 0; \qquad (5)$$
- necessary condition to gain a cusp at X on the line $\mathbf{v}$ for the curve k_π^3 is if the equation (5) is valid for every line $\mathbf{w}$ such that $X \in \mathbf{w}$.

Remark 2 *Generally there are three possible positions for a curve of the 2nd class ζ_1 and its line $\mathbf{v}$:*

a) the curve ζ_1 is a proper curve and the equation $\mathbf{v}^T C_1 \mathbf{w} = 0$ is its the tangent point on the line $\mathbf{v}$;

b) the curve ζ_1 is a singular curve, but $\mathbf{v}$ is not its singular line, i. e. $\zeta_1 = (Z_1) \cup (\hat{Z}_1)$, $Z_1 \in \mathbf{v}$, $\hat{Z}_1 \notin \mathbf{v}$. The point Z_1 is the tangent point at $\mathbf{v}$ and its equation is $\mathbf{v}^T C_1 \mathbf{w} = 0$;

c) the curve ζ_1 is a singular curve and $\mathbf{v}$ is its singular line, i. e. $\zeta_1 = (Z_1) \cup (\hat{Z}_1)$, $Z_1, \hat{Z}_1 \in \mathbf{v}$. The equation $\mathbf{v}^T C_1 \mathbf{w} = 0$ is valid for every line $\mathbf{w}$.

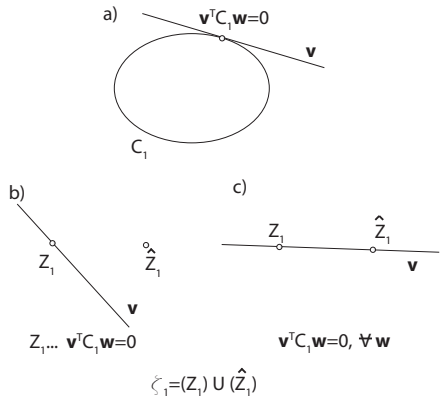


Figure 5: *Positions of the 2nd class curve ζ_1 and its line $\mathbf{v}$*

Furthermore, in respect to the basic elements of the mapping π there are four different positions for a line $\mathbf{v}$ of the curve k_π^3 such that $\mathbf{v} \in \zeta_1$, $P_1 \in \mathbf{v}$:

- $\mathbf{v} \notin \zeta_2, \quad P_2 \notin \mathbf{v}$;
- $\mathbf{v} \in \zeta_2, \quad P_2 \notin \mathbf{v}$;
- $\mathbf{v} \notin \zeta_2, \quad P_2 \in \mathbf{v}$;
- $\mathbf{v} \in \zeta_2, \quad P_2 \in \mathbf{v}$.

Taking in consideration the remark 2 we could discuss all these cases, but in the next section we will present only some of them. By selecting different corresponding pairs (ζ_1, P_1), (ζ_2, P_2) of the projective mapping π we can obtain circular curves of the same type. Therefore, for every type we will present one construction.

Figure 6 represents an example of the entirely circular curve of the 3rd class obtained by the projective mapping π where the corresponding pairs of the mapping are (ζ_1, P_1), (ζ_2, P_2), (ζ_3, P_3), such that curves $\zeta_1 = (Z_1) \cup (\hat{Z}_1)$ and $\zeta_2 = (Z_2) \cup (\hat{Z}_2)$ are singular. The red curve is obtained as a set of tangent points of the curve k^3_π calculated in the software *Wolfram Mathematica*, and the figure is drawn in dynamic software *Geometer's Sketchpad*. As mentioned earlier in Remark 1 it is customary to represent curves as point objects, therefore on the remaining figures in the article curves will be presented in this way.

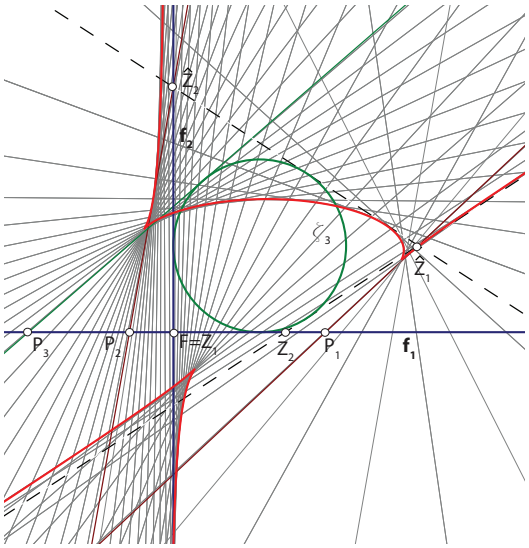


Figure 6: *Circular curve of the 3rd class with the circularity type (2,1) in* $\mathbb{QH}_2$

3 1-circular curves of the 3rd class in $\mathbb{QH}_2$

From the equation (1), as we already mentioned, the curve k^3_π obtained by projective mapping $\pi : [C_1, C_2] \mapsto [P_1, P_2]$ contains nine specific lines, therefore only by picking certain pencils of the 2nd class curves or ranges of points we can ensure the circularity of the curve k^3_π. For instance, if one basic line of the pencils of the 2nd class curves or the basic line of the point range is the absolute line $\mathbf{f_1}$ then the obtained curve k^3_π is 1-circular curve of type $(1,0)$.

Let us observe the case $\mathbf{v} \in \zeta_1$, $P_1 \in \mathbf{v}$, $\mathbf{v} \notin \zeta_2$, $P_2 \notin \mathbf{v}$ when the curve ζ_1 is a proper curve of the 2nd class. From the equation (3) we can conclude that if P_1 is the tangent point of the curve ζ_1 then P_1 is also a tangent point of the curve k^3_π.

Theorem 1 *Let* $[C_1, C_2]$ *be a pencil of 2nd class curves and* $[P_1, P_2]$ *a range of isotropic points in* $\mathbb{QH}_2$. *The result of the projective mapping* $\pi : [C_1, C_2] \mapsto [P_1, P_2]$ *gives a 1-circular curve of the 3rd class* k^3_π *of type (1,0) or (0,1). If the curve of the 2nd class corresponding to the absolute point F is an ellipse, hyperbola or parabola then the remaining two isotropic lines of* k^3_π *are conjugate imaginary, real and distinct or coincide respectively.*

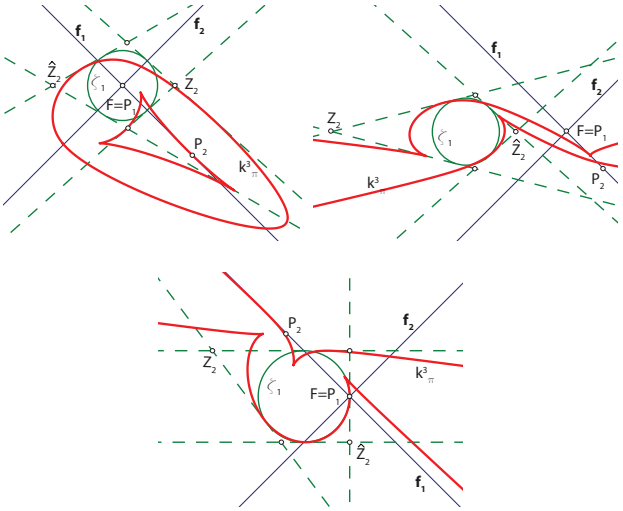


Figure 7: *1-circular curves of the 3rd class of type a, b and c*

Let us observe the case $\mathbf{v} \in \zeta_1, \zeta_2$, $P_1 \in \mathbf{v}$, $P_2 \notin \mathbf{v}$ when the curve ζ_1 is a singular curve of the 2nd class with a singular line $\mathbf{v}$, $\zeta_1 = (Z_1) \cup (\hat{Z}_1)$, $Z_1, \hat{Z}_1 \in \mathbf{v}$. The curves of the pencil $[C_1, C_2]$ are touching at some point on the line $\mathbf{v}$ and the condition (4) is fulfilled, hence the line $\mathbf{v}$ is a double line of the curve k^3_π. The tangent points of the double line are given with the equation (5).

Theorem 2 *Let* $[C_1, C_2]$ *be a pencil of 2nd class curves with a common tangent point on the isotropic line* $\mathbf{v}$, $[P_1, P_2]$ *a range of isotropic points on the absolute line* $\mathbf{f_1}$ *and the curve* k^3_π *the result of the projective mapping* $\pi : [C_1, C_2] \mapsto [P_1, P_2]$ *in* $\mathbb{QH}_2$. *If the absolute point F is the corresponding point to the singular 2nd class curve with the singular line* $\mathbf{v}$, *then the curve* k^3_π *is a 1-circular curve of the 3rd class of type (1,0) with the double line* $\mathbf{v}$.

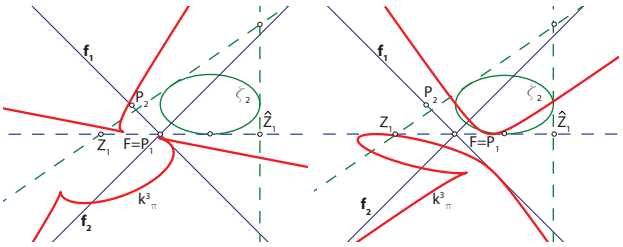


Figure 8: *1-circular curves of the 3rd class of type d and e*

Let us observe the case $P_1, P_2 \in \mathbf{v}$, $\mathbf{v} \in \zeta_1, \zeta_2$. The condition (4) is fulfilled, thus the line $\mathbf{v}$ is the double line of k^3_π. Tangent points on the line $\mathbf{v}$ are given with the equation (5) which in this case is

$$P_2^T \mathbf{w} \cdot \mathbf{v}^T C_1 \mathbf{w} - P_1^T \mathbf{w} \cdot \mathbf{v}^T C_2 \mathbf{w} = 0. \tag{6}$$

One tangent point at the line $\mathbf{v}$ of the curve k^3_π coincides with the tangent point of the curve ζ_1 if and only if P_1 is the tangent point of ζ_1 or curve ζ_1 and ζ_2 are touching.

If the latter case, if the curves ζ_1, ζ_2 are touching then the whole pencil $[C_1, C_2]$ has a common tangent point on the line $\mathbf{v}$. Furthermore, there exists a singular 2nd class curve with the singular line $\mathbf{v}$ and with out loss of generality we can assume it is the curve ζ_1. The equation (6) is of the form

$$P_1^T \mathbf{w} \cdot \mathbf{v}^T C_2 \mathbf{w} = 0,$$

hence one tangent point on the line $\mathbf{v}$ of the curve k^3_π is the common tangent point of $[C_1, C_2]$ while the other one is the point of the range that corresponds to the singular 2nd class curve of $[C_1, C_2]$ with the singular $\mathbf{v}$. These two tangent points can coincide and in that case the line $\mathbf{v}$ is an inflection line of the curve k^3_π.

Theorem 3 *Let $[C_1, C_2]$ be a pencil of special hyperbolas of type $(1,0)$ with a common tangent point on the isotropic line $\mathbf{v}$, $[P_1, P_2]$ a range of points on $\mathbf{v}$ and the curve k^3_π the result of the projective mapping $\pi : [C_1, C_2] \mapsto [P_1, P_2]$ in $\mathbb{QH}_2$. If the corresponding point to the singular 2nd class curve with the singular line $\mathbf{v}$ is the common tangent point of the pencil $[C_1, C_2]$, then the curve k^3_π is 1-circular curve of type (1,0) with the inflection line $\mathbf{v}$.*

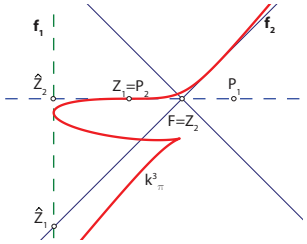


Figure 9: *1-circular curve of the 3rd class of type f*

3.1 2-circular curves of the 3rd class

Theorem 4 *Let $[C_1, C_2]$ be a pencil of circles and $[P_1, P_2]$ a range of points in $\mathbb{QH}_2$. The result of the projective mapping $\pi : [C_1, C_2] \mapsto [P_1, P_2]$ gives a 2-circular curve of the 3rd class k^3_π of type (1,1).*

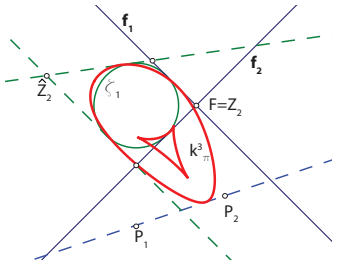


Figure 10: *2-circular curve of the 3rd class of type a*

If in the case $P_1 \in \mathbf{v}$, $P_2 \notin \mathbf{v}$, $\mathbf{v} \in \zeta_1, \zeta_2$ we assume that the curve ζ_1 is a proper curve, then the equation (3) is of the form

$$P_2^T \mathbf{w} \cdot \mathbf{v}^T C_1 \mathbf{w} = 0.$$

Hence, the conclusion is that the tangent point at the line $\mathbf{v}$ of the curve k^3_π coincides with the tangent point of the curve ζ_1. Specially, if the curves of the pencil $[C_1, C_2]$ are touching at a point on the line $\mathbf{v}$ then this common tangent point of the pencil $[C_1, C_2]$ is also the tangent point of the curve k^3_π.

Theorem 5 *Let $[C_1, C_2]$ be a pencil of special parabolas of type (2,0), $[P_1, P_2]$ a range of points and the curve k^3_π the result of the projective mapping $\pi : [C_1, C_2] \mapsto [P_1, P_2]$ in $\mathbb{QH}_2$. Then the curve k^3_π is a 2-circular curve of type (2,0) where the absolute point F is the tangent point at the absolute line $\mathbf{f_1}$.*

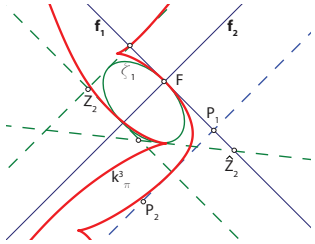


Figure 11: *2-circular curve of the 3rd class of type b*

From the observations before Theorem 2 follows also

Theorem 6 *Let $[C_1, C_2]$ be a pencil of special parabolas of type $(2,0)$, $[P_1, P_2]$ a range of points and the curve k^3_π the result of the projective mapping $\pi : [C_1, C_2] \mapsto [P_1, P_2]$ in $\mathbb{QH}_2$. If the isotropic point of the range $[P_1, P_2]$ incident with the absolute line $\mathbf{f_1}$ corresponds to the singular curve with the singular line $\mathbf{f_1}$ of the pencil $[C_1, C_2]$, then the curve k^3_π is a 2-circular curve of the 3rd class of type (2,0) with the double line $\mathbf{f_1}$.*

In this case the double line of the curve k^3_π can only be an isolated double line or a double tangent.

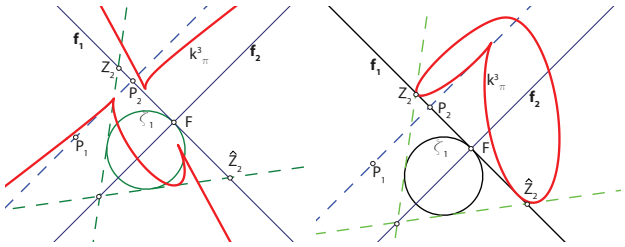


Figure 12: *2-circular curves of the 3rd class of type c and d*

From the observation before Theorem 3 we can ensure that the double line of the curve k^3_π is an inflection line:

Theorem 7 *Let $[C_1,C_2]$ be a pencil of special hyperbola of type $(1,0)$ with a common tangent point on the absolute line $\mathbf{f_1}$, $[P_1,P_2]$ a range of isotropic points on the absolute line $\mathbf{f_1}$ and the curve k^3_π the result of the projective mapping $\pi:[C_1,C_2]\mapsto[P_1,P_2]$ in $\mathbb{QH}_2$. If the corresponding point to the singular curve with the singular line $\mathbf{f_1}$ is the common tangent point of the pencil $[C_1,C_2]$, then the curve k^3_π is a circular curve of type $(2,0)$ with the inflection line $\mathbf{f_1}$.*

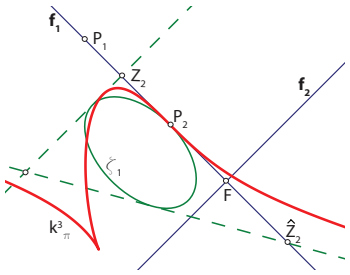


Figure 13: *2-circular curve of the 3rd class of type e*

3.2 3-circular curves or entirely circular curves

Generally, we already concluded that if there exists a point of the range $[P_1,P_2]$ which is the tangent point of its corresponding curve of the 2nd class in the pencil $[C_1,C_2]$, then this point is also a tangent point for k^3_π. Thus, the following theorem is valid:

Theorem 8 *Let $[C_1,C_2]$ be a pencil of the 2nd class curves, $[P_1,P_2]$ a range of isotropic points on the absolute line $\mathbf{f_1}$ and the curve k^3_π the result of the projective mapping $\pi:[C_1,C_2]\mapsto[P_1,P_2]$ in $\mathbb{QH}_2$. If the pencil $[C_1,C_2]$ contains a special parabola of type $(0,2)$ whose corresponding point is the absolute point F, then the curve k^3_π is a 3-circular curve of type $(1,2)$. The absolute point F is the tangent point at the line $\mathbf{f_2}$ of the curve k^3_π.*

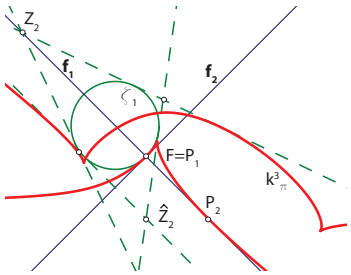


Figure 14: *1-circular curve of the 3rd class of type a*

From the observations before Theorem 3 follows also

Theorem 9 *Let $[C_1,C_2]$ be a pencil of circles, $[P_1,P_2]$ a range of isotropic points on the absolute line $\mathbf{f_1}$ and the curve k^3_π the result of the projective mapping $\pi:[C_1,C_2]\mapsto[P_1,P_2]$ in $\mathbb{QH}_2$. Then the curve k^3_π is an entirely circular curve of the circularity type $(2,1)$, where the absolute line $\mathbf{f_1}$ is a double isolated line or a double tangent line.*

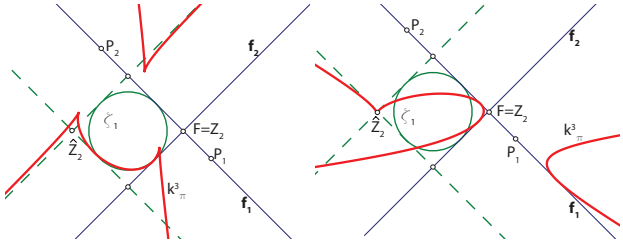


Figure 15: *3-circular curves of the 3rd class of type b and c*

Theorem 10 *Let $[C_1,C_2]$ be a pencil of circles with a common tangent point on the absolute line $\mathbf{f_1}$, $[P_1,P_2]$ a range of isotropic points on the absolute line $\mathbf{f_1}$ and the curve k^3_π the result of the projective mapping $\pi:[C_1,C_2]\mapsto[P_1,P_2]$ in $\mathbb{QH}_2$. If the corresponding point to the singular 2nd class curve with the singular line $\mathbf{f_1}$ is the common tangent point, then the curve k^3_π is a 3-circular curve of type $(2,1)$, where the absolute line $\mathbf{f_1}$ is a inflection line.*

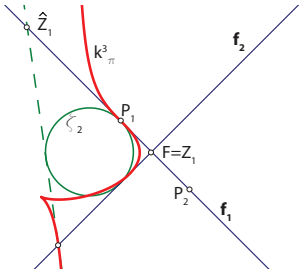


Figure 16: *3-circular curve of the 3rd class of type d*

Theorem 11 *Let $[C_1,C_2]$ be a pencil of special parabolas of type $(2,0)$, $[P_1,P_2]$ a range of isotropic points on the absolute line $\mathbf{f_1}$ and the curve k^3_π the result of the projective mapping $\pi:[C_1,C_2]\mapsto[P_1,P_2]$ in $\mathbb{QH}_2$. The curve k^3_π is*

an entirely circular curve of the 3rd class of the circularity type $(3,0)$ *with the double line* $\mathbf{f_1}$. *The absolute point F is one tangent point on the double line* $\mathbf{f_1}$, *and the other tangent point is the point of the range* $[P_1,P_2]$ *that corresponds to the singular curve of* $[C_1,C_2]$ *with the singular line* $\mathbf{f_1}$. *Specially, if this latter point coincides with F then line* $\mathbf{f_1}$ *is an inflection line.*

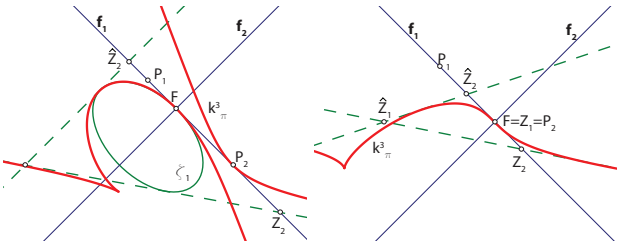


Figure 17: *3-circular curves of the 3rd class of type e and f*

From the Theorem 5 and the observation before we can also conclude

Theorem 12 *Let* $[C_1,C_2]$ *be a pencil of special parabolas of type* $(2,0)$, $[P_1,P_2]$ *a range of points and the curve* k^3_π *the result of the projective mapping* $\pi:[C_1,C_2]\mapsto[P_1,P_2]$ *in* $\mathbb{QH}_2$. *If the corresponding point to the singular curve whose one pencil is* (F) *is the isotropic point of* $[P_1,P_2]$ *incident with the line* $\mathbf{f_1}$, *then the curve* k^3_π *is a 3-circular curve of type* $(3,0)$ *with a cusp at the point F.*

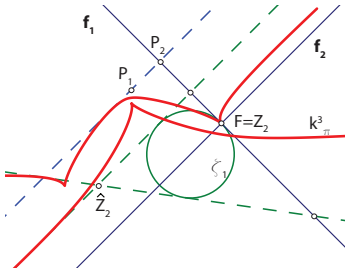


Figure 18: *3-circular curve of the 3rd class of type g*

References

[1] H. Halas, N. Kovačević, A. Sliepčević, Line Inversion in the Quasi-Hyperbolic Plane, *Proceedings ICGG 2014*, Innsbruck, Austria, 739–748.

[2] E. Jurkin, Circular Cubics in pseudo-Euclidean plane, *Novi Sad J. Math.* **44**/2 (2014), 195–206.

[3] E. Jurkin, N. Kovačević, Entirely circular quartics in the pseudo-Euclidean plane, *Acta Math. Hungar.* **134**/4 (2012), 27–45.

[4] E. Jurkin, Circular quartics in the isotropic plane generated by projectively linked pencils of conics, *Acta Math. Hungar.* **130**/1–2 (2011), 35-49.

[5] F. Klein, *Elementary Mathematics from an advanced Standpoint Geometry*, Dover, New York, 2004.

[6] N. M. Makarova, On the projective metrics in plane, *Učenye zap. Mos. Gos. Ped. in-ta* **243** (1965), 274–290. (Russian)

[7] M. D. Milojević, Certain Comparative examinations of plane geometries according to Cayley-Klein, *Novi Sad J. Math.* **29**/3, 1999, 159–167

[8] H. Sachs, *Ebene Isotrope Geometrie*, Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden, 1987.

[9] S. Salmon, *Higher plane curves*, Chelsea Publishing Company, New York, 1879.

[10] A. Sliepčević, I. Božić, H. Halas, Introduction to the Planimetry of the Quasi-Hyperbolic Plane, *KoG* **17** (2013), 58–64.

[11] A. Sliepčević, V. Szirovicza, A classification and construction of entirely circular cubics in the hyperbolic plane, *Acta Math. Hungar.* **104**/3 (2004), 185–201.

[12] D. M. Y Sommerville, Classification of geometries with projective metric, *Proc. Ediburgh Math. Soc.* **28** (1910), 25–41.

[13] I. M. Yaglom, B. A. Rozenfeld, E. U. Yasinskaya, Projective metrics, *Russ. Math Surreys* **19**/5 (1964), 51–113.

Helena Halas
e-mail: hhalas@grad.hr
Faculty of Civil Engineering, University of Zagreb,
10 000 Zagreb, Kačićeva 26, Croatia

Ema Jurkin
email: ejurkin@rgn.hr
Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb,
10 000 Zagreb, Pierottijeva 6, Croatia

Original scientific paper
Accepted 4. 11. 2016.

ANA SLIEPČEVIĆ
IVANA BOŽIĆ DRAGUN

Introduction to Planimetry of Quasi-Elliptic Plane

Introduction to Planimetry of Quasi-Elliptic Plane

ABSTRACT

The quasi-elliptic plane is one of nine projective-metric planes where the metric is induced by the absolute figure $\mathcal{F}_{QE} = \{j_1, j_2, F\}$ consisting of a pair of conjugate imaginary lines j_1 and j_2, intersecting at the real point F. Some basic geometric notions, definitions, selected constructions and a theorem in the quasi-elliptic plane will be presented.

Key words: quasi-elliptic plane, perpendicular points, central line, qe-conic classification, hyperosculating qe-circle, envelope of the cental lines

MSC2010: 51A05, 51M10, 51M15

Uvod u planimetriju kvazieliptičke ravnine

SAŽETAK

Kvazieliptička ravnina jedna je od devet projektivno metričkih ravnina. Apsolutnu figuru $\mathcal{F}_{QE} = \{j_1, j_2, F\}$ određuju dva imaginarna pravca j_1 i j_2 i njihovo realno sjecište F.
U ovom radu definirat ćemo osnove pojmove, prikazati odabrane konstrukcije i dokazati jedan teorem.

Ključne riječi: kvazieliptička ravnina, okomite točke, centrala, klasifikacija qe-konika, hiperoskulacijska qe-kružnica, omotaljka centrala

1 Introduction

This paper begins the study of the quasi-elliptic plane from the constructive and synthetic point of view. We will see although the geometry denoted as quasi-elliptic is dual to Euclidean geometry it is a very rich topic indeed and there are many new and unexpected aspects.
In this paper some basic notations concerning the quasi-elliptic conic and some selected constructions and a theorem will be presented. It is known that there exist nine geometries in plane with projective metric on a line and on a pencil of lines which are denoted as Cayley-Klein projective metrics and they have been studied by several authors, such as [2], [3], [4], [8], [9], [10], [13], [14], [15], [16].
The quasi-elliptic geometry, further in text qe-geometry, has elliptic measure on a line and parabolic measure on a pencil of lines. In the quasi-elliptic plane, further in text qe-plane, the metric is induced by the absolute figure $\mathcal{F}_{QE} = \{j_1, j_2, F\}$, i.e. a pair of conjugate imaginary lines j_1 and j_2, incident with the real point F. The lines j_1 and j_2 are called the **absolute lines**, while the point F is called the **absolute point**. In the Cayley-Klein model of the qe-plane only the points, lines and segments inside of one projective angle between the absolute lines are observed. In this paper all points and lines of the qe-plane embedded in the real projective plane $\mathbb{P}_2(\mathbb{R})$ are observed. It is suitable to obtain a line as a basic element, and a point as a pencil of lines (for example a curve is an envelope of lines; quadratic transformation in the qe-plane maps pencil of lines into the second class curve). Using an elliptic involution on the pencil (F) the absolute triple $\mathcal{F}_{QE} = \{j_1, j_2, F\}$ can be given as follows:

- An elliptic involution on the pencil (F) is determined by two arbitrary chosen pairs of corresponding lines a_1, a_2; b_1, b_2. An elliptic involution (F) has the absolute lines j_1 and j_2 for double lines ([1], p.244-245, [6], p.46).
  Notice that the absolute point F can be finite (Figure 1a) or at infinity (Figure 1c).
  In this paper the model were involutory pair of corresponding lines are perpendicular to each other in Euclidean sense (Figure 1b) is used in a way that only the absolute point F is presented.

- The absolute point F is inside the conic k. Pairs of conjugate lines with respect to a conic k determine aforementioned elliptic involution (F). The absolute lines j_1 and j_2 are double lines for the involution (F) and in this case they are a pair of imaginary tangent lines to k from the absolute point F (Figure 1d).

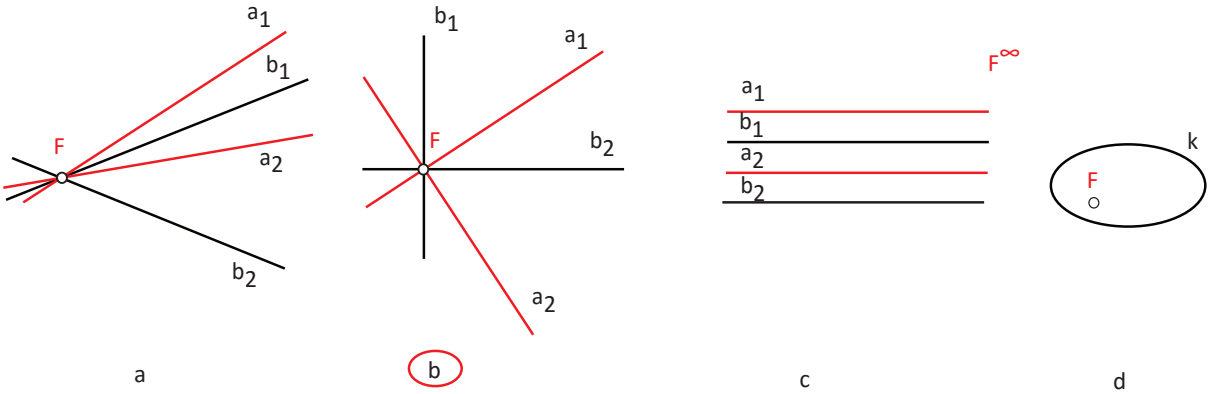


Figure 1

2 Basic notation and selected constructions in the quasi-elliptic plane

For the points and the lines in the qe-plane the following are defined:

- *isotropic lines* - the lines incident with the absolute point F,
- *isotropic points* - the imaginary points incident with one of the absolute line j_1 or j_2,
- *parallel points* - two points incident with the same isotropic line,
- *perpendicular lines* - if at least one of two lines is an isotropic line,
- *perpendicular points* - two points A, A_1 that lie on a pair of corresponding lines a, a_1 of an elliptic (absolute) involution (F).

Remark. The perpendicularity of points in qe-plane is determine by the absolute involution, therefore an elliptic involution (F) is a circular involution in the qe-plane. ([7], p.75)
Notice that the absolute point F is parallel and perpendicular to each point in the qe-plane. Furthermore, in the qe-plane there are no parallel lines.

A brief review of some basic construction

Example 1 *Let the absolute figure $\mathcal{F}_{QE}$ of the qe-plane be given with the involutory pencil (F) (Figure 1b). Let A be the point and p the line which is not incident with the point A in the qe-plane (Figure 2). Construct the point A_1 which is perpendicular to the point A and incident with the given line p.*

Points A, A_1 are perpendicular if they lie on a pair of corresponding lines a, a_1 of an absolute elliptic involution (F), i.e. if they lie on a pair of perpendicular lines in a Euclidean sense ([7], p.71-75).

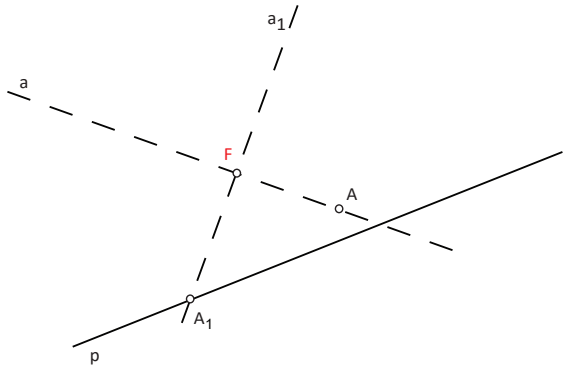


Figure 2

Example 2 *Let the absolute figure $\mathcal{F}_{QE}$ of the qe-plane be given with the involutory pencil (F). Construct the midpoints P_i and the bisectors s_i of a given line segment $\overline{AB}$ $(i = 1,2)$ (Figure 3).*

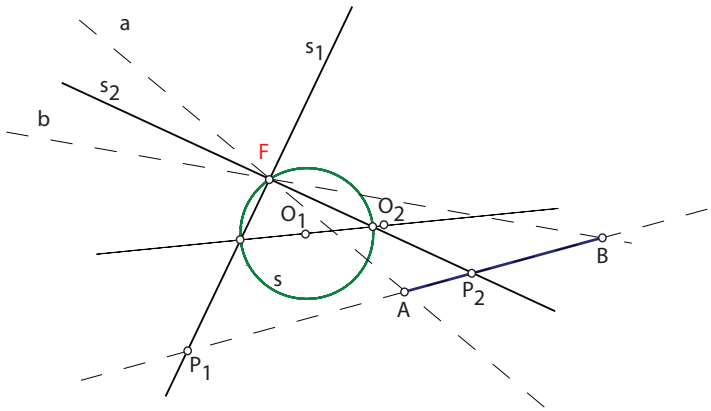


Figure 3

The midpoint of a segment in the qe-plane is dual to an angle bisector in the Euclidean plane, consequently a segment in a qe-plane has two perpendicular midpoints P_1 and P_2 that are in harmonic relation with the points A and B.
A line segment $\overline{AB}$ in the qe-plane has two isotropic bisectors s_1 and s_2 that are a common pair of corresponding lines of two involutions (F) with the center F, denoted as I_1, I_2 . In order to construct the midpoints and bisectors we observe aforementioned involutions (F), a circular involution I_1 is determined by perpendicular corresponding lines in a Euclidean sense and the second hyperbolic involution I_2 is determined by isotropic lines $a = AF$, $b = BF$ as its double lines. The construction is based on the Steiner's construction ([6], p.26, [7], p.74-75). These two pencils will be supplemented by the same Steiner's conic s, which is an arbitrary chosen conic through F. The involutions I_1 and I_2 determine two involutions on the conic s. Let the points O_1 and O_2 be denoted as the centers of these involutions, respectively. The line O_1O_2 intersects the conic s at two points. Isotropic lines s_1 and s_2 through these points are a common pair of these two involutions (F). The intersection points P_1 and P_2 of bisectors s_1 and s_2 with the line AB are midpoints of the line segment $\overline{AB}$.

Example 3 *Let the absolute figure $\mathcal{F}_{QE}$ of the qe-plane be given with the involutory pencil (F). Let two non-isotropic lines a, b be given. Construct an angle bisector between given rays a, b (Figure 4).*

The angle bisector in the qe-plane is dual to a midpoint of a segment in the Euclidean plane. Let V be the vertex of an angle $\angle(a,b)$. Let the isotropic line VF be denoted as f. The angle bisector s is a line in a pencil (V) that is in harmonic relation with triple (a,b,f). The isotropic line f is an isotropic bisector.

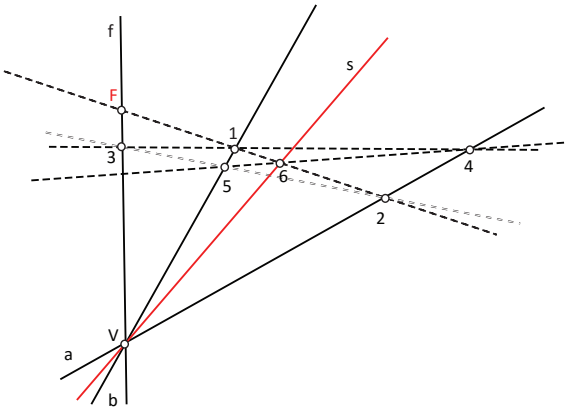


Figure 4

Example 4 *Let the absolute figure $\mathcal{F}_{QE}$ of the qe-plane be given with the involutory pencil (F). Let the lines a, b, c determine a trilateral $\mathcal{ABC}$ with the vertices A, B, C. Construct the ortocentar line of the given trilateral (Figure 5).*

The orthocentar line o of the trilateral in the qe-plane is dual to the orthocenter of a triangle in the Euclidean plane. The points A_1, B_1, C_1 are incident with lines a, b, c and perpendicular to the opposite vertices A, B, C, respectively. The points A_1, B_1, C_1 are collinear and determine a unique ortocentar line.

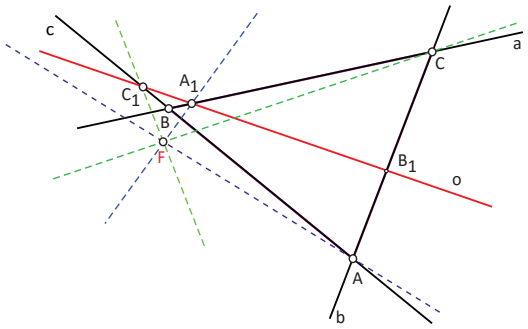


Figure 5

Example 5 *Let the absolute figure $\mathcal{F}_{QE}$ of the qe-plane be given with the involutory pencil (F). Let the lines a, b, c determine a trilateral $\mathcal{ABC}$ with the vertices A, B, C. Construct the centroid line of a trilateral (Figure 6).*

The centroid line o of a trilateral in the qe-plane is dual to the centroid of a triangle in the Euclidean plane. The angel bisectors s_a, s_b, s_c of trilateral intersect opposite sides a, b, c of the trilateral at the points S_A, S_B, S_C, respectively. The points S_A, S_B, S_C are collinear and determine a unique centroid line.

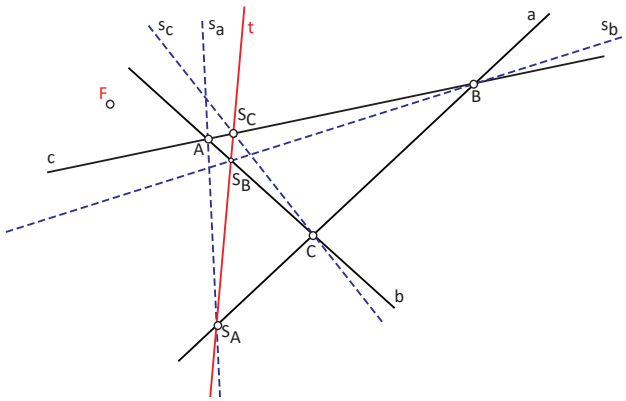


Figure 6

3 Qe-conic classification

There are four types of the second class curves classified according to their position with respect to the absolute figure (Figure 7):

- *qe-hyperbola* (h) - a curve of the second class that has a pair of real and distinct isotropic lines.
  Equilateral qe-hyperbola (h_{EQ}) - a curve of the second class that has isotropic lines as a corresponding lines for the absolute involution (F).
- *qe-ellipse* (e) - a curve of the second class that has a pair of imaginary isotropic lines.
- *qe-parabola* (p) - a curve of the second class where both imaginary isotropic lines coincide.
- *qe-circle* (k) - is a special type of qe-ellipse for which the isotropic lines coincide with the absolute lines j_1 and j_2. In a model of an absolute figure that is used in this paper each qe-conic that has an absolute point F as its Euclidean foci is a qe-circle.

In the projective model of the qe-plane every type of a qe-conic can be represented with every type of Euclidean conics without loss of generality.

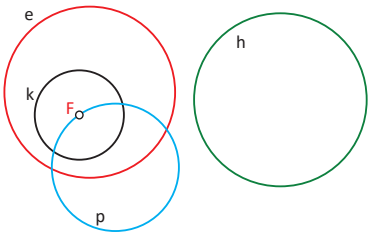


Figure 7

The polar line of the absolute point F with respect to a qe-conic is called the **central line** or the **major diameter** of the second class conic in the qe-plane. The central line of a conic in the qe-plane is dual to a center of a conic in the Euclidean plane. All conics in the qe-plane, except qe-parabolas, have a real non-isotropic central line. The central line of a qe-parabola is isotropic tangent line at the point F.
Dual to the Euclidean diameter of a conic is the point on the central line that is the pole of the isotropic line with respect to a qe-conic. A pair of points incident with the central line that are perpendicular and conjugate with respect to a qe-conic are called the **qe-centers** of the qe-conic. Qe-centers are dual to an axis of Euclidean conic. A qe-ellipse and a qe-hyperbola have two real and distinct qe-centers, while both qe-center of a qe-parabola coincide with the absolute point F.
Each pair of conjugate points incident with the central line with respect to a qe-circle are perpendicular, consequently a qe-circle has infinitely many pairs of qe-centers.
The **isotropic (the minor) diameters** are the lines joining a qe-center to the absolute point F. A qe-ellipse and a qe-hyperbola have two isotropic diameters.
The lines incident with qe-centers of a qe-conic are called the **vertices lines** of a qe-conic in the qe-plane. A qe-hyperbola has two real vertices lines, while a qe-ellipse has four real vertices lines.
A hyperosculating qe-circle of a qe-conic can be constructed only at the vertices lines of a qe-conic.
The intersection points of a qe-conic and vertices lines are called **co-vertices points**.

4 Some construction assignments

Exercise 1 *Construct a qe-circle k determined with the given central line c and the line p (Figure 8).*

In order to construct the qe-circle as a line envelope, a perspective collineation that maps arbitrary chosen qe-circle k_1 into qe-circle k is used. The construction is carried out in the following steps:
The absolute point F is selected for the center of the collineation. Let k_1 be an arbitrary chosen qe-circle with the center F. A polar line c_1 of F, is the central line for chosen qe-circle k_1. Notice that c_1 is the line at infinity. The lines c and c_1 are corresponding lines for the perspective collineation with the center F. Let the point S be the intersection point of the lines p and c. To determine an axis o of the perspective collineation, the point R that is perpendicular to the point S and incident with the line p is constructed. A ray FR of the collineation intersect the qe-circle k_1 at the points R_1 and R_2. Let the line p_1 touch the qe-circle k_1 at a point R_1. The lines p and p_1 are corresponding lines for the perspective collineation with a center F. The axis o passes through the intersection point S_1 of the lines p_1 and p, and it is parallel to c.

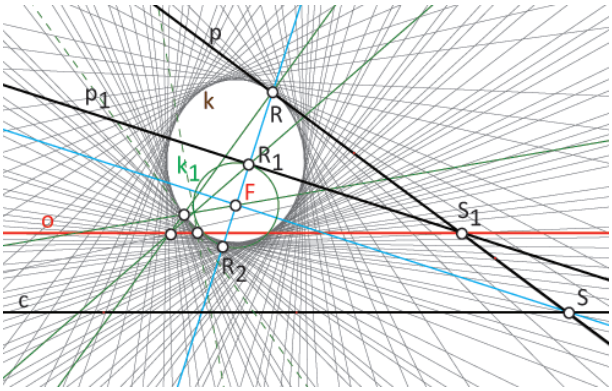


Figure 8

Exercise 2 *Construct the hyperosculating qe-circle of a qe-hyperbola h_1 (Figure 9).*

Let the qe-hyperbola h_1 be given and its central line be denoted as c. A hyperosculating qe-circle of the qe-hyperbola h_1 can be constructed only at the vertices lines. A qe-hyperbola h_1 has two real vertices lines t_1 and t_2. Let the points T_1 and T_2 be co-vertices points. Let the line t_2 and the point T_2 be observed. In order to construct a hyperosculating qe-circle, the point S_2 that is perpendicular to T_2, and incident with the line t_2 is constructed. The central line c_h of a hyperosculating qe-circle is incident with S_2. In order to construct c_h, let the line y_1 of the qe-hyperbola h_1 be arbitrary chosen. The intersection point of h_1 and the line y_1 is denoted as Y_1. The intersection point of lines t_2 and y_1 is denoted as K. The point K_1 is perpendicular to K and incident with joining line T_2Y_1. The line S_2K_1, denoted as c_h, is a central line of a hyperosculating qe-circle. The central line c_h and the line t_2 determine a hyperosculating qe-circle and to construct it the same principle as in Exercise 1 is used.

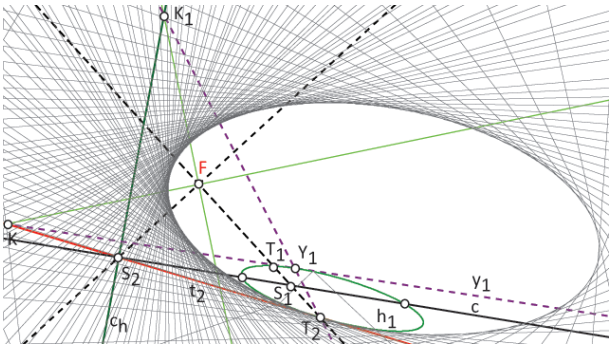


Figure 9

Theorem 1 *Let the lines $\{a,b,c,d\}$ be the base of a pencil of qe-conics in a qe-plane (Figure 10). Then, the envelope of the central lines of all qe-conics in the pencil is a curve of the second class.*

Proof: It is known that the envelope of polar lines of conics in a pencil of conics with respect to a common pole P is a curve of the second class ([6]). Consequently, in the qe-plane, if a common pole P coincides with the absolute point F, than the envelope of its polar lines coincides with the envelope of the central lines in the given pencil. □

In order to construct the envelope of the central lines of all qe-conics in the pencil of qe-conics, denoted as δ_1, we observe involutory pencil (F) of a pairs of isotropic lines of all qe-conics in a pencil. Each qe-conic in a pencil of qe-conics has two real or imaginary isotropic lines. In the given pencil of qe-conics there are three qe-conics degenerated into three pairs of points, denoted as $(1,1')$, $(2,2')$, $(3,3')$. Let the involution (F) be determined with an isotropic lines of any two degenerated qe-conics in a pencil i.e. $(1,1')$, $(3,3')$. The pencil of qe-conics contains two, one or none real qe-parabola.

From the viewpoint of qe-geometry, the envelope δ_1 is a qe-hyperbola if the pencil contains two qe-parabolas. The central lines of these qe-parabolas denoted as, p_1 and p_2 are double lines for the involution (F) and they coincide with the isotropic lines of the envelope δ_1. The envelope is determinate with five lines; the lines p_1, p_2, and central lines of three degenerated qe-conics c_1, c_2, c_3.

The envelope δ_1 is a qe-parabola if the pencil contain one qe-parabola.

The envelope δ_1 is an qe-ellipse if the pencil does not contain qe-parabolas (Figure 10). Double lines for the elliptic involution (F) are imaginary lines.

Pencil will be supplemented by the Steiner's conic s, which is an arbitrary chosen conic through F. Let the point O be denoted as a center of the involution (F).

If the point O is outside the conic s, involutory pencil (F) contains real double lines, and the envelope δ_1 is a qe-hyperbola. If the point O is on the conic s, double lines of involution (F) coincide, and the envelope δ_1 is a qe-parabola.

If the point O is inside the conic s, involutory pencil (F) contains imaginary double lines, and the envelope δ_1 is an qe-ellipse (Figure 10).

If the point O coincides with the center of conic s, double lines of involution (F) coincides with the absolute lines j_1 and j_2, and the envelope δ_1 is a qe-circle (circular involution).

If one of the base lines in a pencil is isotropic line, the pencil of qe-conics contains qe-hyperbolas and one qe-parabola, the envelope δ_1 is a qe-parabola.

If two of the base lines in a pencil are isotropic lines, the envelope δ_1 degenerates into a point.

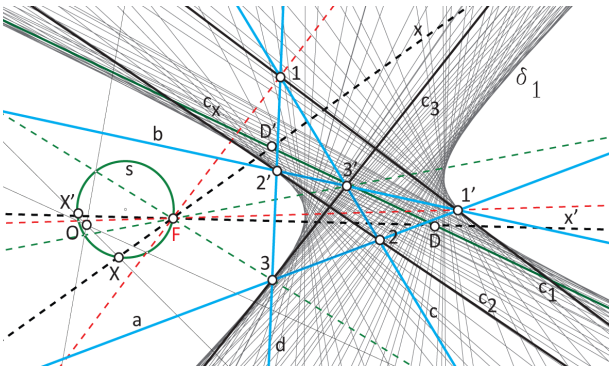


Figure10: *Qe-ellipse - an envelope of the central lines*

Corollary 1 *Let any two degenerated qe-conics in a pencil of qe-conics be given as a pair of perpendicular points i.e. the pencil of equilateral qe-hyperbolas. Than the envelope of the central line is a qe-circle.*

References

[1] H.S.M. COXETER, *Introduction to geometry*, John Wiley & Sons, Inc, Toronto, 1969;

[2] N. KOVAČEVIĆ, E. JURKIN, Circular cubics and quartics in pseudo-Euclidean plane obtained by inversion, *Math. Pannon.* **22**/1 (2011), 1-20;

[3] N. KOVAČEVIĆ, V. SZIROVICZA, Inversion in Minkowskischer geometrie, *Math. Pannon.* **21**/1 (2010), 89-113;

[4] N.M. MAKAROVA, On the projective metrics in plane, *Učenye zap. Mos. Gos. Ped. in-ta*, **243** (1965), 274-290. (Russian);

[5] M. D. MILOJEVIĆ, Certain comparative examinations of plane geometries according to Cayley-Klein, *Novi Sad J.Math.*, **29**/3, (1999), 159-167

[6] V. NIČE, *Uvod u sintetičku geometriju*, Školska knjiga, Zagreb, 1956.;

[7] D. PALMAN, *Projektivne konstrukcije*, Element, Zagreb, 2005;

[8] A. SLIEPČEVIĆ, I. BOŽIĆ, Classification of perspective collineations and application to a conic, *KoG* **15**, (2011), 63-66;

[9] A. SLIEPČEVIĆ, M. KATIĆ ŽLEPALO, Pedal curves of conics in pseudo-Euclidean plane, *Math. Pannon.* **23**/1 (2012), 75-84;

[10] A. SLIEPČEVIĆ, N. KOVAČEVIĆ, Hyperosculating circles of conics in the pseudo-Eucliden plane, *Manuscript*;

[11] D.M.Y SOMMERVILLE, Classification of geometries with projective metric, *Proc. Ediburgh Math. Soc.* **28** (1910), 25-41;

[12] I.M. YAGLOM, B.A. ROZENFELD, E.U. YASINSKAYA, Projective metrics, *Russ. Math Surreys*, **Vol. 19**/5, No. 5, (1964), 51-113;

[13] G. WEISS, A. SLIEPČEVIĆ, Osculating circles of conics in Cayley-Klein planes, *KoG* **13**, (2009), 7-13;

[14] A. SLIEPČEVIĆ, I. BOŽIĆ, H. HALAS, Introduction to the planimetry of the quasi-hyperbolic plane, *KoG* **17**, (2013), 58-64;

[15] M. KATIĆ ŽLEPALO, Curves of centers of conic pencils in pseudo-Euclidean plane, *Proceedings of the 16th International Conference on Geometry and Graphics*, (2014);

[16] A. SLIEPČEVIĆ, I. BOŽIĆ, The analogue of theorems related to Wallace-Simson's line in quasi-hyperbolic plane, *Proceedings of the 16th International Conference on Geometry and Graphics*, (2014);

Ivana Božić Dragun
e-mail: ivana.bozic@tvz.hr
University of Applied Sciences Zagreb,
Avenija V. Holjevca 15, 10 000 Zagreb, Croatia

Ana Sliepčević
email: anasliepcevic@gmail.com
Faculty of Civil Engineering, University of Zagreb,
Kačićeva 26, 10 000 Zagreb, Croatia

Original scientific paper
Accepted 9. 11. 2016.

NGUYEN LE
N J WILDBERGER

Incenter Symmetry, Euler lines, and Schiffler Points

Incenter Symmetry, Euler Lines, and Schiffler Points

ABSTRACT

We look at the four-fold symmetry given by the Incenter quadrangle of a triangle, and the relation with the cirumcircle, which in this case is the nine-point conic of the quadrangle. By investigating Euler lines of Incenter triangles, we show that the classical Schiffler point extends to a set of four Schiffler points, all of which lie on the Euler line. We discover also an additional quadrangle of Incenter Euler points on the circumcircle and investigate its interesting diagonal triangle. The results are framed in purely algebraic terms, so hold over a general bilinear form. We present also a mysterious case of apparent symmetry breaking in the Incenter quadrangle.

Key words: triangle geometry, Euclidean geometry, rational trigonometry, bilinear form, Schiffler points, Euler lines, Incenter hierarchy, circumcircles

MSC2010: 51M05, 51M10, 51N10

"Upisana simetrija", Eulerovi pravci i Schifflerove točke

SAŽETAK

Proučavamo četverostruku simetriju određenu četverovrhom, čiji su vrhovi središta upisanih (pripisanih) kružnica danog trokuta, te vezu s opisanom kružnicom trokuta koja je u ovom slučaju konika devet točaka spomenutog četverokuta. Proučavajući Eulerove pravce takozvanih upisanih trokuta, pokazujemo da je poopćenje klasične Schifflerove točke skup od četiriju točaka koje leže na Eulerovom pravcu. Promatra se četverokut u čijim se vrhovima sijeku Eulerovi pravci upisanih trokuta, te njegov dijagonalni trokut. Kako se koristi algebarski pristup, dobiveni rezultati vrijede za opću bilinearnu formu. Dajemo i primjer svojevrsnog nestanka četverostruke simetrije.

Ključne riječi: geometrija trokuta, euklidska geometrija, racionalna trigonometrija, bilinearna forma, Schifflerove točke, Eulerovi pravci, hijerarhija središta upisanih kružnica, opisane kružnice

1 Introduction

The following is a classical theorem which was first observed by M. Bôcher in 1892. Special cases include the nine-point circle of a triangle, and the nine point hyperbola.

Theorem 1 (Nine point conic) *The six midpoints of a quadrangle (four points) together with the diagonal points lie on a conic.*

This is called the *Nine point conic* of the quadrangle. Bôcher observed that if one of the four points lies on the circumcircle defined by the other three, then the conic is an equilateral hyperbola. If one of the points is the orthocenter of the other three, then the conic is a circle. In Figure 1 we see a general quadrangle $\overline{P_0P_1P_2P_3}$, as well as the six midpoints in dark blue, and the three diagonal points in orange, with these last nine points on the red conic.

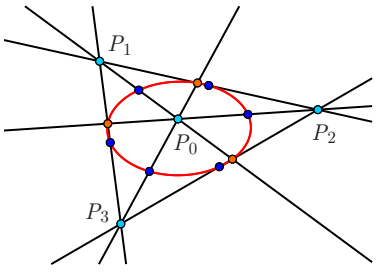


Figure 1: *The Nine point conic of the quadrangle* $\overline{P_1P_2P_3P_4}$

If we consider the above theorem in relation to the Incenter quadrangle $\overline{I_0I_1I_2I_3}$ of a Triangle $\overline{A_1A_2A_3}$, some additional interesting things happen, since this is an orthocentric quadrangle.

Theorem 2 *The Nine point conic of the Incenter quadrangle $\overline{I_0I_1I_2I_3}$ of a Triangle $\overline{A_1A_2A_3}$ is the Circumcircle c of that triangle, and so also the nine-point circle of any three Incenters. Each midpoint of the quadrangle $\overline{I_0I_1I_2I_3}$ is the center of a circle which passes through two Incenters as well as two Points of the Triangle.*

This last lovely fact finds its way routinely into International problem competitions, as has been compiled by E. Chen, who calls it the *Incenter/Excenter lemma* (see [1]). He gives a proof using angle chasing, we will give a more powerful and general argument in the course of this paper. In Figure 2 we see that the **Incenter midpoint** $M = M_{02}$, which is the midpoint of the segment $\overline{I_0I_2}$, is the center of a circle which passes through two points of the triangle, in this case A_1 and A_3, as well as the two Incenters I_0 and I_2.

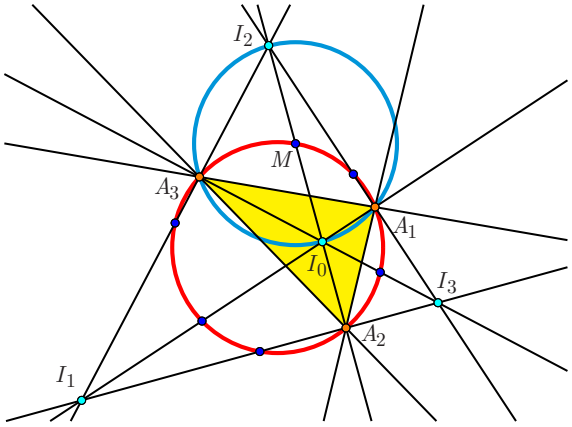


Figure 2: *The Incenter quadrangle and its midpoints*

It is worth noting that obviously each Incenter midpoint lies on an angle bisector, or **Biline**, of the Triangle $\overline{A_1A_2A_3}$, as these are the six lines of the complete quadrangle $\overline{I_0I_1I_2I_3}$.

In C. Kimberling's celebrated list of triangle centers, see [3] and [4], the Incenter I_0 gets pride of place, as the first point X_1 in the entire list. Because his list contains only uniquely defined centers, the other Incenters I_1, I_2 and I_3, which are more usually called excenters, do not get explicit numbered names. In this paper we investigate the fourfold symmetry surrounding Incenter midpoints within the set-up of Rational Trigonometry ([11], [12]), valid for any symmetric bilinear form, as described in [7]. So the theorems in this paper hold also with other bilinear forms, as in Lorentzian planar geometry.

Next to the Incenter, the most famous triangle centers are the Centroid $G = X_2$, the Circumcenter $C = X_3$, and the Orthocenter $H = X_4$, which famously all lie on the Euler line e. In Figure 3 we see both the Euler line and the Incenter quadrangle $\overline{I_0I_1I_2I_3}$ for the Euclidean example that we will exhibit frequently.

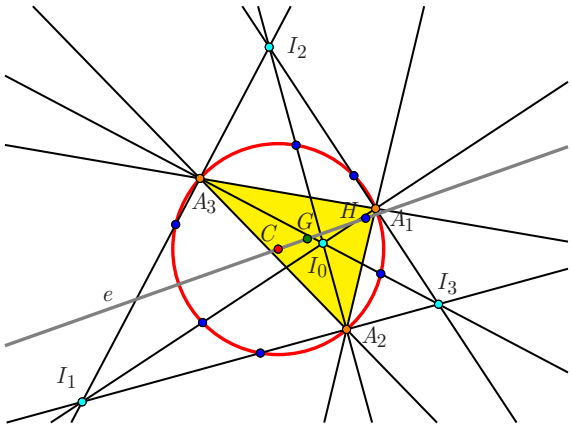


Figure 3: *The Euler line and Incenter quadrangle of $\overline{A_1A_2A_3}$*

The **Schiffler point** $S = X_{21}$ of the triangle $\overline{A_1A_2A_3}$ is another remarkable triangle centre which was discovered more recently by Kurt Schiffler (1896-1986) [9]. This point is the intersection of the Euler lines of the three **Incenter triangles** $\overline{A_1A_2I_0}$, $\overline{A_1A_3I_0}$, $\overline{A_2A_3I_0}$. Pleasantly S lies on the Euler line e of the original triangle $\overline{A_1A_2A_3}$.

This situation is illustrated in Figure 4 which shows the Schiffler point S (in white) of $\overline{A_1A_2A_3}$, the meet of the four **Incenter Euler** lines (in gray), passing through circumcenters (blue) and centroids (green) of the Incenter triangles. Clearly these circumcenters are exactly the midpoints that we observed in the previous diagram. There are several interesting and remarkable properties of the Schiffler point which have been found over the years: see for example ([2], [8], [10]).

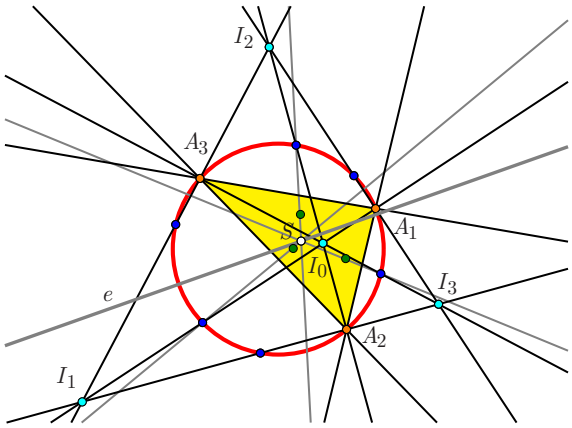


Figure 4: *The Schiffler point S of $\overline{A_1A_2A_3}$*

Since our philosophy, expounded in [7] and [6], is that we ought to consider all four Incenters symmetrically, it is natural for us to expand this story to include Incenter Euler lines from the *other* Incenter triangles obtained by combining two vertices of the original triangle $\overline{A_1A_2A_3}$ and any

one of the Incenters. If we agree that $\{i,j,k\} = \{1,2,3\}$, then such an Incenter triangle $\overline{A_jA_kI_l}$ is determined by the pair of indices (i,l), where i runs through $1,2,3$ and l runs through $0,1,2,3$. So let us denote by e_{il} the Incenter Euler line of the triangle $\overline{A_jA_kI_l}$. *Notice that the Point label comes first, followed by the Incenter label.*
This way we get twelve Incenter Euler lines, not just three. When we look at all of these, we meet some remarkable new phenomenon. The first observation is that the standard Schiffler point $S = S_0$ is now but one of four Schiffler points.

Theorem 3 (Four Schiffler points) *The triples* $S_0 \equiv e_{10}e_{20}e_{30}$, $S_1 \equiv e_{11}e_{21}e_{31}$, $S_2 \equiv e_{12}e_{22}e_{32}$ *and* $S_3 \equiv e_{13}e_{23}e_{33}$ *of Incenter Euler lines are concurrent. These Schiffler points all lie on the Euler line e of the original triangle* $\overline{A_1A_2A_3}$.

The next result shows that there are other interesting concurrences of the Incenter Euler lines. These are also visible in Figure 5.

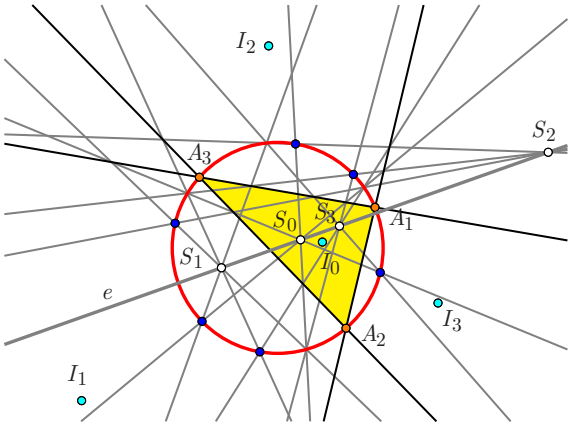


Figure 5: *The four Schiffler points* S_0, S_1, S_2 *and* S_3 *on the Euler line e*

Theorem 4 (Four Incenter Euler points) *The triples* $P_0 \equiv e_{11}e_{22}e_{33}$, $P_1 \equiv e_{10}e_{23}e_{32}$, $P_2 \equiv e_{20}e_{13}e_{31}$ *and* $P_3 \equiv e_{30}e_{12}e_{21}$ *of Euler lines are concurrent. These points all lie on the Circumcircle of the original Triangle.*

These theorems will form the starting points of the investigations of this paper. We will see that the diagonal triangle $\overline{D_1D_2D_3}$ of the quadrangle $\overline{P_0P_1P_2P_3}$ has some remarkable connections with the original triangle $\overline{A_1A_2A_3}$. We call $\overline{D_1D_2D_3}$ the **Diagonal Incenter Euler triangle**. At the end of the paper, we note a remarkable appearance of symmetry breaking in the original Incenter quadrangle $\overline{I_0I_1I_2I_3}$ which is well worth further investigation.
Throughout the paper our emphasis is on explicit formulas that allow us to give general algebraic proofs. We will give diagrams that illustrate the Euclidean case, but it is an essential strength of this approach that the results hold for a general bilinear form, and we will also include a few pictures from the green geometry coming from Chromogeometry (see [13] and [14]).

1.1 Quadrance and spread

In this section we briefly summarize the main facts needed from rational trigonometry in the general affine setting (see [11], [12]). We work in the standard two-dimensional affine or vector space over a field, consisting of affine **points**, or row vectors $v = [x,y]$. Sometimes it will be convenient to represent such a vector projectively, as the projective row vector $[x : y : 1]$; this makes dealing with fractional entries easier. A **line** l is the proportion $l \equiv \langle p : q : r \rangle$, or equivalently a projective column vector $[p : q : r]^T$, provided that p and q are not both zero. Incidence between the point v and the line l above is given by the relation

$$px + qy + r = 0.$$

Our notation is that the line determined by two points A and B is denoted AB, while the point where two non-parallel lines l and m meet is denoted lm. If three lines k,l and m are concurrent at a point A, we will sometimes write $A = klm$.
A metrical structure is determined by a non-degenerate symmetric 2×2 matrix D: this gives a symmetric bilinear form on vectors

$$v \cdot u \equiv vDu^T.$$

Non-degenerate means $\det D \neq 0$, and implies that if $v \cdot u = 0$ for all vectors u, then $v = 0$.
Two vectors v and u are then **perpendicular** precisely when $v \cdot u = 0$. Since the matrix D is non-degenerate, for any vector v there is, up to a scalar, exactly one vector u which is perpendicular to v. Two lines l and m are **perpendicular** precisely when they have perpendicular direction vectors.
The bilinear form determines the **quadrance** of a vector v as

$$Q(v) \equiv v \cdot v$$

and similarly the **quadrance** between points A and B is

$$Q(A,B) \equiv Q\left(\overrightarrow{AB}\right).$$

A vector v is **null** precisely when $Q(v) = v \cdot v = 0$, in other words precisely when v is perpendicular to itself. A line is **null** precisely when it has a null direction vector.

The **spread** between non-null vectors v and u is the number

$$s(v,u) \equiv 1 - \frac{(v \cdot u)^2}{Q(v)\,Q(u)} = 1 - \frac{(v \cdot u)^2}{(v \cdot v)\,(u \cdot u)}$$

and the **spread** between any non-null lines l and m with direction vectors v and u is defined to be $s(l,m) \equiv s(v,u)$.

1.2 Standard coordinates

This paper employs the novel approach to planar affine triangle geometry initiated in [7] and continued in [6], which allows us to frame the subject in a much wider and more general algebraic fashion, valid over an arbitrary field, not of characteristic two.

The basic idea with standard coordinates is to take any particular triangle, and apply a combination of a translation and an invertible linear transformation to send it to the **standard Triangle** $\overline{A_1A_2A_3}$ with

$$A_1 \equiv [0,0], \quad A_2 \equiv [1,0] \quad \text{and} \quad A_3 \equiv [0,1]. \tag{1}$$

Our convention is to use capital letters to refer to objects associated to this standard Triangle. The Lines of the Triangle are

$$l_1 \equiv A_2A_3 = \langle 1:1:-1 \rangle,$$
$$l_2 \equiv A_1A_3 = \langle 1:0:0 \rangle,$$
$$l_3 \equiv A_2A_1 = \langle 0:1:0 \rangle.$$

The **Midpoints** of the Triangle are clearly

$$M_1 = \left[\frac{1}{2}, \frac{1}{2}\right], \quad M_2 = \left[0, \frac{1}{2}\right], \quad M_3 = \left[\frac{1}{2}, 0\right]$$

while the corresponding **Median lines** are

$$d_1 \equiv A_1M_1 = \langle 1:-1:0 \rangle,$$
$$d_2 \equiv A_2M_2 = \langle 1:2:-1 \rangle,$$
$$d_3 \equiv A_3M_3 = \langle 2:1:-1 \rangle.$$

The **Centroid** is the common meet of the Medians, namely

$$G = X_2 = \left[\frac{1}{3}, \frac{1}{3}\right].$$

These objects are defined independent of any metrical structure: they are purely affine notions.

A metrical structure may be imposed by a general invertible 2×2 matrix

$$D \equiv \begin{pmatrix} a & b \\ b & c \end{pmatrix}. \tag{2}$$

We note that the determinant of D is $ac - b^2$. The quantity

$$d \equiv a + c - 2b$$

will also prove to be useful.

Because the effect of a linear transformation on a bilinear form is the familiar congruence, it suffices to understand the particular standard Triangle with respect to such a general quadratic form. This is the basic, but powerful, idea behind standard coordinates. The idea now is to find all relevant information about the original triangle in terms of the corresponding information about the standard Triangle expressed in terms of the numbers a, b and c.

So we have moved from considering *a general triangle with respect to a specific bilinear form* to the more general situation of a *specific triangle with respect to a general quadratic form*. This system of standard coordinates allows a systematic augmentation of Kimberling's *Encyclopedia of Triangle Centers* ([3], [4], [5]) to *arbitrary quadratic forms* and *general fields*.

The **Midlines** m_1, m_2 and m_3 of the Triangle are the lines through the midpoints M_1, M_2 and M_3 perpendicular to the respective sides— these are usually called **perpendicular bisectors**. They are also the altitudes of $\overline{M_1M_2M_3}$ and are given by:

$$m_1 = \langle 2(b-a) : 2(c-b) : a-c \rangle,$$
$$m_2 = \langle 2b : 2c : -c \rangle,$$
$$m_3 = \langle 2a : 2b : -a \rangle.$$

The Midlines m_1, m_2, m_3 meet at the **Circumcenter**

$$C = X_3 = \frac{1}{2(ac-b^2)}\,[c(a-b), a(c-b)]. \tag{3}$$

The **Circumcircle** c of $\overline{A_1A_2A_3}$ is the unique circle with equation $Q(X,C) = R$ that passes through A_1, A_2 and A_3, and this turns out to be the equation in $X = [x,y]$ given by

$$ax^2 + 2bxy + cy^2 - ax - cy = 0. \tag{4}$$

The **Orthocenter** of the Triangle is

$$H = X_4 = \frac{b}{ac-b^2}\,[c-b, a-b].$$

The **Euler line** CG is

$$e = \left\langle 2b^2 - 3ab + ac : -2b^2 + 3cb - ac : b(a-c) \right\rangle. \tag{5}$$

The fact that this line passes through each of C, G and H can be checked by making the following computations via projective coordinates:

$$\left[c(a-b) : a(c-b) : 2\left(ac-b^2\right)\right]$$
$$\left[2b^2 - 3ab + ac : -2b^2 + 3cb - ac : b(a-c)\right]^T = 0,$$
$$[1:1:3]\left[2b^2 - 3ab + ac : -2b^2 + 3cb - ac : b(a-c)\right]^T = 0,$$
$$\left[b(c-b) : b(a-b) : ac-b^2\right]$$
$$\left[2b^2 - 3ab + ac : -2b^2 + 3cb - ac : b(a-c)\right]^T = 0.$$

The existence of Incenters of our standard Triangle however is more subtle: this leads to number theoretic conditions that depend on certain quantities being squares in our field.

2 The four Incenters

A **biline** of the non-null vertex $\overline{l_1 l_2}$ is a line b which passes through $l_1 l_2$ and satisfies $s(l_1,b) = s(b,l_2)$. The existence of Bilines (and hence Incenters) of the standard Triangle depends on number theoretical considerations of a particularly simple kind which we recall from [7].

Theorem 5 (Existence of Triangle bilines) *The Triangle $\overline{A_1A_2A_3}$ has Bilines at each vertex precisely when we can find numbers u,v,w in the field satisfying*

$$ac = u^2, \quad ad = v^2, \quad cd = w^2. \tag{6}$$

In this case we can choose u,v,w so that $acd = uvw$ and

$$du = vw \quad cv = uw \quad \text{and} \quad aw = uv. \tag{7}$$

We are interested in formulas for triangle centers of the standard Triangle $\overline{A_1A_2A_3}$, assuming the existence of Bilines. These formulas will then involve the entries a,b and c of D from (2), as well as the secondary quantities u,v and w. The **quadratic relations** (6 and 7) play a major role in simplifying formulas.
The four Incenters are, from [7],

$$I_0 = \frac{1}{d+v-w}[-w,v], \quad I_1 = \frac{1}{d-v+w}[w,-v],$$
$$I_2 = \frac{1}{d+v+w}[w,v], \quad I_3 = \frac{1}{d-v-w}[-w,-v].$$

It is important to note that I_1, I_2 and I_3 may be obtained from I_0 by changing signs of: both v and w, just w, and just v respectively. This four-fold symmetry will hold more generally and it means that we can generally just record the formulas for objects which are associated to I_0. We refer to this as the **basic u,v,w symmetry**.

2.1 Incenter midpoints

We now look at meets of Midlines and the Circumcircle. Somewhat surprisingly, it turns out that the existence of these meets is entirely aligned with the existence of Incenters.

Theorem 6 (Incenter midpoints) *The three Midlines m_1, m_2 and m_3 meet the Circumcircle c precisely when Incenters exist, that is when we can find u,v and w satisfying the quadratic relations. In this case, the Midline m_1 meets the Circumcircle in points*

$$M_{01} \equiv \frac{1}{2(b-u)}[c-u, a-u], \quad M_{23} \equiv \frac{1}{2(b+u)}[c+u, a+u]$$

which are the midpoints of $\overline{I_0I_1}$ and $\overline{I_2I_3}$ respectively; the Midline m_2 meets the Circumcircle in points

$$M_{13} \equiv \frac{1}{2(b-a+v)}[c, v-a], \quad M_{02} \equiv \frac{1}{2(a-b+v)}[-c, v+a]$$

which are the midpoints of $\overline{I_1I_3}$ and $\overline{I_0I_2}$ respectively; and the Midline m_3 meets the Circumcircle in points

$$M_{03} \equiv \frac{1}{2(b-c+w)}[w-c, a], \quad M_{12} \equiv \frac{1}{2(c-b+w)}[w+c, -a]$$

which are the midpoints of $\overline{I_0I_3}$ and $\overline{I_1I_2}$ respectively.

Proof. The proofs of these are straightforward, as we have the equations of the Midlines and the Circumcircle c, and finding midpoints of a segment just involves taking the averages of the coordinates. However we must be prepared to use the quadratic relations to make simplifications. □

This theorem motivates us to call the points M_{ij} the **Incenter midpoints** of the Triangle.

2.2 Incenter Euler lines

For each Incenter triangle $\overline{A_jA_kI_l}$ we may now compute its Euler line, which we call an **Incenter Euler** line of the original triangle $\overline{A_1A_2A_3}$. This may be done by joining the circumcenter of the Incenter triangle, which is an Incenter midpoint, to the centroid of that Incenter triangle, whose coordinates are just formed by taking affine averages of the points of the given triangle.
For example the Euler line e_{30} of $\overline{A_1A_2I_0}$ is the join of the Incenter midpoint

$$M_{03} = \frac{1}{2(b-c+w)}[w-c, a] = [w-c : a : 2(b-c+w)]$$

and the centroid

$$\frac{1}{3}\left[\frac{d+v-2w}{d+v-w}, \frac{v}{d+v-w}\right] = [d+v-2w : v : 3(d+v-w)].$$

Using a Euclidean cross product and simplifying using the quadratic relations, we find that

$$e_{30} = \left\langle \begin{matrix} 6ab-3ac+2au-3av-4bu+3aw+2bv+2cu-2cv-3a^2 : \\ au-2bc-2ab-2bu+aw-2bv+cu+2bw-cv+4b^2 : \\ ac-2ab-au+av+2bu-2aw-cu+cv+a^2 \end{matrix} \right\rangle.$$

Note that we can obtain e_{i1}, e_{i2}, e_{i3} from e_{i0} by changing the signs of (v,w), (u,w) and (u,v) respectively. So for example by applying the basic u,v,w symmetry we find that

$$e_{31} = \left\langle \begin{matrix} 6ab-3ac+2au+3av-4bu-3aw-2bv+2cu+2cv-3a^2 : \\ au-2bc-2ab-2bu-aw+2bv+cu-2bw+cv+4b^2 : \\ ac-2ab-au-av+2bu+2aw-cu-cv+a^2 \end{matrix} \right\rangle.$$

So it suffices if we exhibit also

$$e_{20} = \left\langle \begin{array}{c} au-2bc-2ab-2bu+aw-2bv+cu+2bw-cv+4b^2 : \\ 6bc-3ac+2au-4bu+2aw+2cu-2bw-3cv+3cw-3c^2 : \\ ac-2bc-au+2bu-aw-cu+2cv-cw+c^2 \end{array} \right\rangle$$

and

$$e_{10} = \left\langle \begin{array}{c} 8ab-3ac+2bc+au-3av-2bu+2aw \\ +4bv+cu-2bw-cv-3a^2-4b^2 : \\ 3ac-2ab-8bc-au+2bu-aw-2bv \\ -cu+4bw+2cv-3cw+4b^2+3c^2 : \\ (a-c)(a-2b+c+v-w) \end{array} \right\rangle.$$

3 Schiffler points

The Incenter Euler lines also figure prominently in the classical Schiffler point. We will now see that there is in fact a four-fold symmetry inherent here.

Theorem 7 (Four Schiffler points) *The triples* $S_0 \equiv e_{10}e_{20}e_{30}$, $S_1 \equiv e_{11}e_{21}e_{31}$, $S_2 \equiv e_{12}e_{22}e_{32}$ *and* $S_3 \equiv e_{13}e_{23}e_{33}$ *of Incenter Euler lines are concurrent. These points all lie on the Euler line.*

Proof. The concurrences of the lines e_{10}, e_{20}, e_{30} is

$$S_0 = \left[\begin{array}{c} \left(2a^2-5ab+6ac+2b^2-7bc+2c^2\right)u \\ -c(5a-5b+2c)+\left(5ac-3ab-2bc+2a^2\right)w \\ -c\left(-10ab+5ac-2bc+5a^2+2b^2\right): \\ \left(2a^2-7ab+6ac+2b^2-5bc+2c^2\right)u \\ +\left(2ab-5ac+3bc-2c^2\right)v+a(5c-5b+2a)w \\ +a\left(2ab-5ac-2b^2+10bc-5c^2\right): \\ \left(6a^2-15ab+16ac+2b^2-15bc+6c^2\right)u \\ +\left(4b^2+9bc-6c^2-13ac\right)v \\ +\left(6a^2-9ab+13ca-4b^2\right)w \\ +\left(4ab^2-13a^2c+22abc-13ac^2-4b^3+4b^2c\right) \end{array} \right].$$

The other three Schiffler points S_1, S_2 and S_3 may be computed to be exactly the corresponding points when we perform the three basic u,v,w symmetries, namely negating v and w to get S_1, negating u and w to get S_2, and negating u and v to get S_3. □

The Euler line e we know is (5), so we can check directly that $eS_0 = 0$ identically, without use of the quadratic relations. The statement also holds for the other Schiffler points.

In Figure 6 we see an example from green geometry with the bilinear form $x_1y_2 + x_2y_1$, showing the four Schiffler points of the triangle $\overline{A_1A_2A_3}$ on the green Euler line e (for more about chromogeometry and geometry in Lorentz spaces see for example [13], [14]).

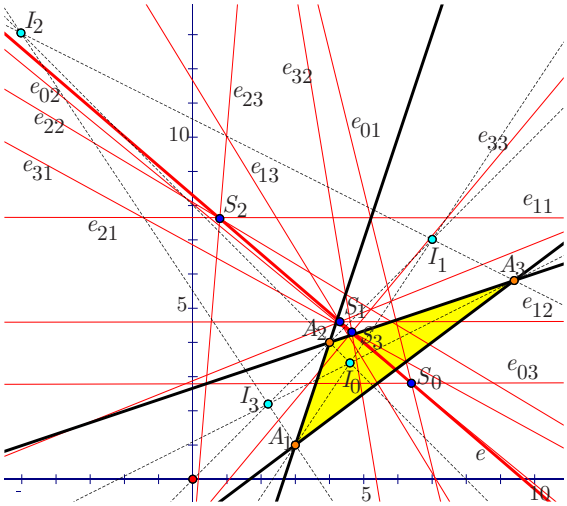


Figure 6: *Green Schiffler points lying on the green Euler line of* $\overline{A_1A_2A_3}$

4 Incenter Euler points

Theorem 8 (Four Incenter Euler points) *The triples* $P_0 \equiv e_{11}e_{22}e_{33}$, $P_1 \equiv e_{10}e_{23}e_{32}$, $P_2 \equiv e_{20}e_{13}e_{31}$, *and* $P_3 \equiv e_{30}e_{12}e_{21}$ *of Euler lines are concurrent. These points all lie on the Circumcircle c of the original triangle.*

Proof. The proof requires using the quadratic relations involving u, v and w. For example to show the concurrency $P_0 \equiv e_{11}e_{22}e_{33}$ we create the determinant of the 3×3 matrix with rows given by the Euler lines. This expression is a polynomial of degree six in a,b,c and u,v and w. By successive applications of the quadratic relations involving u,v and w we can step by step reduce this polynomial until it eventually equals 0. Alternatively we can use the cross product to determine the common meets of these lines: here is the formula for P_0:

$$P_0 = \left[\begin{array}{c} \left(2b^2-5bc-ab+2c^2+2ac\right)u+\left(ac-3bc+2c^2\right)v \\ +(3ac-ab-2bc)w+c\left(a^2-4ab+3ac+2b^2-2bc\right): \\ b(2b-c-a)u+c(a-b)v+a(c-b)w \\ +a\left(2b^2-ac-2bc+c^2\right): \\ b(2b-c-a)u+c(a-b)v+\left(5ac-ab-4b^2\right)w \\ +\left(a^2c-6abc+5ac^2+4b^3-4b^2c\right) \end{array} \right].$$

The formulas for P_1, P_2 and P_3 follow by the basic u,v,w symmetry. The Circumcircle c of the standard Triangle we know has equation $ax^2+2bxy+cy^2-ax-cy=0$. By substitution, we find, after using the quadratic relations, that P_0 satisfies this equation, and the other points are similar. □

We will call the points P_0, P_1, P_2 and P_3 the **Incenter Euler points** of the triangle.

4.1 Lines of the Incenter Euler quadrangle $\overline{P_0P_1P_2P_3}$

The lines of the Incenter Euler quadrangle have the following equations:

$$P_0P_1 = \left\langle \begin{array}{c} c\left(ab-a^2-2ac+2b^2\right)v+a\left(ab-3ca+2b^2\right)w: \\ c\left(2b^2+bc-3ac\right)v+a\left(2b^2-2ac+bc-c^2\right)w: \\ c\left(a^2+3ac-3ba-bc\right)v+a\left(3ac-ba+c^2-3bc\right)w \end{array} \right\rangle,$$

$$P_2P_3 = \left\langle \begin{array}{c} 3a^2cw-2ac^2v-a^2bw-a^2cv-2ab^2w+2b^2cv+abcv: \\ ac^2w-3ac^2v-2ab^2w+bc^2v+2a^2cw+2b^2cv-abcw: \\ 3ac^2v+a^2bw+a^2cv-ac^2w-bc^2v-3a^2cw-3abcv+3abcw \end{array} \right\rangle,$$

$$P_0P_2 = \left\langle \begin{array}{c} 4b^3u-4ab^2u+a^2bu+2ac^2u+2a^2cu+2ab^2w \\ +a^2bw-2b^2cu-3a^2cw-3abcu: \\ -(c-2b)\left(abu-2b^2u+abw+bcu-acw\right): \\ -(a-c)\left(abu-2b^2u+abw+bcu-acw\right) \end{array} \right\rangle,$$

$$P_1P_3 = \left\langle \begin{array}{c} 4b^3u-4ab^2u+a^2bu+2ac^2u+2a^2cu-2ab^2w \\ -a^2bw-2b^2cu+3a^2cw-3abcu: \\ -(c-2b)\left(abu-2b^2u-abw+bcu+acw\right): \\ -(a-c)\left(abu-2b^2u-abw+bcu+acw\right) \end{array} \right\rangle,$$

$$P_0P_3 = \left\langle \begin{array}{c} (a-2b+2c+2w)\left(abu-2b^2u-acv+bcu+bcv\right): \\ (a-2b+c)\left(2ac-13bc+2b^2+10c^2\right)u+ \\ c\left(-4ab+7ac-23bc+10b^2+10c^2\right)v: \\ (a-c)\left(2b^2u+2c^2u+2c^2v-abu+\right. \\ \left.2acu+acv-5bcu-3bcv\right) \end{array} \right\rangle,$$

$$P_1P_2 = \left\langle \begin{array}{c} (a-2b+2c-2w)\left(abu-2b^2u+acv+bcu-bcv\right): \\ (a-2b+c)\left(2ac-13bc+2b^2+10c^2\right)u+ \\ c\left(4ab-7ac+23bc-10b^2-10c^2\right)v: \\ (a-c)\left(2b^2u+2c^2u-2c^2v-abu+\right. \\ \left.2acu-acv-5bcu+3bcv\right) \end{array} \right\rangle.$$

4.2 Diagonal points of the Incenter Euler quadrangle $\overline{P_0P_1P_2P_3}$

Remarkably, the diagonal points of the Incenter Euler quadrangle $\overline{P_0P_1P_2P_3}$ have a particularly simple form, and in fact generally lie on the lines of the original triangle!

Theorem 9 *If $a^2c-ab^2-2abc+ac^2+2b^3-b^2c \neq 0$ and $c-2b \neq 0$ and $a-2b \neq 0$ and $a \neq c$, then the diagonal points of the quadrangle $\overline{P_0P_1P_2P_3}$ are*

$$D_1 \equiv (P_0P_1)(P_2P_3) = \left[\frac{a-2b}{a-c}, \frac{2b-c}{a-c}\right],$$

$$D_2 \equiv (P_0P_2)(P_1P_3) = \left[0, \frac{a-c}{2b-c}\right],$$

$$D_3 \equiv (P_0P_3)(P_1P_2) = \left[\frac{a-c}{a-2b}, 0\right],$$

which lie on the lines L_1, L_2 and L_3 respectively.

Proof. These are calculations that rely on the previous formulas for the Incenter Euler quadrangle lines, and involve simplifications using the quadratic relations, as well as cancellation of the terms that appear in the conditions of the theorem. □

We call D_1, D_2 and D_3 the **Diagonal Incenter Euler points** of the Triangle, and $\overline{D_1D_2D_3}$ the **Diagonal Incenter Euler triangle** of the Triangle $\overline{A_1A_2A_3}$. These two triangles, shown in Figure 7, have a remarkable relationship!

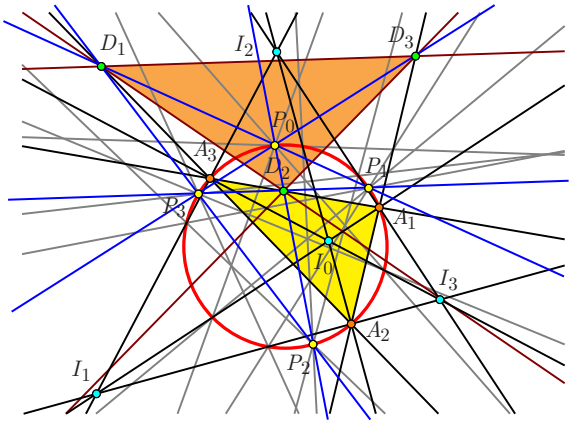


Figure 7: *The Diagonal Incenter Euler triangle $\overline{D_1D_2D_3}$ of the Triangle $\overline{A_1A_2A_3}$*

Theorem 10 *The signed area of the oriented Diagonal Incenter Euler triangle $\overrightarrow{D_1D_2D_3}$ is negative two times the signed area of the oriented original Triangle $\overrightarrow{A_1A_2A_3}$.*

Proof. This is a consequence of the formulas above for D_1, D_2 and D_3, together with the identities

$$\det\begin{pmatrix} \frac{a-2b}{a-c} & \frac{2b-c}{a-c} & 1 \\ 0 & \frac{a-c}{2b-c} & 1 \\ \frac{a-c}{a-2b} & 0 & 1 \end{pmatrix} = -2$$

and

$$\det\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} = 1.$$

□

Theorem 11 *The orthocenter of the Diagonal Incenter Euler triangle $\overline{D_1D_2D_3}$ is the Circumcenter C of the original Triangle $\overline{A_1A_2A_3}$.*

Proof. It is a straightforward calculation to show that the Circumcenter of $\overline{A_1A_2A_3}$ given by (3) is indeed also the orthocenter of $\overline{D_1D_2D_3}$. □

5 The Incenter Euler transformation

We may now define a transformation Γ at the level of triangles, where $\Gamma\left(\overline{A_1A_2A_3}\right)$ is the Diagonal Incenter Euler triangle $\overline{D_1D_2D_3}$. This gives a canonical second triangle associated to a given triangle, with one vertex of the new triangle on each of the lines of the original, where the signed area is multiplied by -2, and where the Circumcenter of the original Triangle becomes the orthocenter of the new triangle.

But now this transformation Γ allows one to transfer whole-scale triangle centers from $\overline{A_1A_2A_3}$ to $\overline{D_1D_2D_3}$. Generally every triangle center of $\overline{A_1A_2A_3}$ will then play a distinguished triangle center role for $\overline{D_1D_2D_3}$. Conceivably there are some particular exceptions, such as when one of the factors $a^2c-ab^2-2abc+ac^2+2b^3-b^2c$ or $c-2b\neq 0$ or $a-2b\neq 0$ or $a\neq c$ is zero.

This implies that Kimberling's list may well have a hall-of mirrors aspect, where once we identify a triangle center say X_i we consider the corresponding point for $\overline{A_1A_2A_3}$ to be a possibly new X_j of $\overline{D_1D_2D_3}$. This gives a natural mapping of Kimberling's list to itself. It seems an interesting question to identify what points go to what points. Could a computer be programmed to answer this question?

6 Incenter Euler line meets on the Lines

We have seen that the Incenter Euler lines meet at Incenter Midpoints (six) , at Incenter Euler points (four) and at Schiffler points (four). But there is more.

Theorem 12 *The Incenter Euler lines also meet at twelve points on the original Lines of the triangle, with four such meets on each Line.*

Proof. The calculation of these points are straightforward, the meets are, using projective coordinates:

$$e_{10}e_{13}=\begin{bmatrix}0:(a-c)(du+(b-c)v):\\ 3(b-c)(a-2b+c)u+\\ \left(2b^2-6bc+3c^2+ac\right)v\end{bmatrix},$$

$$e_{10}e_{12}=\begin{bmatrix}(a-c)(du+(a-b)w):0:\\ 3a^2u+6b^2u+3a^2w+2b^2w-9abu+3acu\\ -6abw-3bcu+acw\end{bmatrix},$$

$$e_{11}e_{12}=\begin{bmatrix}0:-(a-c)(du-(b-c)v):\\ 6b^2u+2b^2v+3c^2u+3c^2v-3abu+3acu\\ +acv-9bcu-6bcv\end{bmatrix},$$

$$e_{11}e_{13}=\begin{bmatrix}(a-c)(du-(a-b)w):0:\\ 3a^2u+6b^2u-3a^2w-2b^2w-9abu+3acu\\ +6abw-3bcu-acw\end{bmatrix},$$

$$e_{20}e_{21}=\left[c(aw-bv+bw-cv):0:2b^2w+acw-3bcv\right],$$

$$e_{22}e_{23}=\left[c(aw+bv+bw+cv):0:2b^2w+acw+3bcv\right],$$

$$e_{20}e_{23}=\begin{bmatrix}-c(au-3bu+bv+2cu-2cv):\\ (c-2b)(bu-cu+cv):\\ -\left(2b^2u+3c^2u-3c^2v+acu-6bcu+3bcv\right)\end{bmatrix},$$

$$e_{21}e_{22}=\begin{bmatrix}c(au-3bu-bv+2cu+2cv):\\ (c-2b)(cu-bu+cv):\\ 2b^2u+3c^2u+3c^2v+acu-6bcu-3bcv\end{bmatrix},$$

$$e_{30}e_{31}=\left[0:a(aw-bv+bw-cv):-2b^2v+3abw-acv\right],$$

$$e_{32}e_{33}=\left[0:(aw+bv+bw+cv):2b^2v+3abw+acv\right],$$

$$e_{31}e_{33}=\begin{bmatrix}(a-2b)(bu-au+aw):\\ -a(2au-3bu-2aw+cu+bw):\\ -\left(3a^2u+2b^2u-3a^2w-6abu+acu+3abw\right)\end{bmatrix},$$

$$e_{30}e_{32}=\begin{bmatrix}(a-2b)(au-bu+aw):\\ a(2au-3bu+2aw+cu-bw):\\ 3a^2u+2b^2u+3a^2w-6abu+acu-3abw\end{bmatrix}.$$

□

7 The mystery of apparent symmetry breaking

There is another very intriguing aspect of this entire story that invites further exploration. The lines of the Diagonal Incenter Euler triangle $\overline{D_1D_2D_3}$ of the Triangle $\overline{A_1A_2A_3}$ with Incenters I_0,I_1,I_2 and I_3 can be easily computed to be

$$D_1D_2=[a+2b-2c:2b-c:c-a]$$
$$D_2D_3=[a-2b:2b-c:c-a]$$
$$D_1D_3=[2b-a:2b-2a+c:a-c].$$

It is first of all remarkable that the formulas for these lines are simple linear expressions in the numbers a,b and c of the matrix for the bilinear form. In Figure 7 we notice that the line D_2D_3 appears to pass through I_3, but the other two lines D_1D_2 and D_1D_3 do not pass through any of the other Incenters. If this were true, it would imply a completely remarkable, even seemingly impossible, symmetry breaking.

Why should the Incenter I_3 be singled out in this fashion? This very curious situation may at first confound the experienced geometer, as it did us when we first observed it. The reader might enjoy creating such a diagram and determining to what extent this phenomenon holds, and trying to find an explanation of it. We will address this challenge in a future paper.

References

[1] W. CHEN, The Incenter/Excenter lemma, `http://www.mit.edu/\symbol{126}evanchen/handouts/Fact5/Fact5.pdf`, 2015.

[2] L. EMELYANOV, T. EMELYANOVA, A Note on the Schiffler Point, *Forum Geom*. **3** (2003), 113–116.

[3] C. KIMBERLING, *Triangle Centers and Central Triangles*, Congressus Numerantium **129**, Utilitas Mathematica Publishing, Winnipeg, MA, 1998.

[4] C. KIMBERLING, *Encyclopedia of Triangle Centers*, `http://faculty.evansville.edu/ck6/encyclopedia/ETC.html`.

[5] C. KIMBERLING, Major Center of Triangles, *Amer. Math. Monthly* **104** (1997), 431–438.

[6] N. LE, N. J. WILDBERGER, Incenter Circles, Chromogeometry, and the Omega Triangle, *KoG* **18** (2014), 5–18.

[7] N. LE, N. J. WILDBERGER, Universal Affine Triangle Geometry and Four-fold Incenter Symmetry, *KoG* **16** (2012), 63–80.

[8] K. L. NGUYEN, On the Complement of the Schiffler point, *Forum Geom.* **5** (2005), 149–164.

[9] K. SCHIFFLER, G. R. VELDKAMP, W. A. SPEK, Problem 1018 and Solution, *Crux Mathematicorum* **12** (1986), 150–152.

[10] C. THAS, On the Schiffler center, *Forum Geom.* **4** (2004), 85–95.

[11] N. J. WILDBERGER, *Divine Proportions: Rational Trigonometry to Universal Geometry*, Wild Egg Books, `http://wildegg.com`, Sydney, 2005.

[12] N. J. WILDBERGER, Affine and projective metrical geometry, `arXiv:math/0612499v1`, 2006.

[13] N. J. WILDBERGER, Chromogeometry, *Mathematical Intelligencer* **32**/1 (2010), 26–32.

[14] N. J. WILDBERGER, Chromogeometry and Relativistic Conics, *KoG* **13** (2009), 43–50.

Nguyen Le
e-mail: nguyenlecm2009@gmail.com
San Francisco State University
San Francisco, CA 94132, United States

N J Wildberger
e-mail: n.wildberger@unsw.edu.au
School of Mathematics and Statistics UNSW
Sydney 2052 Australia

Original scientific paper
Accepted 30. 11. 2016.

GUNTER WEISS

Special Conics in a Hyperbolic Plane

Special Conics in a Hyperbolic Plane

ABSTRACT

In Euclidean geometry we find three types of special conics, which are distinguished with respect to the Euclidean similarity group: circles, parabolas, and equilateral hyperbolas. They have on one hand special elementary geometric properties (c.f. [7]) and on the other they are strongly connected to the "absolute elliptic involution" in the ideal line of the projectively enclosed Euclidean plane. Therefore, in a hyperbolic plane (h-plane) – and similarly in any Cayley-Klein plane – the analogue question has to consider projective geometric properties as well as hyperbolic-elementary geometric properties. It turns out that the classical concepts "circle", "parabola", and "(equilateral) hyperbola" do not suit very well to the many cases of conics in a hyperbolic plane (c.f. e.g. [10]). Nevertheless, one can consider conics in a h-plane systematicly having one ore more properties of the three Euclidean special conics. Place of action will be the "universal hyperbolic plane" π, i.e., the full projective plane endowed with a hyperbolic polarity ruling distance and angle measure.

Key words: conic section, hyperbolic plane, Thales conic, equilateral hyperbola

MSC2010: 51M09

Specijalne konike u hiperboličnoj ravnini

SAŽETAK

U euklidskoj ravnini s obzirom na euklidsku grupu simetrija razlikujemo tri tipa specijalnih konika: kružnice, parabole i specijalne hiperbole. S jedne strane, one imaju specijalno euklidsko svojstvo (vidi [7]), a s druge su strane čvrsto vezane uz apsolutnu eliptičnu involuciju na idealnom pravcu projektivno proširene euklidske ravnine. Zbog toga, u hiperboličnoj ravnini (h-ravnini) – i slično u svakoj Cayley-Kleinovoj ravnini – treba promatrati i projektivna geometrijska svojstva i elementarno-hiperbolična geometrijska svojstva. Pokazuje se da u brojnim slučajevima konika u hiperboličnoj ravnini klasični koncepti "kružnica", "parabola" i "(jednakostranična) hiperbola" nisu primjenjivi (vidi npr. [10]). Unatoč tome, moguće je sustavno promatranje konika u h-ravnini koje imaju jedno ili više svojstava triju euklidskih specijalnih konika. Proučavanje će se vršiti na "univerzalnoj hiperboličnoj ravnini" π, tj. projektivnoj ravnini u kojoj su udaljenost i mjera kuta definirani apsolutnim polaritetom.

Ključne riječi: konika, hiperbolična ravnina, Talesova konika, jednakostranična hiperbola

1 Introduction

We consider conics in a hyperbolic plane (h-plane) having one ore more properties of the three Euclidean special conics "circle", "parabola" and "equilateral hyperbola". In the projectively enclosed and complexified Euclidean plane circles are conics passing through the (complex conjugate) absolute points I, J on the ideal line u of that plane, parabolas touch this absolute line u, and equilateral hyperbolas intersect the absolute line u in points harmonic to I, J. Besides these projective geometric properties, the three special conics have many Euclidean properties and generations.

Circles are e.g. generated as distance curves of a point, the midpoint, but they are also generated by directly congruent pencils of lines, what results in the remarkable inscribed angle theorem and the theorem of Thales as its special case.

Parabolas are e.g. generated as envelope of a leg of a right angle hook sliding along a line, while the other leg passes through a point. The fixed line and point turns out to be vertex tangent and focus of the generated parabola. As a Euclidean conic, a parabola is of course also defined via the Apollonius definition of a conic.

An equilateral hyperbola has orthogonal asymptotes. It is (directly and indirectly) congruent to its conjugate hyperbola, it is generated by indirectly congruent pencils of lines. But the most strange property is that each triangle of points on the hyperbola has its orthocentre on this equilateral hyperbola. The pencil of conics with the vertices of a triangle and its orthocentre as base points consists only of equilateral hyperbolas; (the singular conics are the pairs consisting of the altitude and the corresponding side of the given triangle).

In Figure 1, these three Euclidean cases and their well-known main properties are visualised symbolicly:

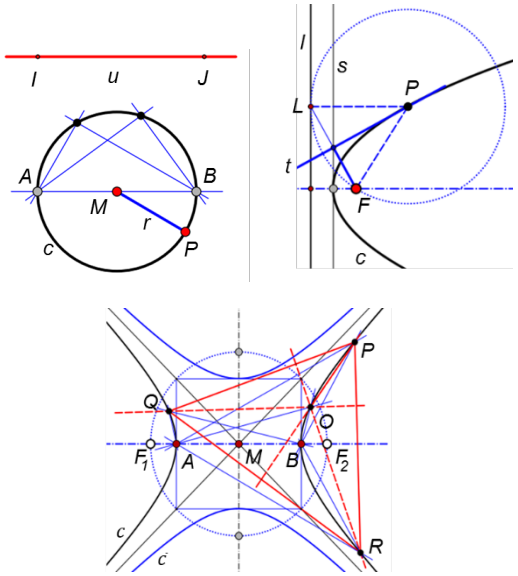


Figure 1: *(Symbolic) visualisation of projective and metric properties of a Euclidean circle, a parabola, and an equilateral hyperbola*

In the following, we shall study conics in a hyperbolic plane being defined by one of the mentioned projective and Euclidean properties. Place of action will be the "universal hyperbolic plane" π, i.e. the full real projective plane endowed with a hyperbolic polarity, see e.g. [15]. This "absolute (regular) hyperbolic polarity" is usually given by the real conic ω and rules orthogonality, h-distance, and h-angle measure. F. Klein's point of view considers h-geometry as a sub-geometry of projective geometry while a puristical point of view allows only the inner domain of ω for being a proper h-plane. In the following, we try to have both points of view in mind, but we will use F. Klein's projective geometric model of a h-plane.

The standard graphic representation of ω is that of a Euclidean circle and this allows us to use e.g. the graphics software "Cinderella" (see [11]), which has the feature "(planar) hyperbolic geometry".

As a first and well-known example we consider circles: In a hyperbolic plane with the (real) "absolute conic" ω a circle c is a conic touching ω twice in algebraic sense (what means disregarding reality and coincidence of the touching points). This projective geometric approach results already in three types of hyperbolic circles: (1) proper circles touching ω in a pair of complex conjugate points, (2) limit circles which hyperosculate ω, and (3) distance circles touching ω in a pair of real distinct points. While the elementary Euclidean definition of a circle as the planar set of points having equal distance from a centre point, the analogue in hyperbolic geometry is true only for h-circles of type (1) and can be modified for h-circles of type (3). For h-circles of type (3) the radius length is not finite, a property, which connects this type rather with Euclidean parabolas than with circles. Euclidean circles can be generated via directly congruent pencils of lines, which expresses the property of a constant angle at circumference and especially the property of Thales. In an h-plane two pencils of orthogonal lines with proper base points generate the so-called "Thales conic" (resp. "Thaloid", as it is called by N. J. Wildberger, see [15]), which is never an h-circle of type (1) and (2), while h-circles of type (3) occur if, and only if the vertices of the h-orthogonal pencils both are ideal points on ω.

Conics with the properties of a Euclidean parabola are treated in [1], where the place of action again is the "universal hyperbolic plane" π. But, for the sake of completeness, we also repeat some of the details here.

A great part of this article will deal with h-conics derived from properties of the Euclidean equilateral hyperbola following the above presented systematic treatment for circles. This results in two special sets of h-conics, the set of "h-equilateral conics" having a harmonic quadrangle of ideal points and the set of h-conics defined by the property that each triangle of conic points has its h-orthocentre also on this conic.

2 Projective geometric classification of h-conics

A given conic c, together with ω, defines as well a pencil of conics $p \cdot c + q \cdot \omega$ as well as a dual pencil of (dual) conics and we distinguish 5 different types of pencils according to the sets of singular conics resp. the sets of common base points and base tangents (see Figures 2 and 3).

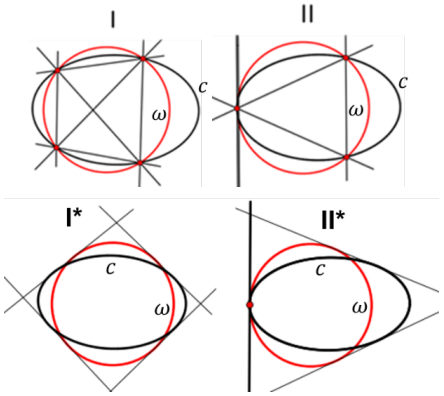


Figure 2: *Conic pencils I and II and its dual pencils I^* and II^**

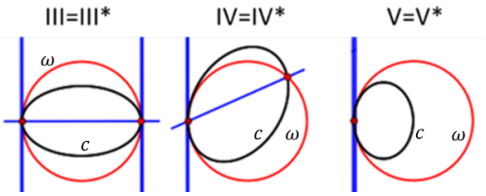


Figure 3: *The self-dual pencils of conics III, IV, and V*

If we take the reality of base points or base lines into account and consider the pencils I and I*, then we have three subcases each, the pencils of type II, II* and III have two such subcases each and there is only one case of pencils of type IV and V. A further distinction can be made concerning the reality of the common polar triangle of c and ω, which acts as the h-midpoint triangle of c: An h-conic can have either one or three real midpoints.

An overview of all possible cases can be found in [5], [6], [9], and [10]. This classification shows that the Euclidean names "ellipse", "parabola" and "hyperbola" need strong modifications (as e.g. "semi-hyperbola", "convex resp. concave hyperbolic parabola" and so on) to express the type of an h-conic, which becomes obvious by its visualisation in some model of the h-plane π.

Pencils $p \cdot c + q \cdot \omega$ of the two subtypes III define h-circles c. They have the well-known property of being distance loci of either points or lines, see Figure 4. h-Conics to type V have no finite radius lenght and are called "limit h-circles" or "horocycles". But as they have similar properties as Euclidean parabolas they can also be considered as special cases of h-parabolas, Figure 5.

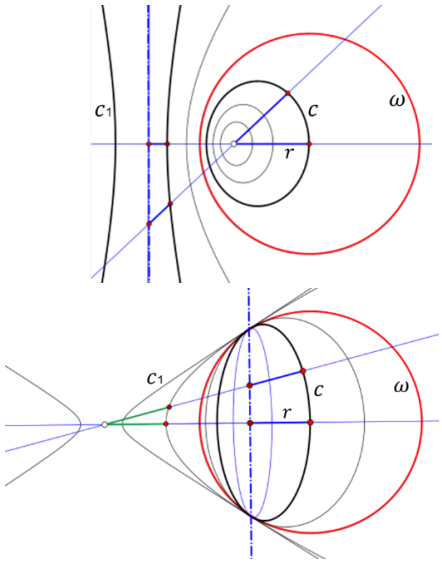


Figure 4: *The different types of h-circles within concentric pencils of h-circles*

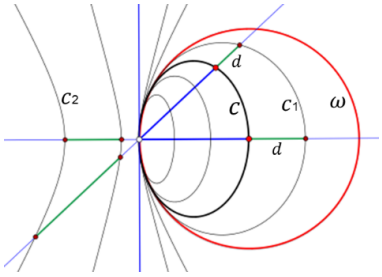


Figure 5: *Pencil of "horocycles" showing the property of Euclidean parabolas, which are translated along their common axis*

Conics c defining pencils of type II and IV can be considered as analogs to Euclidean parabolas and they will be studied in Chapter 5 with respect to their h-metric properties.

Conics c defining pencils of type I are "h-hyperbolas", "semi-hyperbolas" or "h-ellipses" according to the reality of the pencil's base points, which furtheron will be called the "ideal points" of c.

3 Classification with respect to h-orthogonality

We start with an h-conic c, which, together with ω defines a pencil $p \cdot c + q \cdot \omega$ of type I. The quadrangle of its ideal points can be special with respect to the h-orthogonality structure defined by the absolute polarity to ω. There might be h-orthogonal pairs of opposite sides of this base quadrangle (Figure 6). For dual pencils $p \cdot c^* + q \cdot \omega^*$ one can find similar special cases, see Figure 7. Quadrangles (resp. quadrilaterals) with this special property are called "harmonic quadrangles" (resp. "harmonic quadrilaterals") as their (non-trivial) symmetry group is generated by harmonic homologies.

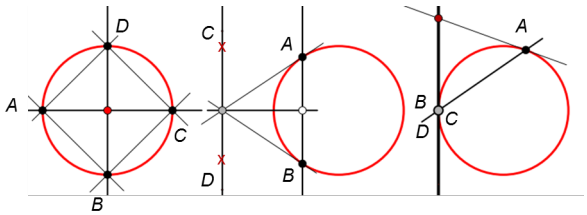


Figure 6: *Pencils $p \cdot c + q \cdot \omega$ with h-orthogonality properties of the base quadrangle (and its degenerate case pencil IV)*

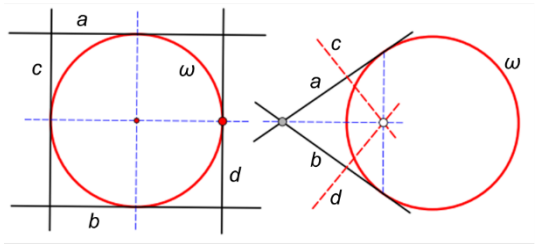


Figure 7: *Dual pencils $p \cdot c^* + q \cdot \omega^*$ of type I^* with h-orthogonality properties of the base quadrilateral*

Figure 6 shows that there occur only h-hyperbolas and semi-hyperbolas and, as a degenerate case, there are h-parabolas, but no h-ellipses. The dual case (Figure 7) contains h-ellipses c, when seen as point conics. It makes sense to define an h-conic possessing a real harmonic ideal quadrangle as "h-equilateral hyperbola" and we will have a closer look to these h-conics in Chapter 6.

4 h-Conics with metric properties of Euclidean circles

We start with the concept of a Euclidean circle and its different Euclidean generations. We have already mentioned that projective h-conics of type III are also a h-distance circles (Figure 4). But they have neither the Thales-property nor the property of the constant angle at circumference. The generation of a conic by h-orthogonal pencils delivers the "Thales conic" x over a segment $[A,B]$, which turns out to be one of the axes of the conic (Figure 8). There is one exeption: Thales-conics over a segment $[A,B]$, $A,B \in \omega$, are h-circles and their radius turns out to be $1/\sqrt{2}$, (Figure 9).

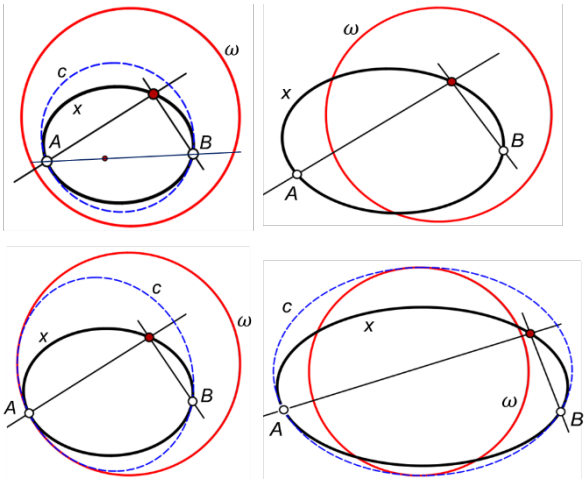


Figure 8: *Different cases of Thales-conics x generated by two h-orthogonal pencils of lines*

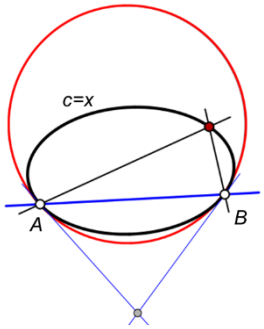


Figure 9: *The exceptional case of a Thales-conic over a segment with endpoints on the absolute conic* ω *in a h-circle of radius* $1/\sqrt{2}$

Connecting the construction of a Thales-conic x with a kinematic mechanism allows us to construct of points and tangents of x, see [14] and Figure 10.

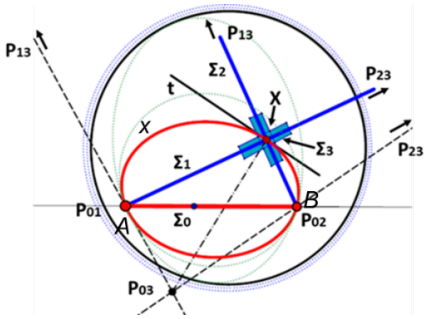


Figure 10: *Kinematic generation of a Thales-conic x applied to construct its tangents*

Arbitrarily chosen direct congruent pencils of lines generate a conic, too. It is simply the Steiner generation of a conic by projective pencils, but this delivers no h-circles, see Figure 11, and it has not the property of inscibed angle theorem either!

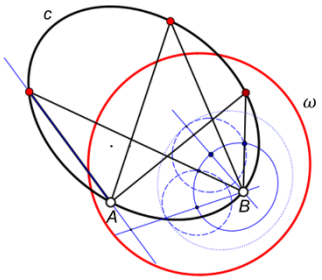


Figure 11: *Conic generated by two directly congurent pencils of lines*

Curves defined by a constant angle at circumference different from a h-right angle, socalled "isoptics of a segment", turn out to be algebraic of degree four! To visualize this one can consider the inverse motion, namely to keep the angle and its legs fixed and move the segment. In the Euclidean plane this motion is the well-known *ellipse motion*,

as all points which are fixed connected with the moving system and which are different from the points on the legs of the fixed angle have ellipses as orbits. In the h-plane these orbits turn out to be of order four, too, see Figure 12.

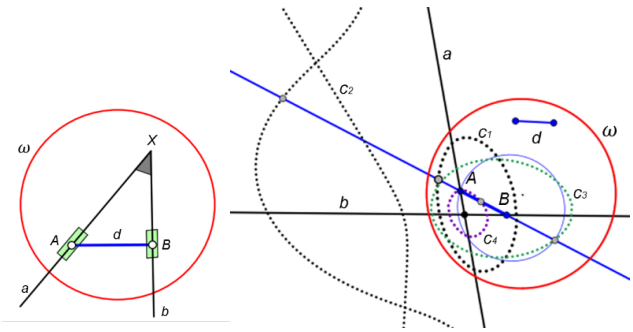


Figure 12: *The h-analog of the Euclidean ellipse motion delivers orbits c_i of degree 4*

Another way to visualise this motion is to start with the Thales-motion and consider the envelop of a line fixed to Thales's right angle hook and passing through its vertex, Figure 13. In the Euclidean case, the envelope is a point of the Thales-circle, in the h-case it is a curve of degree 6.

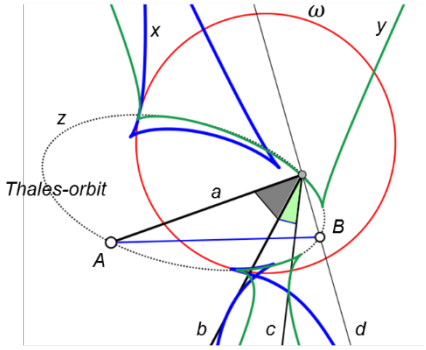


Figure 13: *Moving a fixed angle hook along a Thales-conic such that one leg passes through a bases point A of Thales's construction, the 2nd leg envelops a curve of degree 6.*

5 h-Conics with metric properties of Euclidean parabolas

In this chapter, we strongly refer to [1]. From the projective geometric point of view, we have to distinguish h-parabolas of type II, IV, and V, the latter having also properties of a circle. If one considers the analog of the Euclidean slider crank, there occur h-conics, but they are (in general) not projective h-parabolas, (Figure 14). The proof for the fact that the envelop of the second leg t of the crank slider is a h-conic x is trivial: The line t connects two projectively correlated point series, namely s and the absolute polar line f of F. This line f is also the second vertex tangent of the h-conic x.

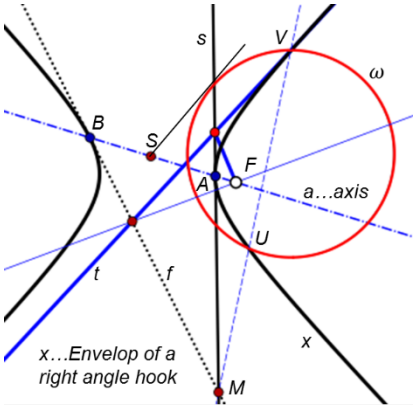


Figure 14: *The h-analog of the Euclidean crank slider motion, defines a conic x with focus F and vertex tangent s.*

Also the h-analog to the construction of a parabola according to Apollonius's definition does in general not deliver h-parabolas in the projective geometric sense, see Figure 15. (Proof: x is Steiner-generated by two projective pencils of lines with centre F, the focus of x, and the absolute Pole L of the directrix line l of x.)

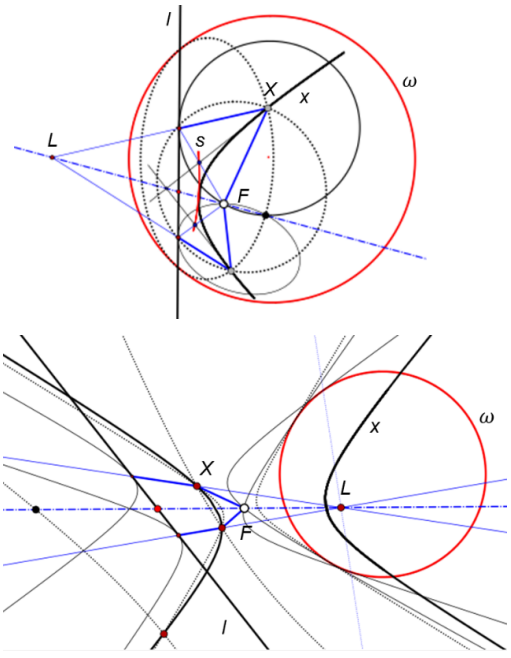


Figure 15: *The h-analog of Apollonius's definition of a Euclidean parabola delivers a conic x.*

Figure 15 also shows that x fulfills a "reflection property" similar to the Euclidean parabola. But while, in the Euclidean case, the diameters of the parabola c are reflected at x, and then pass through the focus of x, in the hyperbolic

case the lines through the absolute pole L of the directrix line l are reflected at x and pass through a point, the h-focus F_1. The points L and F_1 act, therefore, as a pair of foci of x.

As all h-conics with two real focal points have the "reflection property", it seems to be obvious that the projective h-parabolas x have the "reflection property", too, i.e., the h-reflections of "diameters" at x pass through a point, the focus of x. It turns out that this is true for all cases, Figure 16. For the limit case of pencil type V it seems to be trivial, as the diameters intersect the "limit circle" x h-orthogonally, a property, which we connect rather to circles then to a parabola and which justifies to call that special h-conic a *limit circle* (horocycle) and not a *limit parabola*.

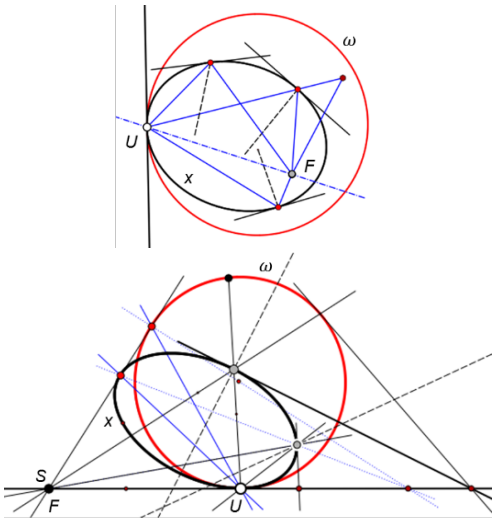


Figure 16: *Reflection property of h-parabolas of the projective types II and IV*

6 h-Conics with properties of Euclidean equilateral hyperbolas

6.1 h-equilateral conics

Here, we continue chapter 3: An "h-equilateral hyperbola" x is an h-conic with a (real) harmonic quadrangle Ω of ideal points, i.e. $x \cap \omega = A,B,C,D$ and $\mathrm{CR}(A,B,C,D) = -1$. Similarly we call an h-equilateral conic x with two real and two conjugate imaginary ideal points an "h-equilateral semi-hyperbola", see Figure 6. For both cases exactly one real pair of sides of the ideal harmonic quadrangle Ω is h-orthogonal and we will call this pair the "asymptotes" a_1, a_2 of x, c.f. also [10]. The other two (real or conjugate imaginary) pairs of sides of the complete quadrangle Ω shall be named as "singular h-equilateral conics" Σ_1, Σ_2. The vertices of the diagonal triangle of Ω are the h-midpoints M_i of x, Figure 17.

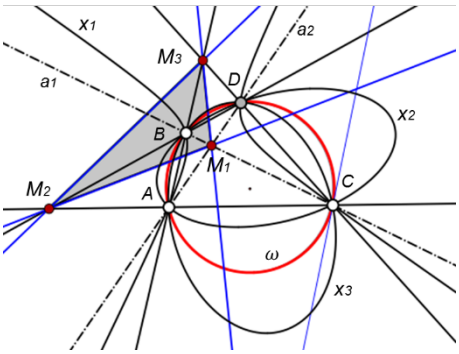


Figure 17: *Concentric h-equilateral hyperbolas x_1, x_2, x_3 with their common asymptotes a_1, a_2, and h-midpoints M_i*

A h-equilateral semi-hyperbola can be visualized in the classical, but projectively closed plane of visual perception as a *Euclidean pencil of circles*, see Figure 18.

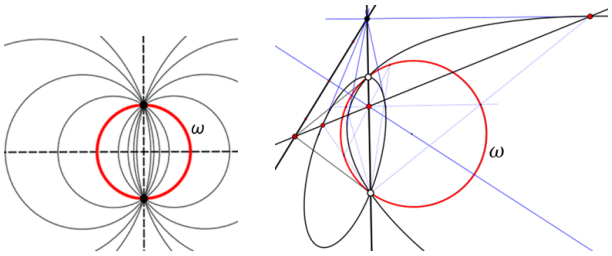


Figure 18: *Left: Euclidean model of h-equilateral semi-hyperbolas, Right: A more projective visualisation of these special h-conics*

Figure 19 shows that even a Thales-conic can be an h-equilateral hyperbola.

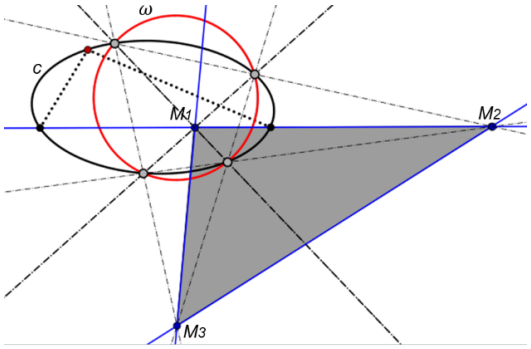


Figure 19: *A Thales-conic can be an h-equilateral hyperbola*

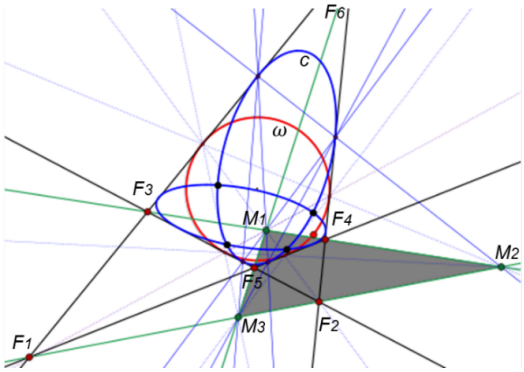


Figure 20: *Dual pencil of h-equilateral dual conics c^**

Figure 20 illustrates the case of a dual pencil of h-equilateral dual conics c^* touching a harmonic (ideal) quadrilateral Ω^*. ("Dual conic" means the set of tangents of a conic c, which is given as a point set.) The three pairs of intersection points of the complete harmonic quadrilateral Ω^* act as the pairs of focal points F_i of the conics c. Again, there exists one real pair of absolute conjugate focal points F_1, F_2, and we call these points the "asymptote points" of c^*. The diagonal points of Ω^* are the midpoints of the conics c to c^*.

If one applies an h-reflection to an h-equilateral hyperbola c in one of its asymptotes, one receives another h-equilateral hyperbola $\overline{c}$, which is h-congruent to c. This repeats a property of Euclidean equilateral hyperbolas. In the Euclidean case c and $\overline{c}$ are "conjugate hyperbolas". This gives a hint to define h-conjugate conics, too, see Figure 21:

Definition 1 *The "h-conjugate conic" y to conic x is concentric with x, has the same ideal points, and thus, the same asymptotes a_1, a_2, and it has the same "axis-quadrangle" Λ. The axis-quadrangle Λ of x has its vertices in the intersection points of vertex tangents w_1, w_2, of x with its asymptotes a_1, a_2.*

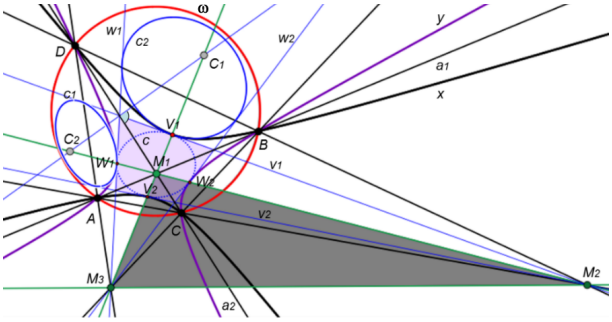


Figure 21: *The h-conjugate conics x and y in the case of x being h-equilateral*

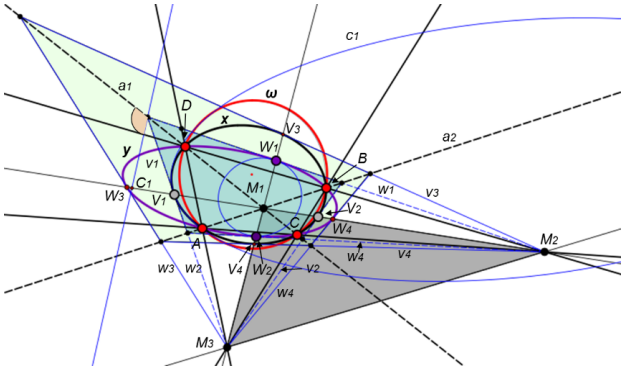


Figure 22: *The h-conjugate conics x and y in the case of x being not h-equilateral*

The special case visualised in Figure 21 also reveals the construction of the hyper-osculating h-circles c_i at the vertices W_1 of x and V_1 of y, which is identical with the Euclidean construction for equilateral hyperbolas. Figure 22 shows the general case of (real) h-conjugate conics. There, too, the classical Euclidean construction of the hyperosculating circles at the vertices is possible. So, we can state the following

Result: If in an arbitrary Cayley-Klein plane (CK-plane) π a conic x has a well-definded real pair of asymptots a_1, a_2, then it has a real CK-conjugate conic y. The CK-normals in vertices of the "axis-quadrangle" of x (defined above) to the asymptotes intersect the axes of x (and y) in the h-centres C_i of the hyperosculating circles c_i of x and y.

6.2 Special h-conics generated by h-congruent pencils of lines

Euclidean equilateral hyperbolas have the property that they can be Steiner-generated by two indirect congruent pencils of lines. Obviously, the result of the analog construction in the h-plane π delivers a conic x, but this conic is, in general, not h-equilateral, see Figure 23. As also a Thales conic can have four real ideal points, see e.g. Figure 19, the sense of the congruence between the two pencils of lines is not essential for receiving a hyperbola as the result of the Steiner-generation.

Remark 1 *In Figure 23 the basis points of both Steiner-generations, the direct and indirect one, are labelled with P, Q. It turns out that the segment $[P,Q]$ is a diameter of the indirectly generated h-conic x as well as of the directly generated h-conic y. It is still an open question, whether any arbitrary h-conic x can be Steiner-generated by h-congruent pencils of lines. In the Euclidean case this is not true.*

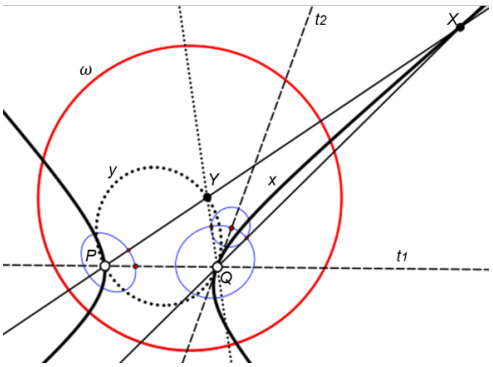


Figure 23: *Generation of a conics x, y via h-congruent pencils of lines. The h-congruence is given by the sense (indirect or direct) and the pair of corresponding lines t_1, t_2. The dotted conic y is the result of directly congruent pencils while the x stems from indirectly congruent pencils.*

6.3 Special h-conics with the "triangle orthocenter property"

It is well-known that Euclidean equilateral hyperbolas x are characterised by the remarkable property that any triangle Δ of hyperbola points has its orthocentre on x. We abbreviate this property as the "triangle orthocentre property" of (Euclidean) equilateral hyperbolas and pose the question whether there exist h-conics with this property, too.
We start with an arbitrary but not right-angled triangle $\Delta = (APQ)$ with O being its h-orthocentre, and we add an arbitrarily chosen fifth point X for defining an h-conic x through these 5 points. We assume x to be regular, see Figure 24.

Theorem 1 *A h-conic x through A,O,P,Q,X, with O the h-orthocentre of $\Delta = (APQ)$ passes also through the h-orthocentre O_X of triangle $\Delta_X = (XPQ)$.*

Proof. Using the labelling of Figure 23 with $a_1 = PA$, $a_2 = QA$, $b_1 = PX$, $b_2 = QX$ and $c_1 = PO$, $c_2 = QO$, $d_1 = PO_X$, $d_2 = QO_X$, we have the following pairs of h-orthogonal lines:

$$a_1 \perp c_2, \quad b_1 \perp d_2, \quad c_1 \perp a_2 \quad \text{and} \quad d_1 \mapsto b_2. \tag{1}$$

The two ordered quadruples (a_1,b_1,c_1,d_1), (c_2,d_2,a_2,b_2) belong to h-orthogonal pencils, which Steiner-generate the Thales conic t over the segment $[P,Q]$. Therefore, we can state that

$$\mathrm{CR}(a_1,b_1,c_1,d_1) = \mathrm{CR}(c_2,d_2,a_2,b_2). \tag{2}$$

By applying permutation rules for cross ratios (see e.g. [2, p. 34]), we infer

$$\mathrm{CR}(a_1,b_1,c_1,d_1) = \mathrm{CR}(c_2,d_2,a_2,b_2) = \mathrm{CR}(a_2,b_2,c_2,d_2), \tag{3}$$

such that also the ordering $a_1 \mapsto a_2$, $b_1 \mapsto b_2$, $c_1 \mapsto c_2$, $d_1 \mapsto d_2$ defines projective pencils, which Steiner-generate the conic x through P,Q,A,X,O. Since (2) holds, we also have $d_1 \cap d_2 = O_X \in x$. □

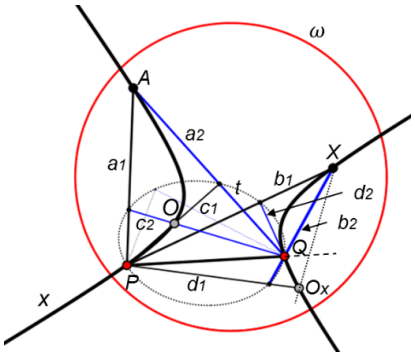


Figure 24: *An h-conic x through A,O,P,Q,X, with O the h-orthocentre of $\Delta = (APQ)$ passes also through the h-orthocentre O_X of the triangle $\Delta_X = (XPQ)$.*

Applying Theorem 1 to different points $X_i \in x$ allows us to go from the basic triangle (A,P,Q) to any other triangle (X_1,X_2,X_3) of conic points, and also this new triangle must have its h-orthocentre on x. So we can state

Theorem 2 *A (regular) h-conic x passing through A,O,P,Q, with O being the h-orthocentre of triangle (A,P,Q) has the triangle orthocentre property, i.e. any triangle of points of x has its h-orthocentre on x.*

One can extend Theorem 1 by the following statement, (see Figure 25):

Theorem 3 *Any conic x through points A,B,C,D and passing through the h-orthocentre O_1 of triangle (ABC) passes also through the h-orthocentres O_i of (BCD), (ABD), and (ACD). Especially, if $A,B,C,D \in \omega$ and (A,B,C,D) is not harmonic, then the diagonal triangle of the quadrangle $(O_1,\ldots,O_4)$ coincides with that of (A,B,C,D), which is the midpoint triangle of x.*

Proof. The first part of theorem 3 is simply a consequence of theorem 2. Now we consider a quadrangle $\Omega = (A,B,C,D)$ of ideal points. Let O_1,O_2 be the h-orthocentres of the triangles (ABC) and (ABD), see Figure 26. The quadrangle Ω admits the h-reflections in the sides of its diagonal triangle (M_1,M_2,M_3). Thereby, the h-symmetry σ with centre $Z := M_3$ and axis $z := M_1M_2$

maps triangle (ABC) to the triangle (ABD) and, as σ is an h-congruence, it maps also O_1 to O_2 which implies that O_1, O_2, Z are collinear. Applying the other possible h-symmetries with centres M_i and axes M_jM_k to the h-orthocenters of the remaining partial triangles Ω completes the proof. □

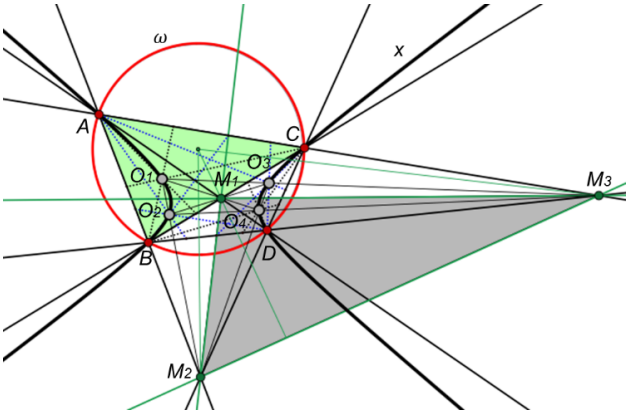


Figure 25: *An h-conic x through A,B,C,D,O_1 with O_1 the h-orthocentre of (ABC) also passes through the h-orthocentres O_i of the remaining partial triangles of (A,B,C,D).*

Remark 2 *The h-symmetry argument used in the proof above suggests an extension of Theorem 3: Each quadrangle (A,B,C,D) admitting three h-symmetries would suit as start figure such that the quadrangle of h-orthocentres of the partial triangles has the diagonal triangle in common with (A,B,C,D).*

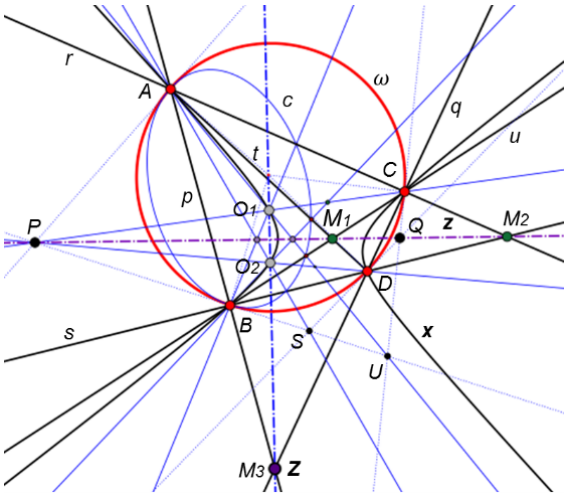


Figure 26: *Applying the h-symmetry $\sigma\colon (A,B,C,D) \mapsto (A,B,D,C)$ proves that O_1, O_2, Z are collinear.*

Remark 3 *Both, the set of h-equilateral hyperbolas possessing one pair of h-orthogonal asymptotes and the set of conics with the triangle orthocentre property ("top-conics") are four-parametric with a three-parametric family of h-conics having both properties, while in Euclidean geometry the two four-parametric families coincide.*

7 Final remarks and conclusion

Conics in Euclidean and non-Euclidean geometries are already widely studied since decades, see e.g. [3], [5], [6] and also the reference list in the monograph on conics [7]. Many references mainly deal with the classification and normal form problem and less with explicite constructions or properties of conics, see e.g. [10], [12]. Explicit constructions can be found e.g. in [1] and [14].

This article aims at a systematic treatment of what can be called "special conics" in a hyperbolic plane. This means that we have to base the investigation on usual classifications of conics from the (projective) universal hyperbolic point of view as well as on the basis of special properties which are non-Euclidean adaptions of properties one can find at Euclidean conics. As one can interpret many of these adaptions simply as Steiner-generations of conics (or its dual), one can widely omit calculations and use synthetic reasoning instead.

Because of the used projective geometric point of view, it is an easy task to transfer the presented results resp. the constructions also to *elliptic geometry*. In an elliptic plane (or its Euclidean spherical model), there are no parabolas even so the constructions for Euclidean parabolas can be performed. Each general conic in the elliptic plane is an *ellipse*, but a spherical conic allows both, the Apollonius-definition of an ellipse and (seen from the complementary side in the spherical model) also that of a hyperbola. As special projective types of (real) conics, one finds one type of "e-circles". All the other metric definitions (as e.g. by the triangle-orthocentre-property) deliver "e-ellipses" with special properties or curves of higher degree.

We conclude with open questions:

It remains open, whether there exist additional special conics in hyperbolic geometry, which have properties one did not consider in Euclidean geometry. One such property which makes no sense in Euclidean geometry but is meaningful in hyperbolic and elliptic planes, is the dual to the Apollonius definition of a conic:

"The tangents of a conic intersect two given lines in angles of constant sum."

For elliptic resp. spherical geometry, this results in a nice application: Given two lines a,b intersecting in C, find points $B \in a$, $A \in b$, such that the spherical triangle ABC has a given area.

As a second open problem occurs, whether each h-conic can be Steiner-generated via two *congruent* pencils of lines. For h-special hyperbolas the generation via congru-

ent pencils leads to pencil vertices on a diameter of the hyperbola. As there is a 5-parametric set of h-conics in the h-plane, and there is also a 5-parametric set of congruent pencils, (namely 2 times two for the vertices and one for the rotation given by start line t_2 to a fixed start line t_1), this question might be answered with "yes", even so it is wrong in Euclidean geometry. But if "yes" is true, how can one find these vertices and the angle of rotation to a given h-conic?

A third question concerns the "h-isoptic curves of a segment", which generalise the incribed angle theorem in Euclidean geometry. Is it possible that the h-isoptic curve, which is irreducible of degree 4 in geneneral, can be reducible in some special cases? This would be similar to the Euclidean case, where the resulting curve of degree 4 always splits into two circular arcs?

A fourth problem might concern the hyperbolic versions of some generalisations of Euclidean Thales-constructions as presented in [13].

Final remark. *Even so the topic of dealing with special conics a specific CK-plane only seems to be what can be called "advanced elementary geometry", it could stimulate research of conics - namely as curves of degree 2 - in arbitrary metric planes, so-called Minkowski planes. Until now "Minkowski conics" are defined only via the Apollonius definition, see e.g.* [4], [8].

Acknowledgement

The author thanks Prof. Molnár and Prof. Horváth [9] for hints and references and Prof. Norman Wildberger for fruitful discussions.

References

[1] A. ALKHALDI, N. J. WILDBERGER, The Parabola in Universal Hyperbolic Geometry I, *KoG* **17** (2013), 14–42.

[2] H. BRAUNER, *Geometrie projektiver Räume II*, BI-Wissenschaftsverlag, Mannheim, Wien, Zürich 1976.

[3] P. CHAO, J. ROSENBERG, Different definitions of conic sections in hyperbolic geometry, `https://arxiv.org/pdf/1603.09285v1.pdf`

[4] A. FANKHÄNEL, On Conics in Minkowski Planes, *Extracta Math.* **27**/1 (2012), 13–29.

[5] K. FLADT, Die allgemeine Kegelschnittsgleichung in der ebenen hyperbolischen Geometrie, *Journ. f. d. Reine u. Angew. Math.* **197** (1957), 121–139.

[6] K. FLADT, Die allgemeine Kegelschnittsgleichung in der ebenen hyperbolischen Geometrie II, *Journ. f. d. Reine u. Angew. Math.* **199** (1958), 203–207.

[7] G. GLAESER, H. STACHEL, B. ODEHNAL, *The Universe of Conics*, Springer Spectrum, Berlin-Heidelberg, 2016.

[8] Á. G. HORVÁTH, H. MARTINI, Conics in normed planes, *Extracta Math.* **26**/1 (2011), 29–43.

[9] Á. G. HORVÁTH, Constructive curves in non-Euclidean planes, *19th Scientific-Professional Colloquium on Geometry and Graphics*, Starigrad-Paklenica, 2016.

[10] E. MOLNÁR, Kegelschnitte auf der metrischen Ebene, *Acta Math. Acad. Sci. Hungar.* **31**/3-4 (1978), 317–343.

[11] U. RICHTER-GEBERT, U. H. KORTENKAMP, Cinderella, `https://plus.maths.org/content/interactive-geometry-software-cinderella-version-12`.

[12] A. SLIEPČEVIĆ, I. BOŽIĆ, H. HALAS, Introduction to the Planimetry of the Quasi-Hyperbolic Plane,*KoG* **17** (2013), 58–64.

[13] G. WEISS, F. GRUBER, Den Satz von Thales verallgemeinern - aber wie?, *KoG* **12** (2008), 7–18.

[14] G. WEISS, Elementary Constructions for Conics in Hyperbolic and Elliptic Planes, *KoG* **19** (2015), 24–31.

[15] N. J. WILDBERGER, Universal Hyperbolic Geometry I: Trigonometry, *Geometriae Dedicata* **163**/1 (2013), 215–274.

Gunter Weiss
e-mail: weissgunter@hotmail.com

University of Technology Vienna,
Karlsplatz 13, 1040 Vienna, Austria

University of Technology Dresden,
Helmholtzstr. 10, 01069 Dresden, Germany

Original scientific paper
Accepted 5. 12. 2016.

SEBASTIAN BLEFARI
N J WILDBERGER

Quadrangle Centroids in Universal Hyperbolic Geometry

Quadrangle Centroids in Universal Hyperbolic Geometry

ABSTRACT

We study relations between the eight projective quadrangle centroids of a quadrangle in universal hyperbolic geometry which are analogs of the barycentric centre of a Euclidean quadrangle. We investigate the number theoretical conditions for such centres to exist, and show that the eight centroids naturally form two quadrangles which together with the original one have three-fold perspective symmetries. The diagonal triangles of these three quadrangles are also triply perspective.

Key words: Universal hyperbolic geometry, projective geometry, centroids, quadrangles, diagonal triangles, perspectivities

MSC2010: 51M05, 51M10, 51N10

Težišta četverokuta u univerzalnoj hiperboličnoj geometriji

SAŽETAK

Promatramo veze između osam projektivnih težišta četverokuta u univerzalnoj hiperboličnoj geometriji koji su analogoni baricentričnom središtu euklidskog četverokuta. Određujemo teoretske uvjete postojanja tih središta i pokazujemo da osam težišta tvore dva četverokuta koji zajedno s danim četverokutom imaju trostruku perspektivnu simetriju. I dijagonalni trokuti ovih četverokuta su trostruko perspektivni.

Ključne riječi: Univerzalna hiperbolična geometrija, projektivna geometrija, težišta, četverovrsi, dijagonalni trokuti, perspektiviteti

1 Introduction: classical theories of the centroid

In this paper we investigate quadrangle centroids in the general setting of universal hyperbolic geometry (UHG) using the novel algebraic orientation of *standard quadrangle coordinates*. We will associate to a hyperbolic quadrangle with midpoints two other perspectively related *quadrangles of centroids* as in Figure 1, and will identify a number of perspective relations between these quadrangles, and also of their associated diagonal triangles.

Recall that the three median lines of a Euclidean or affine triangle $\overline{ABC}$ are concurrent at the **centroid** G**,** which can be viewed as either the centre of mass of the three points of the triangle, or as the centre of mass of a uniform lamina, or mass distribution, on the triangle – the two notions coincide.

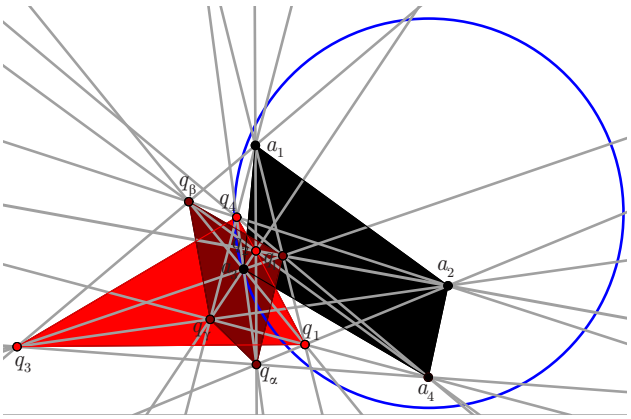


Figure 1: *Three perspectively related quadrangles in UHG*

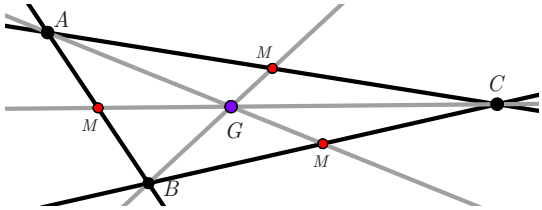


Figure 2: *The centroid of a triangle*

For a Euclidean or affine quadrangle the story is a bit more subtle; there are two notable centroid points. For simplicity let's assume temporarily that the Euclidean quadrangle $\square \equiv \overline{ABCD}$ is convex, and consider the four subtriangles

$$\triangle_A \equiv BCD, \quad \triangle_B \equiv ACD, \quad \triangle_C \equiv ABD \quad \text{and} \quad \triangle_D \equiv ABC$$

with respective centroids G_A, G_B, G_C, and G_D. If we consider the quadrangle $\square$ to have a uniformly distributed mass, then the **type I centroid** G_1 is the point at which this mass balances. The lines G_AG_C and G_BG_D are both lines of balance of the uniform mass distribution, so their meet

$$G_1 = (G_AG_C)\,(G_BG_D)$$

is also a point of balance, as in Figure 3.

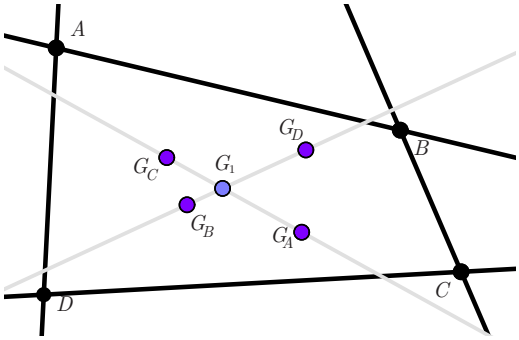


Figure 3: *The type 1 centroid of a quadrangle*

On the other hand the **type II centroid**, or **barycenter**, G_2 is the center of mass of the four vertices. This is the common meet of the lines of balance AG_A, BG_B, CG_C and DG_D as in Figure 4.

The barycenter G_2 can also be found as the meet of the three **bimedian lines**, which are the joins of the midpoints of opposite sides, which we write as the triple concurrence

$$G_2 = (M_{AB}M_{CD})\,(M_{AD}M_{BC})\,(M_{AC}M_{BD})$$

as in Figure 5.

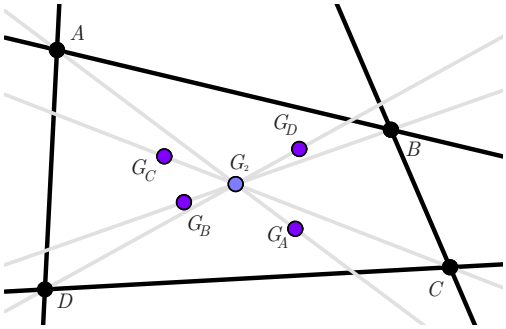


Figure 4: *The type II centroid or barycenter of a quadrangle*

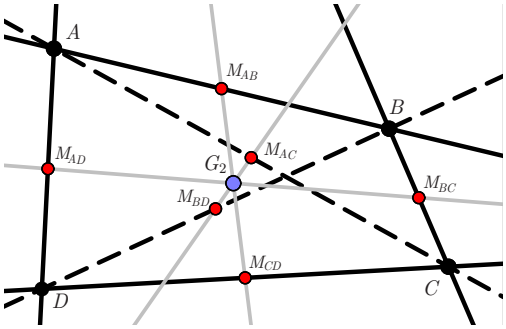


Figure 5: *Bimedian line construction of type II centroid*

Both these ideas extend to non-convex shapes, but on account of the dependence on the order of the points, there are two other possible type I centroids, namely the points $(G_AG_B)\,(G_CG_D)$ and $(G_AG_D)\,(G_BG_C)$, depending on the quadrangle's orientation. In contrast, the barycenter G_2 has a purely affine nature and is uniquely defined even in the non-convex situation.

How do these notions extend to hyperbolic, elliptic and other non-Euclidean geometries? Both the books of Sommerville [13] and Coolidge [5], after constructing a projective metric in a Cayley Klein geometry (as in [10], [11]), discuss how a side has two midpoints, a three-point system has four centroids, and finally that a four-point system has eight centroids. The absence of proofs suggests that this was a reasonably well-known 19th century or early 20th century configuration.

UHG is a broad algebraic generalization of hyperbolic geometry which stems from rational trigonometry ([15]) and is related to, but distinctly different from, the more familiar Cayley Klein geometry; and since we want this paper to be largely self contained we will review this; the reader can refer to papers [16], [17], [18], [19] and [20] for further details. We then study quadrangle centroids using the purely algebraic approach of UHG, which works over general fields as well as arbitrary non-degenerate symmetric

bilinear forms, and allows us to give concrete computational proofs of results. The generality of this theory is actually a powerful aspect to the arguments that we employ, which require specific transformations to place particular points in standard positions to simplify the algebra, and then the structure constants of the general bilinear form become ingredients in our formulas.

In this general situation there are either zero, one or two midpoints of a side [4], and which case it is depends on number theoretic considerations. So we need to understand the explicit algebraic consequences of the assumptions that ensure existence of quadrilateral centroids. This is highly dependent on the underlying field, so things becomes clearer if we do not assume an algebraically closed field, where such subtleties are largely lost, and indeed where there are serious logical hurdles. So our results are meaningful over, for example, the rational numbers, and also over finite fields, where additional combinatorial aspects arise that are largely invisible to classical synthetic geometry.

Our goal is to prove theorems using explicit general formulas that may help future researchers to make further algebraic explorations in as wide a context as possible. While hyperbolic triangles have seen some recent study, (see [14], [19] and [20]), opportunities for investigation with hyperbolic quadrangles are also rich.

2 Universal hyperbolic geometry and quadrangles

We will now review briefly the projective metrical framework of universal hyperbolic geometry (see [17], [20]) and introduce basic notation for triangles and quadrangles. The general projective linear algebraic setting covers both elliptic, hyperbolic and more general metrical geometries, and works over a general field, not of characteristic two. UHG shares the underlying framework of Cayley Klein geometry—a projective plane with a distinguished conic or its algebraic equivalent, namely a symmetric bilinear form. But it utilises purely algebraic metrical notions, namely *projective quadrance* and *spread,* instead of *hyperbolic distance* and *angle*.

Projective quadrance and spread can be introduced also in a projective framework using cross ratios as described in [17], but we will frame them in the context of a three-dimensional (affine) vector space over a field not of characteristic two. To differentiate between affine and projective linear algebra we will use the convention of writing familiar affine vectors and matrices with round brackets, and projective vectors and matrices, which are determined only up to a (non-zero) multiple, with square brackets.

A **(projective) point** is a *non-zero* projective row vector a shown in either of two ways:

$$a = \begin{bmatrix} x & y & z \end{bmatrix} \equiv [x : y : z]$$

while a **(projective) line** is a *non-zero* projective column vector L shown in either of two ways:

$$L = \begin{bmatrix} l \\ m \\ n \end{bmatrix} \equiv \langle l : m : n \rangle.$$

The point $a = [x : y : z]$ and line $L = \langle l : m : n \rangle$ are **incident** precisely when

$$0 = aL \equiv \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} l \\ m \\ n \end{bmatrix} = [x : y : z] \langle l : m : n \rangle = lx + my + nz.$$

Note that the product of two projective matrices is only determined up to scalars, but in particular the condition of having a zero product is well-defined.

The **join** a_1a_2 of distinct points $a_1 \equiv [x_1 : y_1 : z_1]$ and $a_2 \equiv [x_2 : y_2 : z_2]$ is then the line

$$\begin{aligned} a_1a_2 &\equiv [x_1 : y_1 : z_1] \times [x_2 : y_2 : z_2] \\ &= \langle y_1z_2 - z_1y_2 : z_1x_2 - x_1z_2 : x_1y_2 - y_1x_2 \rangle \end{aligned}$$

and the **meet** L_1L_2 of two distinct lines $L_1 = \langle l_1 : m_1 : n_1 \rangle$ and $L_2 = \langle l_2 : m_2 : n_2 \rangle$ is the point

$$\begin{aligned} L_1L_2 &\equiv \langle l_1 : m_1 : n_1 \rangle \times \langle l_2 : m_2 : n_2 \rangle \\ &= [m_1n_2 - n_1m_2 : n_1l_2 - l_1n_2 : l_1m_2 - m_1l_2]. \end{aligned}$$

The cross product here is the usual Euclidean cross product, which is well-defined on projective points and lines.

We'll say a **side** $\overline{a_1a_2} = \{a_1, a_2\}$ is a set of two points; a **vertex** $\overline{L_1L_2} = \{L_1, L_2\}$ is a set of two lines; a **triangle** $\overline{a_1a_2a_3} \equiv \{a_1, a_2, a_3\}$ is a set of three non-collinear points; and a **quadrangle** $\overline{a_1a_2a_3a_4} \equiv \{a_1, a_2, a_3, a_4\}$ is a set of four points, where no three are collinear. The quadrangle $\overline{a_1a_2a_3a_4} \equiv \{a_1, a_2, a_3, a_4\}$ has six distinct sides $\overline{a_1a_2}$, $\overline{a_3a_4}$, $\overline{a_1a_3}$, $\overline{a_2a_4}$, $\overline{a_1a_4}$, and $\overline{a_2a_3}$, with corresponding lines $L_{12} \equiv a_1a_2$, $L_{34} \equiv a_3a_4$, $L_{13} \equiv a_1a_3$, $L_{24} \equiv a_2a_4$, $L_{14} \equiv a_1a_4$, and $L_{23} \equiv a_2a_3$.

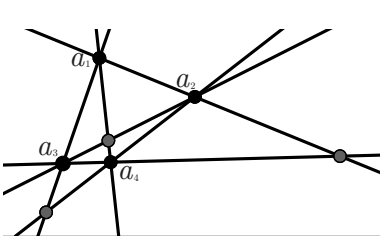


Figure 6: *A projective quadrangle and its diagonal points*

The four **subtriangles** of the quadrangle $\overline{a_1a_2a_3a_4}$ are

$$\triangle_1 \equiv \overline{a_2a_3a_4}, \quad \triangle_2 \equiv \overline{a_1a_3a_4},$$
$$\triangle_3 \equiv \overline{a_1a_2a_4}, \quad \triangle_4 \equiv \overline{a_1a_2a_3}.$$

Two sides of a quadrangle are **opposite** if they have no point in common. Each of the six sides of the quadrangle are contained in exactly two distinct subtriangles, while opposite sides are precisely the pairs of sides which don't belong to any one of the subtriangles.

The **diagonal points** of the quadrangle $\overline{a_1a_2a_3a_4}$ are the meets of the associated lines of opposite sides, that is

$$d_\alpha \equiv L_{12}L_{34}, \quad d_\beta \equiv L_{13}L_{24} \quad \text{and} \quad d_\gamma \equiv L_{14}L_{23}. \tag{1}$$

Then $\overline{d_\alpha d_\beta d_\gamma}$ is the **diagonal triangle** of the standard quadrangle. The indices α, β and γ correspond to the three ways of dividing four points into two pairs, these play a distinguished role in the subject.

3 Projective bilinear forms and metrical structure

We now impose metrical structure through a general symmetric bilinear form, possibly non-Euclidean, on an underlying three-dimensional vector space, as studied for example in [12]. This generalises the notion of a dot product, and is a standard notion in projective geometry, where a bilinear form can be used to define objects such as conics sections and quadrics, as described for example in [21] and [11]. The bilinear form induces metrical structure to the associated projective space, which can be viewed as one-dimensional subspaces of the vector space. This gives a far-reaching methodology for imbedding measurement into purely algebraic geometry. For a given affine vector v we will denote the associated projective vector by $a = [v] = [\lambda v]$ where $\lambda \neq 0$, and similarly for an affine matrix A we denote the associated projective matrix as $\mathbf{A} = [A] = [\mu A]$ where $\mu \neq 0$.

Let us fix a general invertible symmetric 3×3 matrix A and its adjugate B in the general form

$$A = \begin{pmatrix} a & d & f \\ d & b & g \\ f & g & c \end{pmatrix} \quad \text{and}$$
$$B = \begin{pmatrix} g^2 - bc & dc - fg & fb - dg \\ dc - fg & f^2 - ac & ga - df \\ fb - dg & ga - df & d^2 - ab \end{pmatrix} \tag{2}$$

and denote by $\mathbf{A}$ and $\mathbf{B}$ the associated symmetric projective matrices. By assumption the determinant

$$D \equiv \det A = abc + 2fdg - ag^2 - bf^2 - cd^2$$

is non-zero. The **dual** of a projective point a is then defined to be the line $a^\perp \equiv \mathbf{A}a^T$, and the **dual** of the projective line L is the point $L^\perp \equiv L^T\mathbf{B}$. Since the projective matrices $\mathbf{A}$ and $\mathbf{B}$ are inverses, and both are symmetric,

$$(aL)^T = (a\mathbf{A}\mathbf{B}L)^T = L^T\mathbf{B}\mathbf{A}a^T = L^\perp a^\perp,$$

and so the point a is incident with the line L precisely when their duals are incident, and in addition $\left(a^\perp\right)^\perp = a$ and $\left(L^\perp\right)^\perp = L$. Hence duality preserves projective theorems.

Two points a_1 and a_2 are **perpendicular,** written as $a_1 \perp a_2$, precisely when one is incident with the dual of the other, that is when $0 = a_1a_2^\perp = a_1\mathbf{A}a_2^T$. Symmetrically two lines L_1 and L_2 are **perpendicular,** written as $L_1 \perp L_2$, precisely when one is incident with the dual of the other, that is when $0 = L_1^\perp L_2 = L_1^T\mathbf{B}L_2$.

A point a is **null** precisely when it is self perpendicular, and a line L is **null** precisely when it is self perpendicular, that is when respectively

$$a\mathbf{A}a^T = 0 \quad \text{and} \quad L^T\mathbf{B}L = 0.$$

The **null conic** (or **absolute**) then consists of null points, and the null lines are the tangents to this conic. The above notions of perpendicularity and duality are symmetric, and algebraically capture the synthetic notion of polarity with respect to the null conic in the projective plane.

The standard cases in UHG are those of *hyperbolic* and *elliptic geometries* [19], which arise respectively from the choices

$$\mathbf{A} = \mathbf{J} \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \mathbf{I} \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

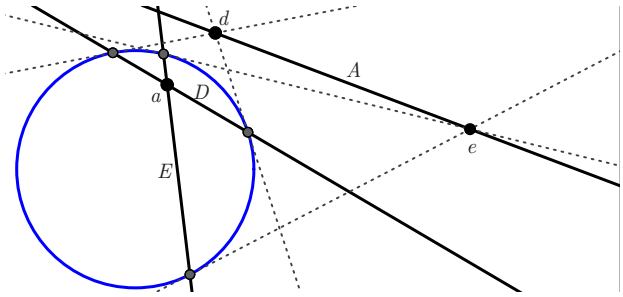


Figure 7: *Polars of points and lines, interior and exterior*

Figures in this paper come from the hyperbolic situation where the null conic has equation $x^2 + y^2 - z^2 = 0$, which meets the viewing plane $z = 1$ in the usual unit circle $x^2 + y^2 = 1$ shown in blue in Figure 7. In this Figure we see that the duals of the points a, d and e are the lines A, D and E respectively, so this is just the classical pole/polar

duality with respect to the circle. The points a and d are perpendicular, as are a and e, and so are the dual lines.

The inverse pair of symmetric projective matrices **A** and **B** coming from (2) that determine perpendicularity and duality in UHG also allow us to define the **(projective) quadrance** $q(a_1,a_2)$ between the two points a_1 and a_2, and the **(projective) spread** $S(L_1,L_2)$ between the two lines L_1 and L_2 as the respective quantities

$$q(a_1,a_2) \equiv 1 - \frac{\left(a_1\mathbf{A}a_2^T\right)^2}{\left(a_1\mathbf{A}a_1^T\right)\left(a_2\mathbf{A}a_2^T\right)} \qquad \text{and}$$

$$S(L_1,L_2) \equiv 1 - \frac{\left(L_1^T\mathbf{B}L_2\right)^2}{\left(L_1^T\mathbf{B}L_1\right)\left(L_2^T\mathbf{B}L_2\right)}.$$

By homogeneity these are well-defined numbers, even if the various ingredients are only defined up to non-zero multiplicative scalars. It should be noted that in rational trigonometry, the notions of quadrance and spread are affine versions of the above definitions: but they are different.

In the special case of hyperbolic geometry, these metrical measurements are closely related to the hyperbolic distance $d(a_1,a_2)$ and the angle $\theta(L_1,L_2)$ as described in [17] provided we restrict our attention to the interior of the null conic, and the relations are

$$q(a_1,a_2) = -\sinh^2(d(a_1,a_2)),$$
$$S(L_1,L_2) = \sin^2(\theta(L_1,L_2)).$$

With the algebraic orientation of UHG, the avoidance of all transcendental functions means that calculations can be performed with full precision, that we can work over a general field, that quadrance and spread really mirror the fundamental projective duality between points and lines, and that metrical theorems extend to arbitrary non-degenerate symmetric bilinear forms. These are very serious advantages! As explained in [17] and [18], projective quadrance and spread may also be defined in terms of cross ratios, independent of the relative positions of the points and lines with respect to the null conic. This is also related to the approach of Brauner in [3].

A **hyperbolic circle** with fixed center a and quadrance k is the locus of points x which satisfy $q(a,x) = k$. This is a conic. In Figure 8 we see circles centered at the external point a and their quadrances; these are conics which are tangential to the null conic and include what in the classical literature are called "equi-distant curves". This notation is not really necessary, as these conics are just circles.

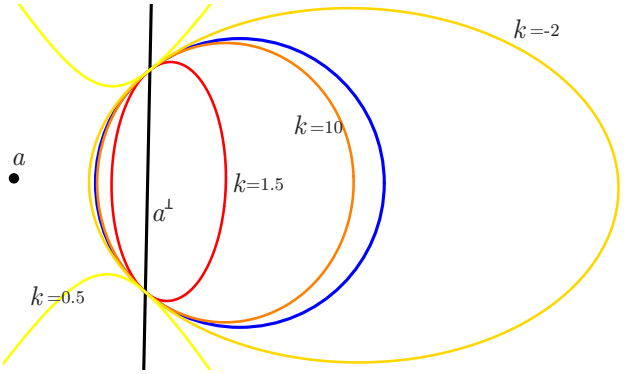


Figure 8: *Hyperbolic circles with center a and different quadratnces*

4 Midpoints

As defined in [17], the point m is a **midpoint** of the side $\overline{a_1a_2}$ precisely when m is a point incident with the line a_1a_2 which satisfies

$$q(a_1,m) = q(m,a_2).$$

A **midline** M of the side $\overline{a_1a_2}$ is a line passing through a midpoint m which is perpendicular to a_1a_2. In Figure 9 we see two midpoints of the side $\overline{a_1a_2}$, and a synthetic construction of them, along with the associated midlines. Note that such a construction does not always work: the relative positions of the side $\overline{a_1a_2}$ and the null conic determine when this happens.

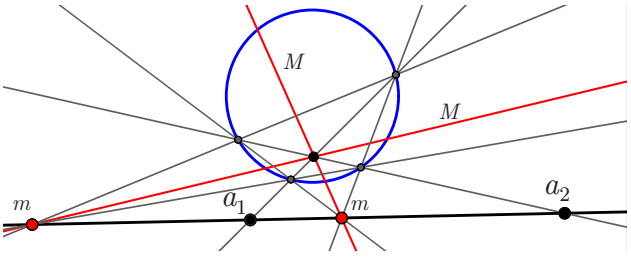


Figure 9: *Midpoints and midlines of a side*

The following theorem was given by Wildberger [17] in the purely hyperbolic case. Since midpoints are at the heart of this paper, we give a more general proof which applies to an arbitrary bilinear form.

Theorem 1 (Side midpoints) *Suppose that a_1 and a_2 are non-null, non-perpendicular points, forming a non-null side $\overline{a_1a_2}$. Then $\overline{a_1a_2}$ has a non-null midpoint m precisely when $1-q(a_1,a_2)$ is a square (within the field), and in this case there are exactly* two *perpendicular midpoints m.*

Proof. We prove the theorem first for specific points and a general bilinear form. Let $a_1 \equiv [u_1] = [1:1:1]$ and $a_2 \equiv [u_2] = [-1:-1:1]$ where $u_1 \equiv (1,1,1)$ and $u_2 \equiv (-1,-1,1)$ are affine row vectors, and the general bilinear form be given by the invertible pair of symmetric projective matrices $\mathbf{A}$ and $\mathbf{B}$ with affine representatives given by (2). By the definition of quadrance

$$1 - q(a_1,a_2) = \frac{\left(a_1\mathbf{A}a_2^T\right)^2}{\left(a_1\mathbf{A}a_1^T\right)\left(a_2\mathbf{A}a_2^T\right)} = \frac{(c-b-a-2d)^2}{A_1A_2}$$

where

$$A_1 \equiv u_1Au_1^T = a+b+c+2(d+f+g)$$
$$A_2 \equiv u_2Au_2^T = a+b+c+2(d-f-g).$$

Since a_1 and a_2 are non-null points, each of A_1 and A_2 are nonzero. An arbitrary point m on the line $a_1a_2 \equiv \langle 1:-1:0\rangle$ has the form $m = [x-y:x-y:x+y]$, which is null precisely when

$$m\mathbf{A}m^T = (a+b+2d)(x-y)^2 + c(x+y)^2 + 2(f+g)(x^2-y^2) = 0$$

in which case the quadrance between m and any other point is undefined. Thus we assume that m is non-null, which gives the quadrances

$$q(a_1,m) = \frac{4y^2\left(c(a+b+d)-(f+g)^2\right)}{A_1\left((a+b+2d)(x-y)^2 + c(x+y)^2 + 2(f+g)(x^2-y^2)\right)}$$
$$q(a_2,m) = \frac{4x^2\left(c(a+b+d)-(f+g)^2\right)}{A_2\left((a+b+2d)(x-y)^2 + c(x+y)^2 + 2(f+g)(x^2-y^2)\right)}.$$

By assumption $\overline{a_1a_2}$ is non-null, and so the above expressions are equal precisely when $y^2A_2 = x^2A_1$ or

$$\frac{1}{A_1A_2} = \frac{y^2}{A_1^2x^2}.$$

This occurs precisely when $1-q(a_1,a_2)$ is a square. In such a case if we define a number σ_{12} by

$$\frac{1}{A_1A_2} = \sigma_{12}^2$$

then we get solutions $x = 1$ and $y = \pm\sigma_{12}A_1$, so the midpoints are

$$m \equiv [u_1 \pm \sigma_{12}A_1 \cdot u_2] = [1 \pm \sigma_{12}A_1 : 1 \pm \sigma_{12}A_1 : 1 \mp \sigma_{12}A_1].$$

These two points are perpendicular since $m_+\mathbf{A}m_-^T = 0$ by a computation.

Despite the proof being for a specific side, the general bilinear form allows us to use the *Fundamental theorem of projective geometry* to transform this side to a general side with any specific bilinear form. This gives us the result. □

What we have shown in the proof is that for points a_1 and a_2 if we can find affine vectors v and u such that $a_1 = [v]$, $a_2 = [u]$ and $vAv^T = uAu^T$, then the midpoints are $[v+u]$ and $[v-u]$. This is a particularly useful observation.

We point out an important variant of midpoint introduced in [20]: the point s is a **sydpoint** of the side $\overline{a_1a_2}$ precisely when s is a point incident with the line a_1a_2 which satisfies

$$q(a_1,s) = -q(s,a_2).$$

Alkhaldi and Wildberger have shown that the side $\overline{a_1a_2}$ has *sydpoints* precisely when the number $q(a_1,a_2)-1$ is a square. Thus over the rationals a side approximately has two midpoints or two sydpoints [20]. Over a finite field things can behave rather differently, depending on the field. For example over $\mathbb{F}_5$ a side has midpoints precisely when it has sydpoints, as the only squares are 1 and -1. It is remarkable that virtually everything in this paper can be extended to the situation where the sides of a quadrangle have either midpoints or sydpoints, but we will leave that to a future discussion.

4.1 Centroids and Circumcenters

We now review familiar facts about medians and centroids of triangles from Cayley Klein geometry and classical hyperbolic geometry, as described for example in [13] and [6], but in the setting of UHG. This way we get a seamless extension of these notions also to points outside the traditional hyperbolic disk; in Cayley Klein geometry it can be awkward to smoothly make this extension on account of the more limited nature of distance and angle, and their dependence on transcendental functions.

We assume that we have a (hyperbolic) triangle $\overline{a_1a_2a_3}$ all of whose sides have midpoints m as in Figure 10.

The **median lines** (or just **medians**) D of $\overline{a_1a_2a_3}$ are then the joins of corresponding midpoints of sides with opposite points of the triangle. There are six medians, two passing through each point a_i. The following results are well-known, although the terminology is somewhat novel, see for example the above references and [19].

Theorem 2 (Centroids) *The median lines D of the triangle $\overline{a_1a_2a_3}$ are concurrent three at a time, meeting at four distinct* ***centroid points*** *g.*

We note that each centroid is associated with a distinct set of three midpoints used to construct it, one from each side.

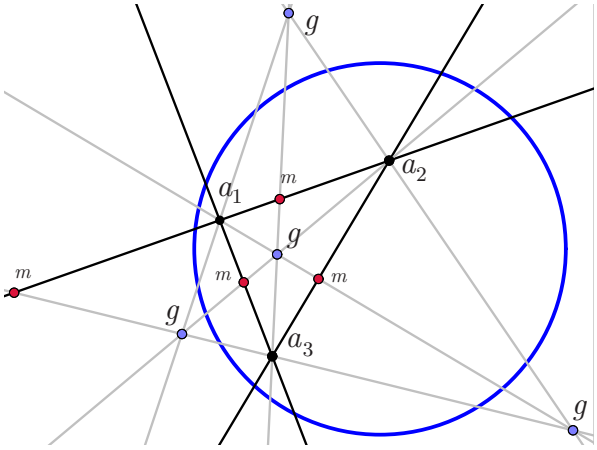


Figure 10: *Centroids of a triangle* $\overline{a_1a_2a_3}$

Theorem 3 (Circumcenters) *The six midpoints m of the triangle* $\overline{a_1a_2a_3}$ *are collinear three at a time, lying on four distinct* ***circumlines*** *C. The six midlines M of* $\overline{a_1a_2a_3}$ *are concurrent three at a time, meeting at four distinct* ***circumcenters*** *c which are dual to the circumlines C. The circumcenters are the centers of the four hyperbolic circles which go through the points of the triangle.*

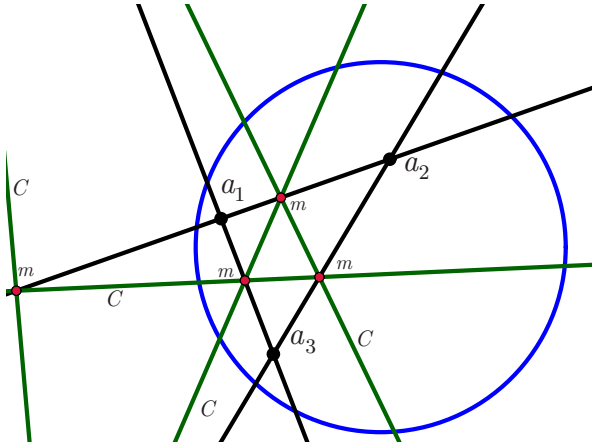


Figure 11: *Circumlines of a triangle* $\overline{a_1a_2a_3}$

The centroid and circumcenter configurations are closely related through perspectivities. As noted in [5], the median/centroid configuration is that of a quadrangle, as there are four points, and six lines concurrent at the points in threes, giving a Desargues configuration.

The quadrangle of centroids has the original triangle as its diagonal triangle. So if we look at a subtriangle of this centroid quadrangle, then it is perspective with the original triangle at the fourth centroid point, as illustrated in Figure 12.

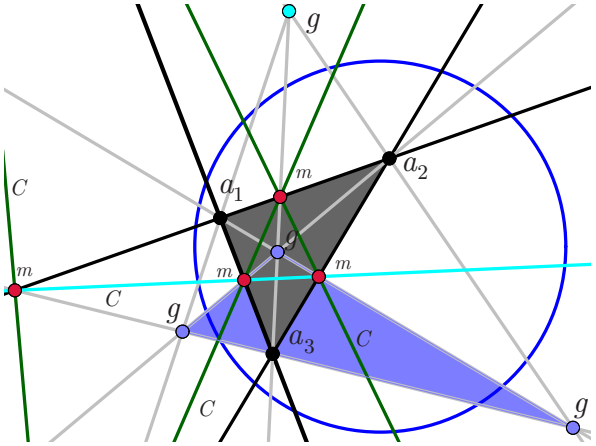


Figure 12: *The triangle* $\overline{a_1a_2a_3}$ *is perspective with a subtriangle of centroids from a Circumline C.*

From Desargues' theorem we know that these two triangles are perspective from a line which is precisely a circumline. This can be done for each centroid, giving a unique correspondence between centroids and circumcenters (through the circumlines). In this paper we will concentrate on the centroid story, as it is quite rich and interesting in its own right, and leave the circumcenter picture for another occasion.

5 Standard quadrangles

The fundamental theorem of projective geometry states that *any two projective frames in a projective plane can be mapped onto each other by exactly one invertible projective linear transformation*. This allows us to transform the study of a general quadrangle under a fixed bilinear form to that of a fixed standard quadrangle with a general bilinear form. This powerful technique for tackling metrical structure was employed in [19] and [20] in the context of hyperbolic triangle geometry, in [1] and [2] to study hyperbolic parabolas, and used in the papers [9] and [8], as well as the thesis of Nguyen Le [7] in the Euclidean case.

In this paper the **standard quadrangle** $\square \equiv \overline{a_1a_2a_3a_4}$ is given by the four points

$$a_1 \equiv [1:1:1], \quad a_2 \equiv [-1:-1:1],$$
$$a_3 \equiv [1:-1:1], \quad a_4 \equiv [-1:1:1]$$

with corresponding affine vectors

$$u_1 \equiv (1,1,1), \quad u_2 \equiv (-1,-1,1),$$
$$u_3 \equiv (1,-1,1), \quad u_4 \equiv (-1,1,1).$$

After a suitable projective transformation, any given quadrangle can be transformed to the standard quadrangle, and since the metrical structure will change correspondingly by

a congruence, we may assume that this is then given by the general pair of matrices A and B of (2). We will also use a consistent notational convention throughout this paper, which will emphasize the quadrangle point(s) used to construct an object, so that for example whenever an object has a 1 in its label, it relates to the point a_1 in some way.

We make some assumptions to avoid degenerate situations. The first is that none of the points a_1, a_2, a_3, a_4 or lines

$$L_{12} \equiv a_1a_2 = \langle 1 : -1 : 0\rangle, \quad L_{13} \equiv a_1a_3 = \langle 1 : 0 : -1\rangle,$$
$$L_{14} \equiv a_1a_4 = \langle 0 : 1 : -1\rangle, \quad L_{34} \equiv a_3a_4 = \langle 1 : 1 : 0\rangle,$$
$$L_{24} \equiv a_2a_4 = \langle 1 : 0 : 1\rangle, \quad L_{23} \equiv a_2a_3 = \langle 0 : 1 : 1\rangle$$

of $\square$ are null. The second is that the four points of the quadrangle *do not lie on a single hyperbolic circle*. The reason for this will be explained in the last section.

Let us also record that the diagonal points of the quadrangle are

$$d_\alpha \equiv L_{12}L_{34} = [0 : 0 : 1], \quad d_\beta \equiv L_{13}L_{24} = [0 : 1 : 0],$$
$$d_\gamma \equiv L_{14}L_{23} = [1 : 0 : 0]. \tag{3}$$

It will be useful to define the numbers

$$A_1 \equiv u_1Au_1^T = a+b+c+2(d+f+g),$$
$$A_2 \equiv u_2Au_2^T = a+b+c+2(d-f-g),$$
$$A_3 \equiv u_3Au_3^T = a+b+c+2(-d+f-g),$$
$$A_4 \equiv u_4Au_4^T = a+b+c+2(-d-f+g)$$

which by assumption are all non-zero. We are going to express things in terms of the entries a, b, c, d, f and g of the matrix A of (2) which determines the metrical structure.

Theorem 4 (Quadrangle quadrances) *The quadrances of the standard quadrangle $\overline{a_1a_2a_3a_4}$ are*

$$q(a_1,a_2) = 4\frac{c(a+b+2d)-(f+g)^2}{A_1A_2},$$
$$q(a_3,a_4) = 4\frac{c(a+b-2d)-(f-g)^2}{A_3A_4},$$
$$q(a_1,a_3) = 4\frac{b(a+c+2f)-(d+g)^2}{A_1A_3},$$
$$q(a_2,a_4) = 4\frac{b(a+c-2f)-(d-g)^2}{A_2A_4},$$
$$q(a_1,a_4) = 4\frac{a(b+c+2g)-(d+f)^2}{A_1A_4},$$
$$q(a_2,a_3) = 4\frac{a(b+c-2g)-(d-f)^2}{A_2A_3}.$$

These satisfy

$$1-q(a_1,a_2) = \frac{(c-b-a-2d)^2}{A_1A_2},$$
$$1-q(a_3,a_4) = \frac{(c-b-a+2d)^2}{A_3A_4},$$
$$1-q(a_1,a_3) = \frac{(a-b+c+2f)^2}{A_1A_3},$$
$$1-q(a_2,a_4) = \frac{(a-b+c-2f)^2}{A_2A_4},$$
$$1-q(a_1,a_4) = \frac{(b-a+c+2g)^2}{A_1A_4},$$
$$1-q(a_2,a_3) = \frac{(b-a+c-2g)^2}{A_2A_3}.$$

Proof. Computations yield these results. $\square$

5.1 Conditions for midpoints, or sigma relations

Now we also want to impose the conditions that ensure that all six sides $\overline{a_ia_j}$ for $i \neq j$ have midpoints, so that each of the subtriangles of $\overline{a_1a_2a_3a_4}$ has centroids. To find elegant proofs of our theorems we will require some careful bookkeeping with regards to the labelling of midpoints. There is an interesting combinatorial aspect to this.

From the Side midpoint theorem we know that the non-null side $\overline{a_ia_j}$ has midpoints precisely when $1-q(a_i,a_j)$ is a square. From the second half of the previous theorem, the condition for all six sides having midpoints is equivalent to the existence of six non-zero **sigma values**

$$\sigma_{12}, \sigma_{34}, \sigma_{13}, \sigma_{24}, \sigma_{14}, \sigma_{23}$$

in the chosen field, satisfying the following **quadratic relations**

$$\frac{1}{A_1A_2} = \sigma_{12}^2, \quad \frac{1}{A_1A_3} = \sigma_{13}^2, \quad \frac{1}{A_1A_4} = \sigma_{14}^2,$$
$$\frac{1}{A_3A_4} = \sigma_{34}^2, \quad \frac{1}{A_2A_4} = \sigma_{24}^2, \quad \frac{1}{A_2A_3} = \sigma_{23}^2.$$

We can further take the product of these quadratic relations in threes, say

$$\frac{1}{A_1A_2} = \sigma_{12}^2, \quad \frac{1}{A_1A_3} = \sigma_{13}^2 \quad \text{and} \quad \frac{1}{A_2A_3} = \sigma_{23}^2$$

and by possibly changing the sign of any or all of the *sigma values* to arrange the following four **cubic relations**

$$\frac{1}{A_1A_2A_3} = \sigma_{12}\sigma_{23}\sigma_{13} \equiv \sigma_4, \quad \frac{1}{A_1A_2A_4} = \sigma_{12}\sigma_{24}\sigma_{14} \equiv \sigma_3,$$
$$\frac{1}{A_1A_3A_4} = \sigma_{13}\sigma_{34}\sigma_{14} \equiv \sigma_2, \quad \frac{1}{A_2A_3A_4} = \sigma_{23}\sigma_{34}\sigma_{24} \equiv \sigma_1.$$

There are three ways of expressing each A_i using the quadratic and cubic relations, given by

$$A_1 = \frac{\sigma_{23}}{\sigma_{12}\sigma_{13}} = \frac{\sigma_{24}}{\sigma_{12}\sigma_{14}} = \frac{\sigma_{34}}{\sigma_{13}\sigma_{14}},$$
$$A_2 = \frac{\sigma_{13}}{\sigma_{12}\sigma_{23}} = \frac{\sigma_{14}}{\sigma_{12}\sigma_{24}} = \frac{\sigma_{34}}{\sigma_{23}\sigma_{24}},$$
$$A_3 = \frac{\sigma_{12}}{\sigma_{13}\sigma_{23}} = \frac{\sigma_{14}}{\sigma_{13}\sigma_{34}} = \frac{\sigma_{24}}{\sigma_{23}\sigma_{34}},$$
$$A_4 = \frac{\sigma_{12}}{\sigma_{14}\sigma_{24}} = \frac{\sigma_{13}}{\sigma_{14}\sigma_{34}} = \frac{\sigma_{23}}{\sigma_{24}\sigma_{34}}.$$

These show how *each point is in exactly three subtriangles*. For example A_1 has three representations, each relating to the subtriangles $\triangle_2$, $\triangle_3$ and $\triangle_4$ which contain a_1.

From this point on most (if not all) of the expressions and formulas involve these sigma values and relations defined in this section, so it is valuable to appreciate the symmetries involved, which boil down to the following relations

$$\frac{\sigma_{23}}{\sigma_{13}} = \frac{\sigma_{24}}{\sigma_{14}}, \quad \frac{\sigma_{23}}{\sigma_{12}} = \frac{\sigma_{34}}{\sigma_{14}}, \quad \frac{\sigma_{24}}{\sigma_{12}} = \frac{\sigma_{34}}{\sigma_{13}}$$

or simply

$$\sigma_{12}\sigma_{34} = \sigma_{13}\sigma_{24} = \sigma_{14}\sigma_{23} \equiv \eta.$$

Let's also introduce

$$\sigma_{234} \equiv \sigma_{12}\sigma_{13}\sigma_{14}, \quad \sigma_{134} \equiv \sigma_{12}\sigma_{23}\sigma_{24},$$
$$\sigma_{124} \equiv \sigma_{13}\sigma_{23}\sigma_{34}, \quad \sigma_{123} \equiv \sigma_{14}\sigma_{24}\sigma_{34}.$$

5.2 Midpoints of the quadrangle

By the Side midpoint theorem, for a side $\overline{a_i a_j}$ with midpoints, where $a_i = [v_i]$ and $a_j = [v_j]$, we are able to normalize the affine vectors v_i and v_j such that the midpoints have the simple forms

$$m^{(ij)} \equiv [v_i + v_j] \quad \text{and} \quad m^{(ji)} \equiv [v_i - v_j].$$

Despite the arbitrariness of the ordering we will use the convention that the side corresponding to the points a_i and a_j for $i < j$ will be given as $\overline{a_i a_j}$ and *not as* $\overline{a_j a_i}$, and so the midpoint $m^{(ij)} = [v_i + v_j]$ can be seen as having **positive** (or **ascending**) **orientation** and the midpoint $m^{(ji)} = [v_i - v_j]$ as having **negative** (or **descending) orientation**.

In the end of the proof of the Side midpoints theorem we saw that the midpoints for the side $\overline{a_1 a_2}$ could also be written as

$$m \equiv [v_1 \pm \sigma_{12} A_1 v_2] = [1 \pm \sigma_{12} A_1 : 1 \pm \sigma_{12} A_1 : 1 \mp \sigma_{12} A_1].$$

In light of the above discussion, we'll say that $m^{(12)} = [v_1 + \sigma_{12} A_1 v_2]$ has **ascending orientation** and $m^{(21)} = [v_1 - \sigma_{12} A_1 v_2]$ has **descending orientation**.

But these aren't the only forms of the midpoints; the η sigma relations allow us to rewrite the midpoints in two different forms. For example the side $\overline{a_1 a_2}$ midpoints can be rewritten as

$$\begin{aligned} m &= [\sigma_{13} \pm \sigma_{23} : \sigma_{13} \pm \sigma_{23} : \sigma_{13} \mp \sigma_{23}] \\ &= [\sigma_{14} \pm \sigma_{24} : \sigma_{14} \pm \sigma_{24} : \sigma_{14} \mp \sigma_{24}]. \end{aligned}$$

There are exactly two sigma representations for every midpoint, as *every side is in exactly two subtriangles*. For example the side $\overline{a_1 a_2}$ is in the subtriangles $\triangle_3$ and $\triangle_4$ which have corresponding midpoint representations

$$[\sigma_{14} \pm \sigma_{24} : \sigma_{14} \pm \sigma_{24} : \sigma_{14} \mp \sigma_{24}] \quad \text{and}$$

$$[\sigma_{13} \pm \sigma_{23} : \sigma_{13} \pm \sigma_{23} : \sigma_{13} \mp \sigma_{23}]$$

respectively. For this reason we will list the remaining midpoints of the sides in direct reference to the subtriangles in the next section.

6 Subtriangles and their centroids

We now give some precise notational conventions and listings for the subtriangles and their centroids.

6.1 Subtriangle 1

Subtriangle $\triangle_1$ has lines

$$L_{34} \equiv a_3 a_4 = \langle 1 : 1 : 0 \rangle, \quad L_{24} \equiv a_2 a_4 = \langle 1 : 0 : 1 \rangle,$$
$$L_{23} \equiv a_2 a_3 = \langle 0 : 1 : 1 \rangle$$

which correspond to the three sides of the quadrangle associated with the points a_2, a_3 and a_4. Note that the subscripts here really describe sets, although we write them as lists for brevity. As stated before, each midpoint of the sides of the quadrangle has exactly two sigma representations corresponding to the two subtriangles that a side belongs to.

In this light the sigma representations for the midpoints of these sides corresponding to the subtriangle $\triangle_1$, are given as follows;

$$\begin{aligned} m^{(34)} &\equiv [\sigma_{23} - \sigma_{24} : \sigma_{24} - \sigma_{23} : \sigma_{23} + \sigma_{24}], \\ m^{(43)} &\equiv [\sigma_{23} + \sigma_{24} : -\sigma_{24} - \sigma_{23} : \sigma_{23} - \sigma_{24}], \\ m^{(24)} &\equiv [-\sigma_{23} - \sigma_{34} : \sigma_{34} - \sigma_{23} : \sigma_{34} + \sigma_{23}], \\ m^{(42)} &\equiv [\sigma_{23} - \sigma_{34} : \sigma_{34} + \sigma_{23} : \sigma_{34} - \sigma_{23}], \\ m^{(23)} &\equiv [\sigma_{34} - \sigma_{24} : -\sigma_{24} - \sigma_{34} : \sigma_{34} + \sigma_{24}], \\ m^{(32)} &\equiv [\sigma_{34} + \sigma_{24} : \sigma_{24} - \sigma_{34} : \sigma_{34} - \sigma_{24}]. \end{aligned}$$

The corresponding medians are

$$D_1^{(34)} \equiv a_2 m^{(34)} = \langle -\sigma_{24} : \sigma_{23} : \sigma_{23} - \sigma_{24} \rangle,$$
$$D_1^{(43)} \equiv a_2 m^{(43)} = \langle \sigma_{24} : \sigma_{23} : \sigma_{23} + \sigma_{24} \rangle,$$
$$D_1^{(24)} \equiv a_3 m^{(24)} = \langle \sigma_{34} : \sigma_{23} + \sigma_{34} : \sigma_{23} \rangle,$$
$$D_1^{(42)} \equiv a_3 m^{(42)} = \langle \sigma_{34} : \sigma_{34} - \sigma_{23} : -\sigma_{23} \rangle,$$
$$D_1^{(23)} \equiv a_4 m^{(23)} = \langle \sigma_{24} + \sigma_{34} : \sigma_{34} : \sigma_{24} \rangle,$$
$$D_1^{(32)} \equiv a_4 m^{(32)} = \langle \sigma_{24} - \sigma_{34} : -\sigma_{34} : \sigma_{24} \rangle.$$

These medians are concurrent three at a time giving the four subtriangle centroids for $\triangle_1$:

$$g_1^1 \equiv D_1^{(34)} D_1^{(24)} D_1^{(23)} = \begin{bmatrix} \sigma_{23}\sigma_{24} - \sigma_{23}\sigma_{34} + \sigma_{24}\sigma_{34} : \\ \sigma_{23}\sigma_{24} + \sigma_{23}\sigma_{34} - \sigma_{24}\sigma_{34} : \\ \sigma_{23}\sigma_{24} - \sigma_{23}\sigma_{34} - \sigma_{24}\sigma_{34} \end{bmatrix},$$
$$g_1^2 \equiv D_1^{(34)} D_1^{(42)} D_1^{(32)} = \begin{bmatrix} \sigma_{23}\sigma_{24} + \sigma_{23}\sigma_{34} - \sigma_{24}\sigma_{34} : \\ \sigma_{23}\sigma_{24} - \sigma_{23}\sigma_{34} + \sigma_{24}\sigma_{34} : \\ -\sigma_{23}\sigma_{24} + \sigma_{23}\sigma_{34} + \sigma_{24}\sigma_{34} \end{bmatrix},$$
$$g_1^3 \equiv D_1^{(43)} D_1^{(24)} D_1^{(32)} = \begin{bmatrix} \sigma_{23}\sigma_{24} + \sigma_{23}\sigma_{34} + \sigma_{24}\sigma_{34} : \\ \sigma_{23}\sigma_{24} - \sigma_{23}\sigma_{34} - \sigma_{24}\sigma_{34} : \\ -\sigma_{23}\sigma_{24} + \sigma_{23}\sigma_{34} - \sigma_{24}\sigma_{34} \end{bmatrix},$$
$$g_1^4 \equiv D_1^{(43)} D_1^{(42)} D_1^{(23)} = \begin{bmatrix} \sigma_{23}\sigma_{24} - \sigma_{23}\sigma_{34} - \sigma_{24}\sigma_{34} : \\ \sigma_{23}\sigma_{24} + \sigma_{23}\sigma_{34} + \sigma_{24}\sigma_{34} : \\ -\sigma_{23}\sigma_{24} - \sigma_{23}\sigma_{34} + \sigma_{24}\sigma_{34} \end{bmatrix}.$$

In Figure 13 we see how these centroids appear.

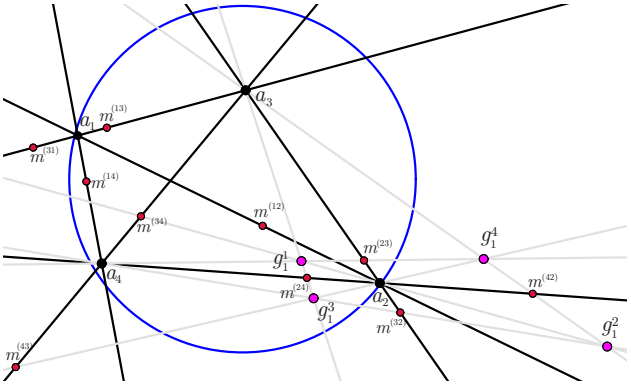


Figure 13: *Labelling for the centroids of* $\triangle_1 = \overline{a_2 a_3 a_4}$

The labelling of a centroid reflects both the subtriangle it is associated with and the three distinct midpoints used in its construction, at least one of which is of positive orientation. The centroid g_1^2 for example is associated with the midpoints $m^{(32)}$, $m^{(42)}$ and $m^{(34)}$ used to construct it, exactly one of which is positively oriented, namely $m^{(34)}$, which lies on the median $D_1^{(34)} = a_2 m^{(34)}$. Thus the superscript of the centroid g_1^2 comes from the point a_2 which is incident with the median through the positively oriented midpoint.

Similarly the centroids g_1^3 and g_1^4 are associated with precisely one positively oriented midpoint $m^{(24)}$ and $m^{(23)}$ respectively, which lie on the medians $D_1^{(24)} = a_3 m^{(24)}$ and $D_1^{(23)} = a_4 m^{(23)}$ respectively, and so are associated with the points a_3 and a_4 respectively. Finally the centroid g_1^1 is distinct from the other centroids in the sense that it is associated with *three* positively oriented midpoints, and is thus associated with the point a_1. This method for labelling the centroids can be extended to the rest of the subtriangles in the natural way. So our centroid labelling depends on and is consistent with the prior labelling of midpoints. We now proceed to the other subtriangles, where the pattern is the same. But it is useful to see all of these laid out explicitly.

6.2 Subtriangle 2

Subtriangle $\triangle_2$ has lines

$$L_{14} \equiv a_1 a_4 = \langle 0 : 1 : -1 \rangle,$$
$$L_{13} \equiv a_1 a_3 = \langle 1 : 0 : -1 \rangle,$$
$$L_{34} \equiv a_3 a_4 = \langle 1 : 1 : 0 \rangle,$$

with midpoints

$$m^{(14)} \equiv [\sigma_{13} - \sigma_{34} : \sigma_{13} + \sigma_{34} : \sigma_{13} + \sigma_{34}],$$
$$m^{(41)} \equiv [\sigma_{13} + \sigma_{34} : \sigma_{13} - \sigma_{34} : \sigma_{13} - \sigma_{34}],$$
$$m^{(13)} \equiv [\sigma_{14} + \sigma_{34} : \sigma_{14} - \sigma_{34} : \sigma_{14} + \sigma_{34}],$$
$$m^{(31)} \equiv [\sigma_{14} - \sigma_{34} : \sigma_{14} + \sigma_{34} : \sigma_{14} - \sigma_{34}],$$
$$m^{(34)} \equiv [\sigma_{13} - \sigma_{14} : \sigma_{14} - \sigma_{13} : \sigma_{13} + \sigma_{14}],$$
$$m^{(43)} \equiv [\sigma_{13} + \sigma_{14} : -\sigma_{13} - \sigma_{14} : \sigma_{13} - \sigma_{14}],$$

medians

$$D_2^{(14)} \equiv a_3 m^{(14)} = \langle \sigma_{13} + \sigma_{34} : \sigma_{34} : -\sigma_{13} \rangle,$$
$$D_2^{(41)} \equiv a_3 m^{(41)} = \langle \sigma_{34} - \sigma_{13} : \sigma_{34} : \sigma_{13} \rangle,$$
$$D_2^{(13)} \equiv a_4 m^{(13)} = \langle \sigma_{34} : \sigma_{14} + \sigma_{34} : -\sigma_{14} \rangle,$$
$$D_2^{(31)} \equiv a_4 m^{(31)} = \langle \sigma_{34} : \sigma_{34} - \sigma_{14} : \sigma_{14} \rangle,$$
$$D_2^{(34)} \equiv a_1 m^{(34)} = \langle \sigma_{13} : -\sigma_{14} : \sigma_{14} - \sigma_{13} \rangle,$$
$$D_2^{(43)} \equiv a_1 m^{(43)} = \langle \sigma_{13} : \sigma_{14} : -\sigma_{14} - \sigma_{13} \rangle,$$

and centroids

$$g_2^2 \equiv D_2^{(14)}D_2^{(13)}D_2^{(34)} = \begin{bmatrix} \sigma_{13}\sigma_{14}+\sigma_{13}\sigma_{34}-\sigma_{14}\sigma_{34} : \\ \sigma_{13}\sigma_{14}-\sigma_{13}\sigma_{34}+\sigma_{14}\sigma_{34} : \\ \sigma_{13}\sigma_{14}+\sigma_{13}\sigma_{34}+\sigma_{14}\sigma_{34} \end{bmatrix},$$
$$g_2^1 \equiv D_2^{(41)}D_2^{(31)}D_2^{(34)} = \begin{bmatrix} \sigma_{13}\sigma_{14}-\sigma_{13}\sigma_{34}+\sigma_{14}\sigma_{34} : \\ \sigma_{13}\sigma_{14}+\sigma_{13}\sigma_{34}-\sigma_{14}\sigma_{34} : \\ \sigma_{13}\sigma_{14}-\sigma_{13}\sigma_{34}-\sigma_{14}\sigma_{34} \end{bmatrix},$$
$$g_2^4 \equiv D_2^{(41)}D_2^{(13)}D_2^{(43)} = \begin{bmatrix} \sigma_{13}\sigma_{14}+\sigma_{13}\sigma_{34}+\sigma_{14}\sigma_{34} : \\ \sigma_{13}\sigma_{14}-\sigma_{13}\sigma_{34}-\sigma_{14}\sigma_{34} : \\ \sigma_{13}\sigma_{14}+\sigma_{13}\sigma_{34}-\sigma_{14}\sigma_{34} \end{bmatrix},$$
$$g_2^3 \equiv D_2^{(14)}D_2^{(31)}D_2^{(43)} = \begin{bmatrix} \sigma_{13}\sigma_{14}-\sigma_{13}\sigma_{34}-\sigma_{14}\sigma_{34} : \\ \sigma_{13}\sigma_{14}+\sigma_{13}\sigma_{34}+\sigma_{14}\sigma_{34} : \\ \sigma_{13}\sigma_{14}-\sigma_{13}\sigma_{34}+\sigma_{14}\sigma_{34} \end{bmatrix}.$$

6.3 Subtriangle 3

Subtriangle $\triangle_3$ has lines

$$L_{12} \equiv a_1a_2 = \langle 1 : -1 : 0\rangle, \quad L_{24} \equiv a_2a_4 = \langle 1 : 0 : 1\rangle,$$
$$L_{14} \equiv a_1a_4 = \langle 0 : 1 : -1\rangle$$

with midpoints

$$m^{(12)} \equiv [\sigma_{14}-\sigma_{24} : \sigma_{14}-\sigma_{24} : \sigma_{14}+\sigma_{24}],$$
$$m^{(21)} \equiv [\sigma_{14}+\sigma_{24} : \sigma_{14}+\sigma_{24} : \sigma_{14}-\sigma_{24}],$$
$$m^{(24)} \equiv [-\sigma_{12}-\sigma_{14} : \sigma_{14}-\sigma_{12} : \sigma_{12}+\sigma_{14}],$$
$$m^{(42)} \equiv [\sigma_{12}-\sigma_{14} : \sigma_{14}+\sigma_{12} : \sigma_{14}-\sigma_{12}],$$
$$m^{(14)} \equiv [\sigma_{12}-\sigma_{24} : \sigma_{12}+\sigma_{24} : \sigma_{12}+\sigma_{24}],$$
$$m^{(41)} \equiv [\sigma_{12}+\sigma_{24} : \sigma_{12}-\sigma_{24} : \sigma_{12}-\sigma_{24}],$$

medians

$$D_3^{(12)} \equiv a_4m^{(12)} = \langle \sigma_{24} : \sigma_{14} : \sigma_{24}-\sigma_{14}\rangle,$$
$$D_3^{(21)} \equiv a_4m^{(21)} = \langle \sigma_{24} : -\sigma_{14} : \sigma_{24}+\sigma_{14}\rangle,$$
$$D_3^{(24)} \equiv a_1m^{(24)} = \langle \sigma_{12} : -\sigma_{12}-\sigma_{14} : \sigma_{14}\rangle,$$
$$D_3^{(42)} \equiv a_1m^{(42)} = \langle \sigma_{12} : \sigma_{14}-\sigma_{12} : -\sigma_{14}\rangle,$$
$$D_3^{(14)} \equiv a_2m^{(14)} = \langle \sigma_{12}+\sigma_{24} : -\sigma_{12} : \sigma_{24}\rangle,$$
$$D_3^{(41)} \equiv a_2m^{(41)} = \langle \sigma_{24}-\sigma_{12} : \sigma_{12} : \sigma_{24}\rangle,$$

and centroids

$$g_3^3 \equiv D_3^{(12)}D_3^{(24)}D_3^{(14)} = \begin{bmatrix} \sigma_{12}\sigma_{14}-\sigma_{12}\sigma_{24}-\sigma_{14}\sigma_{24} : \\ \sigma_{12}\sigma_{14}-\sigma_{12}\sigma_{24}+\sigma_{14}\sigma_{24} : \\ \sigma_{12}\sigma_{14}+\sigma_{12}\sigma_{24}+\sigma_{14}\sigma_{24} \end{bmatrix},$$
$$g_3^4 \equiv D_3^{(12)}D_3^{(42)}D_3^{(41)} = \begin{bmatrix} \sigma_{12}\sigma_{14}-\sigma_{12}\sigma_{24}+\sigma_{14}\sigma_{24} : \\ \sigma_{12}\sigma_{14}-\sigma_{12}\sigma_{24}-\sigma_{14}\sigma_{24} : \\ \sigma_{12}\sigma_{14}+\sigma_{12}\sigma_{24}-\sigma_{14}\sigma_{24} \end{bmatrix},$$
$$g_3^1 \equiv D_3^{(21)}D_3^{(24)}D_3^{(41)} = \begin{bmatrix} \sigma_{12}\sigma_{14}+\sigma_{12}\sigma_{24}+\sigma_{14}\sigma_{24} : \\ \sigma_{12}\sigma_{14}+\sigma_{12}\sigma_{24}-\sigma_{14}\sigma_{24} : \\ \sigma_{12}\sigma_{14}-\sigma_{12}\sigma_{24}-\sigma_{14}\sigma_{24} \end{bmatrix},$$
$$g_3^2 \equiv D_3^{(21)}D_3^{(42)}D_3^{(14)} = \begin{bmatrix} \sigma_{12}\sigma_{14}+\sigma_{12}\sigma_{24}-\sigma_{14}\sigma_{24} : \\ \sigma_{12}\sigma_{14}+\sigma_{12}\sigma_{24}+\sigma_{14}\sigma_{24} : \\ \sigma_{12}\sigma_{14}-\sigma_{12}\sigma_{24}+\sigma_{14}\sigma_{24} \end{bmatrix}.$$

6.4 Subtriangle 4

Finally subtriangle $\triangle_4$ has lines

$$L_{23} \equiv a_2a_3 = \langle 0 : 1 : 1\rangle, \quad L_{13} \equiv a_1a_3 = \langle 1 : 0 : -1\rangle,$$
$$L_{12} \equiv a_1a_2 = \langle 1 : -1 : 0\rangle,$$

with midpoints

$$m^{(23)} \equiv [\sigma_{13}-\sigma_{12} : -\sigma_{12}-\sigma_{13} : \sigma_{12}+\sigma_{13}],$$
$$m^{(32)} \equiv [\sigma_{12}+\sigma_{13} : \sigma_{12}-\sigma_{13} : \sigma_{13}-\sigma_{12}],$$
$$m^{(13)} \equiv [\sigma_{12}+\sigma_{23} : \sigma_{12}-\sigma_{23} : \sigma_{12}+\sigma_{23}],$$
$$m^{(31)} \equiv [\sigma_{12}-\sigma_{23} : \sigma_{12}+\sigma_{23} : \sigma_{12}-\sigma_{23}],$$
$$m^{(12)} \equiv [\sigma_{13}-\sigma_{23} : \sigma_{13}-\sigma_{23} : \sigma_{13}+\sigma_{23}],$$
$$m^{(21)} \equiv [\sigma_{13}+\sigma_{23} : \sigma_{13}+\sigma_{23} : \sigma_{13}-\sigma_{23}],$$

medians

$$D_4^{(23)} \equiv a_1m^{(23)} = \langle -\sigma_{12}-\sigma_{13} : \sigma_{12} : \sigma_{13}\rangle,$$
$$D_4^{(32)} \equiv a_1m^{(32)} = \langle \sigma_{12}-\sigma_{13} : -\sigma_{12} : \sigma_{13}\rangle,$$
$$D_4^{(13)} \equiv a_2m^{(13)} = \langle -\sigma_{12} : \sigma_{12}+\sigma_{23} : \sigma_{23}\rangle,$$
$$D_4^{(31)} \equiv a_2m^{(31)} = \langle \sigma_{12} : \sigma_{23}-\sigma_{12} : \sigma_{23}\rangle,$$
$$D_4^{(12)} \equiv a_3m^{(12)} = \langle \sigma_{13} : \sigma_{23} : \sigma_{23}-\sigma_{13}\rangle,$$
$$D_4^{(21)} \equiv a_3m^{(21)} = \langle -\sigma_{13} : \sigma_{23} : \sigma_{23}+\sigma_{13}\rangle,$$

and centroids

$$g_4^4 \equiv D_4^{(12)} D_4^{(13)} D_4^{(23)} = \begin{bmatrix} \sigma_{12}\sigma_{13} - \sigma_{12}\sigma_{23} + \sigma_{13}\sigma_{23} : \\ \sigma_{12}\sigma_{13} - \sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{23} : \\ \sigma_{12}\sigma_{13} + \sigma_{12}\sigma_{23} + \sigma_{13}\sigma_{23} \end{bmatrix},$$

$$g_4^3 \equiv D_4^{(12)} D_4^{(31)} D_4^{(32)} = \begin{bmatrix} \sigma_{12}\sigma_{13} - \sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{23} : \\ \sigma_{12}\sigma_{13} - \sigma_{12}\sigma_{23} + \sigma_{13}\sigma_{23} : \\ \sigma_{12}\sigma_{13} + \sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{23} \end{bmatrix},$$

$$g_4^2 \equiv D_4^{(21)} D_4^{(13)} D_4^{(32)} = \begin{bmatrix} \sigma_{12}\sigma_{13} + \sigma_{12}\sigma_{23} + \sigma_{13}\sigma_{23} : \\ \sigma_{12}\sigma_{13} + \sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{23} : \\ \sigma_{12}\sigma_{13} - \sigma_{12}\sigma_{23} + \sigma_{13}\sigma_{23} \end{bmatrix},$$

$$g_4^1 \equiv D_4^{(21)} D_4^{(31)} D_4^{(23)} = \begin{bmatrix} \sigma_{12}\sigma_{13} + \sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{23} : \\ \sigma_{12}\sigma_{13} + \sigma_{12}\sigma_{23} + \sigma_{13}\sigma_{23} : \\ \sigma_{12}\sigma_{13} - \sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{23} \end{bmatrix}.$$

7 Finding the quadrangle centroids

What we would like to do is use centroids of subtriangles to construct centroids of the quadrangle $\overline{a_1a_2a_3a_4}$. In Figure 14 we see some of the sixteen subtriangle centroids, using different colours for each subtriangle. It turns out that these sixteen centroids enjoy some interesting relations.

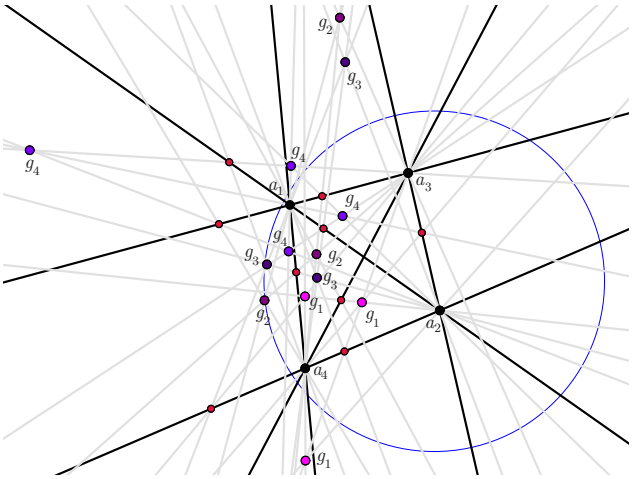


Figure 14: *Subtriangle centroids of the quadrangle* $\overline{a_1a_2a_3a_4}$

7.1 Midpoint consistencies

Define the **set of associated midpoints** (or **associated midpoints**) S_j^i for a centroid point g_j^i of a Triangle $\triangle_i$ to be the (unique) set of three distinct midpoints used to construct it. For example the centroid g_4^4 of the subtriangle $\triangle_4$ is constructed from the median lines $D_4^{(12)}$, $D_4^{(13)}$, and $D_4^{(23)}$, and hence

$$S_4^4 \equiv \left\{ m^{(12)}, m^{(13)}, m^{(23)} \right\}.$$

A set $\left\{ g_1^{i_1}, g_2^{i_2}, g_3^{i_3}, g_4^{i_4} \right\}$ containing one centroid from each subtriangle of the quadrangle $\overline{a_1a_2a_3a_4}$ is said to be **midpoint consistent** if the union of the associated midpoint sets

$$S_m \equiv \cup_k S_k^{i_k}$$

contains exactly one midpoint from each of the six sides of the quadrangle. For example the set

$$S = \left\{ g_1^1, g_2^2, g_3^3, g_4^4 \right\}$$

is midpoint consistent since the union of associated midpoints is

$$S_m = \left\{ m^{(12)}, m^{(34)}, m^{(13)}, m^{(24)}, m^{(14)}, m^{(23)} \right\}$$

which has exactly six elements, one from each side.

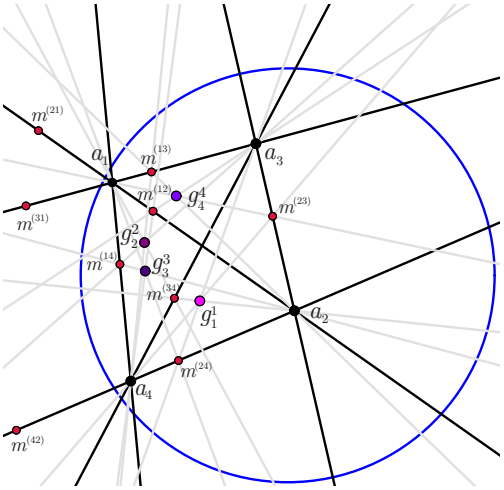


Figure 15: *A midpoint consistent set of subtriangle centroids with six associated midpoints*

On the other hand the set

$$S = \left\{ g_1^1, g_2^2, g_3^4, g_4^3 \right\}$$

is not midpoint consistent, since the union of associated midpoints

$$S_m = \left\{ m^{(12)}, m^{(34)}, m^{(13)}, m^{(31)}, m^{(42)}, m^{(14)}, m^{(41)}, m^{(32)} \right\}$$

contains both midpoints for the sides $\overline{a_1a_3}$ and $\overline{a_1a_4}$, and in addition contains eight points.

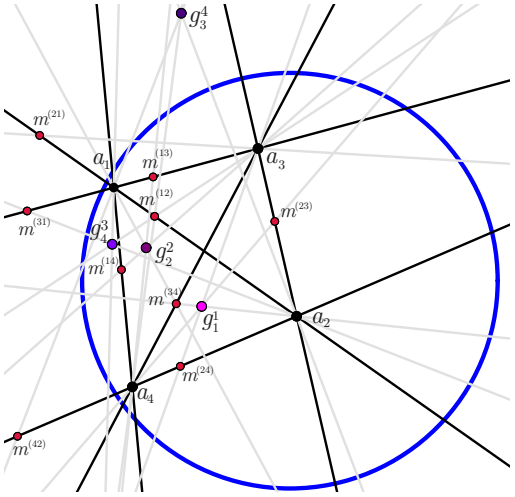


Figure 16: *A set of subtriangle centroids which are not midpoint consistent*

Theorem 5 (Midpoint consistent centroids) *There are exactly eight distinct sets of subtriangle centroids*

$$S = \left\{g_1^{i_1}, g_2^{i_2}, g_3^{i_3}, g_4^{i_4}\right\}$$

which are midpoint consistent.

Proof. We can see this by trying to construct a midpoint consistent set of subtriangle centroids $S = \left\{g_1^{i_1}, g_2^{i_2}, g_3^{i_3}, g_4^{i_4}\right\}$. First let's choose $i_1 = 1$, that is let g_1^1 be in the set. The centroid g_1^1 has associated midpoints $S_1^1 \equiv \left\{m^{(23)}, m^{(24)}, m^{(34)}\right\}$, and so for any other centroid in S, the midpoints $m^{(32)}, m^{(42)}$, and $m^{(43)}$ cannot be in their respective set of associated midpoints.

So looking at the centroids $g_2^{i_2}$ of the subtriangle $\triangle_2$ and their associated midpoints

$$\begin{aligned}
g_2^2 : S_2^2 &\equiv \left\{m^{(13)}, m^{(14)}, m^{(34)}\right\},\\
g_2^1 : S_2^1 &\equiv \left\{m^{(31)}, m^{(41)}, m^{(34)}\right\},\\
g_2^4 : S_2^4 &\equiv \left\{m^{(31)}, m^{(14)}, m^{(43)}\right\},\\
g_2^3 : S_2^3 &\equiv \left\{m^{(13)}, m^{(41)}, m^{(43)}\right\},
\end{aligned}$$

we see that the centroids g_2^2 and g_2^1 are the only options for S. If we choose $i_2 = 2$, then the set

$$\left\{m^{(34)}, m^{(13)}, m^{(24)} m^{(14)}, m^{(23)}\right\} \subseteq S_m$$

which forces $i_3 = 3$, and $i_4 = 4$, as $S_3^3 \equiv \left\{m^{(12)}, m^{(14)}, m^{(24)}\right\}$ and $S_4^4 \equiv \left\{m^{(12)}, m^{(13)}, m^{(23)}\right\}$. Else if $i_2 = 1$, then the set

$$\left\{m^{(34)}, m^{(31)}, m^{(24)}, m^{(41)}, m^{(23)}\right\} \subseteq S_m$$

which forces $i_3 = i_4 = 1$, as $S_3^1 \equiv \left\{m^{(21)}, m^{(31)}, m^{(24)}\right\}$ and $S_4^1 \equiv \left\{m^{(21)}, m^{(31)}, m^{(23)}\right\}$. Therefore there are two distinct midpoint consistent sets of subtriangle centroids which contain the centroid g_1^1. The above method can be used for any choice of $i_1 \in \{1,2,3,4\}$. Hence there are in total exactly eight distinct sets of subtriangle centroids which are midpoint consistent which can be constructed in this manner. Each of these sets are distinct and this construction in fact covers all the cases of midpoint consistent sets. □

Corollary 1 *The complete list of midpoint consistent sets is given below:*

S :	S_m :
$S_\iota \equiv \{g_1^1, g_2^2, g_3^3, g_4^4\}$	$\left\{m^{(12)}, m^{(34)}, m^{(13)}, m^{(24)}, m^{(14)}, m^{(23)}\right\}$
$S_\alpha \equiv \{g_1^2, g_2^1, g_3^4, g_4^3\}$	$\left\{m^{(12)}, m^{(34)}, m^{(31)}, m^{(42)}, m^{(41)}, m^{(32)}\right\}$
$S_\beta \equiv \{g_1^3, g_2^4, g_3^1, g_4^2\}$	$\left\{m^{(21)}, m^{(43)}, m^{(13)}, m^{(24)}, m^{(41)}, m^{(32)}\right\}$
$S_\gamma \equiv \{g_1^4, g_2^3, g_3^2, g_4^1\}$	$\left\{m^{(21)}, m^{(43)}, m^{(31)}, m^{(42)}, m^{(14)}, m^{(23)}\right\}$
$S_1 \equiv \{g_1^1, g_2^1, g_3^1, g_4^1\}$	$\left\{m^{(21)}, m^{(34)}, m^{(31)}, m^{(24)}, m^{(41)}, m^{(23)}\right\}$
$S_2 \equiv \{g_1^2, g_2^2, g_3^2, g_4^2\}$	$\left\{m^{(21)}, m^{(34)}, m^{(13)}, m^{(42)}, m^{(14)}, m^{(32)}\right\}$
$S_3 \equiv \{g_1^3, g_2^3, g_3^3, g_4^3\}$	$\left\{m^{(12)}, m^{(43)}, m^{(31)}, m^{(24)}, m^{(14)}, m^{(32)}\right\}$
$S_4 \equiv \{g_1^4, g_2^4, g_3^4, g_4^4\}$	$\left\{m^{(12)}, m^{(43)}, m^{(13)}, m^{(42)}, m^{(41)}, m^{(23)}\right\}$.

These midpoint consistent sets of subtriangle centroids are quite pleasant. Indeed the labels almost chose themselves. The S_α, S_β, and S_γ sets represent the α, β, and γ pairings of the subtriangle centroids respectively, while the S_ι set represents the ι pairing or *identity pairing* of subtriangle centroids. The S_1, S_2, S_3 and S_4 sets contain the subtriangle centroids associated with the points a_1, a_2, a_3 and a_4 respectively.

It turns out that these sets of midpoint consistent subtriangle centroids behave in a very similar way as do quadrangle barycentres in the Euclidean case.

7.2 Quadrangle centroids

Define the **g-lines** to be the join $a_i g_i$ of the points a_i with the centroids g_i of the corresponding subtriangle $\triangle_i$. Thus a quadrangle has sixteen distinct g-lines, four passing through each point a_i.

Theorem 6 (Quadrangle centroids) *The four g-lines associated with a midpoint consistent set of subtriangle centroids are concurrent, producing eight distinct* ***quadrangle centroids*** *q.*

Proof. Consider the midpoint consistent set of subtriangle centroids

$$S_\iota = \{g_1^1, g_2^2, g_3^3, g_4^4\}.$$

The set S_ι has the four associated g-lines

$$a_1 g_1^1 = \langle \sigma_{23}(\sigma_{24}+\sigma_{34}) : -\sigma_{24}(\sigma_{23}+\sigma_{34}) : \sigma_{34}(\sigma_{24}-\sigma_{23}) \rangle,$$
$$a_2 g_2^2 = \langle \sigma_{14}(\sigma_{13}+\sigma_{34}) : -\sigma_{13}(\sigma_{14}+\sigma_{34}) : \sigma_{34}(\sigma_{14}-\sigma_{13}) \rangle,$$
$$a_3 g_3^3 = \langle \sigma_{14}(\sigma_{12}+\sigma_{24}) : \sigma_{24}(\sigma_{12}+\sigma_{14}) : \sigma_{12}(\sigma_{24}-\sigma_{14}) \rangle,$$
$$a_4 g_4^4 = \langle \sigma_{23}(\sigma_{12}+\sigma_{13}) : \sigma_{13}(\sigma_{12}+\sigma_{23}) : \sigma_{12}(\sigma_{23}-\sigma_{13}) \rangle.$$

Now let $q_\iota \equiv \left(a_1 g_1^1\right)\left(a_2 g_2^2\right)$, then we find that

$$q_\iota = \left\langle \begin{array}{c} \sigma_{23}(\sigma_{24}+\sigma_{34}) : \\ -\sigma_{24}(\sigma_{23}+\sigma_{34}) : \\ \sigma_{34}(\sigma_{24}-\sigma_{23}) \end{array} \right\rangle \times \left\langle \begin{array}{c} \sigma_{14}(\sigma_{13}+\sigma_{34}) : \\ -\sigma_{13}(\sigma_{14}+\sigma_{34}) : \\ \sigma_{34}(\sigma_{14}-\sigma_{13}) \end{array} \right\rangle$$
$$= \left[\begin{array}{c} \sigma_{24}\sigma_{34}(\sigma_{13}-\sigma_{14})(\sigma_{23}+\sigma_{34}) \\ -\sigma_{13}\sigma_{34}(\sigma_{23}-\sigma_{24})(\sigma_{14}+\sigma_{34}) : \\ \sigma_{23}\sigma_{34}(\sigma_{13}-\sigma_{14})(\sigma_{24}+\sigma_{34}) \\ -\sigma_{14}\sigma_{34}(\sigma_{23}-\sigma_{24})(\sigma_{13}+\sigma_{34}) : \\ \sigma_{14}\sigma_{24}(\sigma_{23}+\sigma_{34})(\sigma_{13}+\sigma_{34}) \\ -\sigma_{13}\sigma_{23}(\sigma_{24}+\sigma_{34})(\sigma_{14}+\sigma_{34}) \end{array} \right]$$
$$= \left[\begin{array}{c} -\sigma_{34}(\eta(\sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}-2\sigma_{34})+\sigma_3+\sigma_4) : \\ -\sigma_{34}(\eta(\sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}+2\sigma_{34})-\sigma_3-\sigma_4) : \\ -\sigma_{34}(\eta(\sigma_{13}-\sigma_{14}+\sigma_{23}-\sigma_{24})+\sigma_3-\sigma_4) \end{array} \right]$$

where $\eta = \sigma_{12}\sigma_{34} = \sigma_{13}\sigma_{24} = \sigma_{14}\sigma_{23}$. Now as $\sigma_{34} \neq 0$ we can use the fact that $\eta = \sigma_{12}\sigma_{34}$ and divide by $(\sigma_{34})^2$ to get the representation

$$q_\iota = \left[\begin{array}{c} \sigma_{12}(\sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}-2\sigma_{34})+\sigma_{13}\sigma_{23}+\sigma_{14}\sigma_{24} : \\ \sigma_{12}(\sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}+2\sigma_{34})-\sigma_{13}\sigma_{23}-\sigma_{14}\sigma_{24} : \\ \sigma_{12}(\sigma_{13}-\sigma_{14}+\sigma_{23}-\sigma_{24})+\sigma_{13}\sigma_{23}-\sigma_{14}\sigma_{24} \end{array} \right].$$

Recalling that A_3 and A_4 have representations $\sigma_{12}/(\sigma_{13}\sigma_{23})$ and $\sigma_{12}/(\sigma_{14}\sigma_{24})$ respectively, divide by σ_{12} to get the alternate expression

$$q_\iota = \left[\begin{array}{c} \sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}-2\sigma_{34}+1/A_3+1/A_4 : \\ \sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}+2\sigma_{34}-1/A_3-1/A_4 : \\ \sigma_{13}-\sigma_{14}+\sigma_{23}-\sigma_{24}+1/A_3-1/A_4 \end{array} \right].$$

It is then a calculation that q_ι is also incident with the g-lines $a_3 g_3^3$ and $a_4 g_4^4$, so that the four g-lines are concurrent at the point q_ι. The representation of q_ι above seems to be more associated with the points a_3 and a_4 when looking at the subscripts of the sigma values, and in particular the terms A_3 and A_4 are involved in the equation. This is due in part to the two g-lines we chose at the start, and so since there are six ways to choose two from four there are also six equal representations for the centroid q_ι. The five remaining representations of q_ι are

$$\left(a_3 g_3^3\right)\left(a_4 g_4^4\right) = \left[\begin{array}{c} \sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}-2\sigma_{12}+1/A_1+1/A_2 : \\ -\sigma_{13}+\sigma_{14}+\sigma_{23}-\sigma_{24}-2\sigma_{12}+1/A_1+1/A_2 : \\ \sigma_{13}+\sigma_{14}-\sigma_{23}-\sigma_{24}+1/A_1-1/A_2 \end{array} \right],$$

$$\left(a_1 g_1^1\right)\left(a_3 g_3^3\right) = \left[\begin{array}{c} \sigma_{12}+\sigma_{14}+\sigma_{23}+\sigma_{34}-2\sigma_{24}-1/A_2-1/A_4 : \\ \sigma_{12}+\sigma_{14}-\sigma_{23}-\sigma_{34}-1/A_2+1/A_4 : \\ \sigma_{12}+\sigma_{14}+\sigma_{23}+\sigma_{34}+2\sigma_{24}+1/A_2+1/A_4 \end{array} \right],$$

$$\left(a_2 g_2^2\right)\left(a_4 g_4^4\right) = \left[\begin{array}{c} -\sigma_{12}-\sigma_{14}-\sigma_{23}-\sigma_{34}+2\sigma_{13}+1/A_1+1/A_3 : \\ -\sigma_{12}+\sigma_{14}-\sigma_{23}+\sigma_{34}+1/A_1-1/A_3 : \\ \sigma_{12}+\sigma_{14}+\sigma_{23}+\sigma_{34}+2\sigma_{13}+1/A_1+1/A_3 \end{array} \right],$$

$$\left(a_1 g_1^1\right)\left(a_4 g_4^4\right) = \left[\begin{array}{c} \sigma_{12}+\sigma_{13}-\sigma_{24}-\sigma_{34}-1/A_2+1/A_3 : \\ \sigma_{12}+\sigma_{13}+\sigma_{24}+\sigma_{34}-2\sigma_{23}-1/A_2-1/A_3 : \\ \sigma_{12}+\sigma_{13}+\sigma_{24}+\sigma_{34}+2\sigma_{23}+1/A_2+1/A_3 \end{array} \right],$$

$$\left(a_2 g_2^2\right)\left(a_3 g_3^3\right) = \left[\begin{array}{c} -\sigma_{12}+\sigma_{13}-\sigma_{24}+\sigma_{34}+1/A_1-1/A_4 : \\ -\sigma_{12}-\sigma_{13}-\sigma_{24}-\sigma_{34}+2\sigma_{14}+1/A_1+1/A_4 : \\ \sigma_{12}+\sigma_{13}+\sigma_{24}+\sigma_{34}+2\sigma_{14}+1/A_1+1/A_4 \end{array} \right].$$

These can be transformed from one to another through an appropriate multiplication of a non-zero scalar.

We can similarly do this for each of the other midpoint consistent sets of subtriangle centroids producing eight distinct quadrangle centroids. □

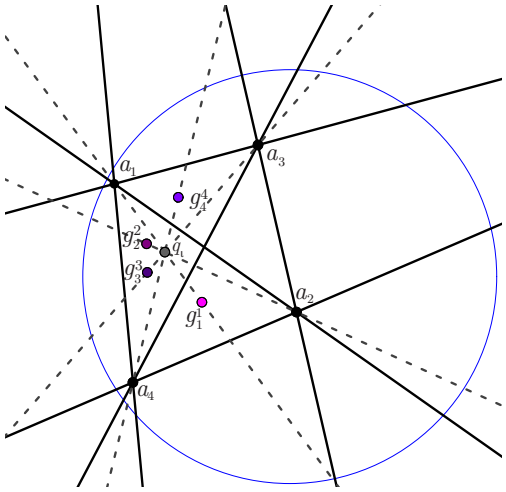


Figure 17: *Concurrent g-lines associated with a midpoint consistent set of subtriangle centroids*

The six different representations of a quadrangle centroid give an equal weighting between all the four points of the quadrangle. Despite this, we will break this symmetry to save space, and write all the quadrangle centroids in the representation derived from the meet of the g-lines $a_1 g_1^{i_1}$ and $a_2 g_2^{i_2}$ which are associated with the points a_1 and a_2. These representations have a bias associated with the points a_3 and a_4.

The eight distinct quadrangle centroids are then

$$q_\iota \equiv \left[\begin{array}{c} \sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}-2\sigma_{34}+1/A_3+1/A_4 : \\ \sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}+2\sigma_{34}-1/A_3-1/A_4 : \\ \sigma_{13}-\sigma_{14}+\sigma_{23}-\sigma_{24}+1/A_3-1/A_4 \end{array} \right],$$

$$q_\alpha \equiv \begin{bmatrix} \sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}+2\sigma_{34}-1/A_3-1/A_4 : \\ \sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}-2\sigma_{34}+1/A_3+1/A_4 : \\ \sigma_{13}-\sigma_{14}+\sigma_{23}-\sigma_{24}-1/A_3+1/A_4 \end{bmatrix},$$

$$q_\beta \equiv \begin{bmatrix} \sigma_{13}+\sigma_{14}+\sigma_{23}+\sigma_{24}+2\sigma_{34}+1/A_3+1/A_4 : \\ \sigma_{13}+\sigma_{14}+\sigma_{23}+\sigma_{24}-2\sigma_{34}-1/A_3-1/A_4 : \\ \sigma_{13}+\sigma_{14}-\sigma_{23}-\sigma_{24}+A_3-A_4 \end{bmatrix},$$

$$q_\gamma \equiv \begin{bmatrix} \sigma_{13}+\sigma_{14}+\sigma_{23}+\sigma_{24}-2\sigma_{34}-1/A_3-1/A_4 : \\ \sigma_{13}+\sigma_{14}+\sigma_{23}+\sigma_{24}+2\sigma_{34}+1/A_3+1/A_4 : \\ \sigma_{13}+\sigma_{14}-\sigma_{23}-\sigma_{24}-1/A_3+1/A_4 \end{bmatrix},$$

$$q_1 \equiv \begin{bmatrix} \sigma_{13}-\sigma_{14}+\sigma_{23}-\sigma_{24}+2\sigma_{34}-1/A_3-1/A_4 : \\ \sigma_{13}-\sigma_{14}+\sigma_{23}-\sigma_{24}-2\sigma_{34}+1/A_3+1/A_4 : \\ \sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}-1/A_3+1/A_4 \end{bmatrix},$$

$$q_2 \equiv \begin{bmatrix} \sigma_{13}-\sigma_{14}+\sigma_{23}-\sigma_{24}-2\sigma_{34}+1/A_3+1/A_4 : \\ \sigma_{13}-\sigma_{14}+\sigma_{23}-\sigma_{24}+2\sigma_{34}-1/A_3-1/A_4 : \\ \sigma_{13}-\sigma_{14}-\sigma_{23}+\sigma_{24}+1/A_3-1/A_4 \end{bmatrix},$$

$$q_3 \equiv \begin{bmatrix} \sigma_{13}+\sigma_{14}-\sigma_{23}-\sigma_{24}-2\sigma_{34}-1/A_3-1/A_4 : \\ \sigma_{13}+\sigma_{14}-\sigma_{23}-\sigma_{24}+2\sigma_{34}+1/A_3+1/A_4 : \\ \sigma_{13}+\sigma_{14}+\sigma_{23}+\sigma_{24}-1/A_3+1/A_4 \end{bmatrix},$$

$$q_4 \equiv \begin{bmatrix} \sigma_{13}+\sigma_{14}-\sigma_{23}-\sigma_{24}+2\sigma_{34}+1/A_3+1/A_4 : \\ \sigma_{13}+\sigma_{14}-\sigma_{23}-\sigma_{24}-2\sigma_{34}-1/A_3-1/A_4 : \\ \sigma_{13}+\sigma_{14}+\sigma_{23}+\sigma_{24}+1/A_3-1/A_4 \end{bmatrix}.$$

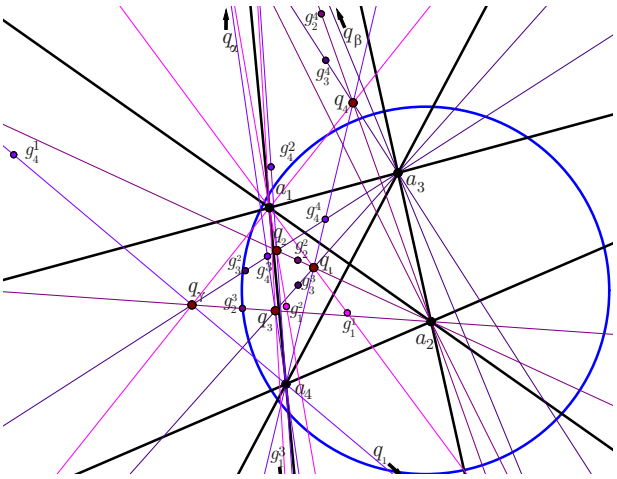


Figure 18: *The quadrangle centroids produced by g-lines*

From the list of midpoint consistent sets of subtriangle centroids and from the Quadrangle Centroid theorem we see that each g-line passes through exactly two quadrangle centroids. Moreover each g-line that passes through each of the quadrangle centroids $q_\iota, q_\alpha, q_\beta$ and q_γ is incident with one of the quadrangle centroids q_1, q_2, q_3, and q_4 and *not* any of the other quadrangle centroids $q_\iota, q_\alpha, q_\beta$ and q_γ. This is a symmetric relation, and so it is useful to consider the eight centroids as two quadrangles

$$\square_a \equiv \overline{q_\iota q_\alpha q_\beta q_\gamma} = \{q_\iota, q_\alpha, q_\beta, q_\gamma\}$$

and

$$\square_b \equiv \overline{q_1 q_2 q_3 q_4} = \{q_1, q_2, q_3, q_4\}.$$

We record this in a theorem.

Theorem 7 (Quadrangle centroid perspectivities) *The three quadrangles $\square$, $\square_a$ and $\square_b$ are pair-wise perspective in four ways, where the points of perspectivity are exactly the points of the third quadrangle.*

This is seen in Figure 19. The grey lines, which show the perspectivities, are exactly the g-lines of the quadrangle $\square = \overline{a_1 a_2 a_3 a_4}$.

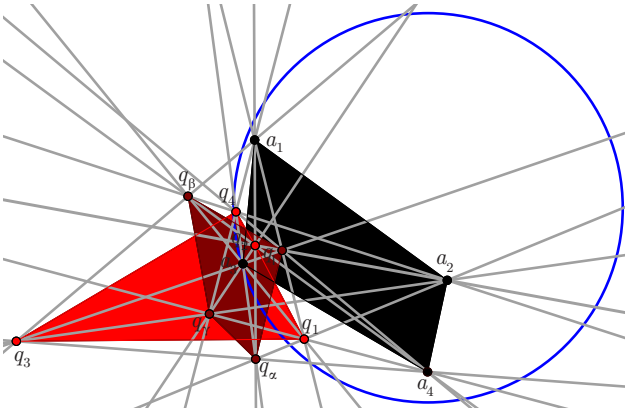


Figure 19: *Three perspective quadrangles $\square$, $\square_a$ and $\square_b$*

8 Bimedian Lines

Just as in the Euclidean case, there is more than one way to find the barycentric centroid of a quadrangle; we can also look at meets of the bimedian lines of the quadrangle.

A **bimedian line** $B_{\{ij,k\ell\}}$ is the join of two midpoints $m^{(ij)}$ and $m^{(k\ell)}$ from opposite sides $\overline{a_i a_j}$ and $\overline{a_k a_\ell}$ of the quadrangle, where $\{1,2,3,4\} = \{i,j,k,\ell\}$. By calculation, the bimedian lines $B_{\{ij,k\ell\}}$ of the quadrangle $\overline{a_1 a_2 a_3 a_4}$ are given as follows- note the pleasant linear aspect of these expressions.

The bimedian lines corresponding to the α opposite sides are:

$$B_{\{12,34\}} \equiv \langle \sigma_{13}-\sigma_{24} : \sigma_{23}-\sigma_{14} : \sigma_{23}+\sigma_{14}-\sigma_{13}-\sigma_{24} \rangle,$$
$$B_{\{12,43\}} \equiv \langle \sigma_{13}+\sigma_{24} : \sigma_{23}+\sigma_{14} : \sigma_{23}-\sigma_{14}-\sigma_{13}+\sigma_{24} \rangle,$$
$$B_{\{21,34\}} \equiv \langle \sigma_{13}+\sigma_{24} : -\sigma_{23}-\sigma_{14} : \sigma_{14}-\sigma_{23}-\sigma_{13}+\sigma_{24} \rangle,$$
$$B_{\{21,43\}} \equiv \langle \sigma_{24}-\sigma_{13} : \sigma_{23}-\sigma_{14} : \sigma_{23}+\sigma_{14}+\sigma_{13}+\sigma_{24} \rangle.$$

The bimedian lines corresponding to the β opposite sides are:

$$B_{\{13,24\}} \equiv \langle \sigma_{34}-\sigma_{12} : \sigma_{14}+\sigma_{34}+\sigma_{23}+\sigma_{12} : \sigma_{23}-\sigma_{14} \rangle,$$
$$B_{\{13,42\}} \equiv \langle \sigma_{34}+\sigma_{12} : \sigma_{14}-\sigma_{23}+\sigma_{34}-\sigma_{12} : -\sigma_{14}-\sigma_{23} \rangle,$$
$$B_{\{31,24\}} \equiv \langle \sigma_{34}+\sigma_{12} : \sigma_{23}-\sigma_{14}+\sigma_{34}-\sigma_{12} : \sigma_{14}+\sigma_{23} \rangle,$$
$$B_{\{31,42\}} \equiv \langle \sigma_{12}-\sigma_{34} : \sigma_{23}+\sigma_{14}-\sigma_{34}-\sigma_{12} : \sigma_{23}-\sigma_{14} \rangle.$$

The bimedian lines corresponding to the γ opposite sides are:

$$B_{\{14,23\}} \equiv \langle \sigma_{12}+\sigma_{34}+\sigma_{13}+\sigma_{24} : \sigma_{34}-\sigma_{12} : \sigma_{24}-\sigma_{13} \rangle,$$
$$B_{\{14,32\}} \equiv \langle \sigma_{12}-\sigma_{34}+\sigma_{24}-\sigma_{13} : -\sigma_{34}-\sigma_{12} : \sigma_{24}+\sigma_{13} \rangle,$$
$$B_{\{41,23\}} \equiv \langle \sigma_{34}-\sigma_{12}+\sigma_{24}-\sigma_{13} : \sigma_{34}+\sigma_{12} : \sigma_{24}+\sigma_{13} \rangle,$$
$$B_{\{41,32\}} \equiv \langle \sigma_{34}+\sigma_{12}-\sigma_{24}-\sigma_{13} : \sigma_{34}-\sigma_{12} : \sigma_{13}-\sigma_{24} \rangle.$$

Theorem 8 (Quadrangle bimedian centroids) *The bimedian lines $B_{\{ij,k\ell\}}$ of the quadrangle are concurrent three at a time at the quadrangle centroids.*

Proof. Given that we have equations of all the bimedian lines and the quadrangle centroids, this is a calculation involving the various sigma relations. □

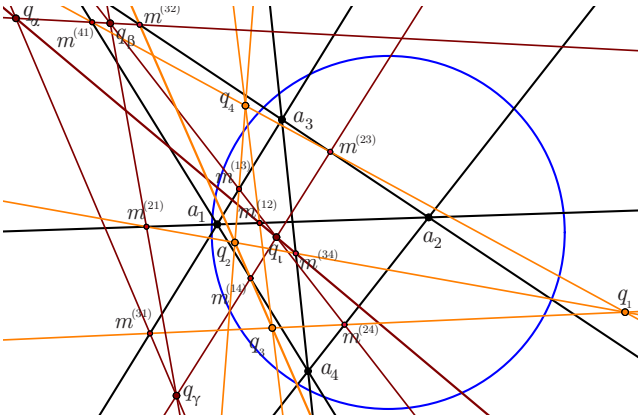


Figure 20: *The bimedian lines of a quadrangle determine the quadrangle centroids*

To record which triples of bimedian lines are concurrent, we once again can look at midpoint consistent sets of subtriangle centroids, and then look at the bimedian lines induced from the union of the associated midpoints.

Returning to the example previous we have the midpoint consistent set of subtriangle centroids $S_\iota = \{g_1^1, g_2^2, g_3^3, g_4^4\}$, where the union of associated midpoints is $\{m^{(12)}, m^{(34)}, m^{(13)}, m^{(24)}, m^{(14)}, m^{(23)}\}$. By definition of a midpoint consistent set, this union of the sets of associated midpoints contains precisely one midpoint for every side. Hence the set naturally corresponds to the bimedian lines $B_{\{12,34\}}$, $B_{\{13,24\}}$, and $B_{\{14,23\}}$. These are in fact concurrent at the point q_ι.

As each subtriangle centroid is in precisely two midpoint consistent sets, we get that each bimedian line is incident with precisely two quadrangle centroids, namely if a bimedian is incident with a quadrangle centroid from $\square_a$ then it is *not* incident with a quadrangle centroid from $\square_b$. So the twelve bimedian lines of the quadrangle $\overline{a_1a_2a_3a_4}$ are *exactly the lines of the sides of the quadrangles $\square_a$ and $\square_b$.*

More precisely the relations are

$$q_\iota q_\alpha = B_{\{12,34\}}, \qquad q_\beta q_\gamma = B_{\{21,43\}},$$
$$q_\iota q_\beta = B_{\{13,24\}}, \qquad q_\alpha q_\gamma = B_{\{31,42\}},$$
$$q_\iota q_\gamma = B_{\{14,23\}}, \qquad q_\alpha q_\beta = B_{\{41,32\}}$$

and

$$q_1 q_2 = B_{\{21,34\}}, \qquad q_3 q_4 = B_{\{12,43\}},$$
$$q_1 q_3 = B_{\{31,24\}}, \qquad q_2 q_4 = B_{\{13,42\}},$$
$$q_1 q_4 = B_{\{14,32\}}, \qquad q_2 q_3 = B_{\{41,23\}}.$$

We can once again recognise the α, β and γ pairings in these relations. For example in $\square_a$ the line $q_\iota q_\alpha$ is a bimedian line from the α opposite sides, while the line $q_\alpha q_\gamma$ is a bimedian line from the β opposite sides. The sides of $\square_a$ correspond to the bimedian lines constructed from midpoints with the same orientation. In contrast the sides of $\square_b$ correspond to the bimedian lines constructed from the midpoints with opposite orientations.

9 Diagonal triangles and perspectivities

We have observed that from the quadrangle $\overline{a_1a_2a_3a_4}$ we can find eight centroids, which separate into two distinct quadrangles, and that these three quadrangles have a threefold perspective relation. There is in addition a strong correlation between diagonal triangles of these three quadrangles which manifests itself algebraically in a very elegant way.

Recall that the diagonal triangle of the quadrangle $\overline{a_1a_2a_3a_4}$ is given by (1). Now from the equations of the bimedian lines, we can determine that the diagonal triangles

of the quadrangles $\square_a$ and $\square_b$ are given by the points

$$d_\alpha^a \equiv B_{\{12,34\}}B_{\{21,43\}} = \begin{bmatrix} (\sigma_{13}+\sigma_{24})(\sigma_{14}-\sigma_{23}) : \\ (\sigma_{13}-\sigma_{24})(\sigma_{14}+\sigma_{23}) : \\ (\sigma_{13}-\sigma_{24})(\sigma_{14}-\sigma_{23}) \end{bmatrix},$$

$$d_\beta^a \equiv B_{\{13,24\}}B_{\{31,42\}} = \begin{bmatrix} (\sigma_{12}+\sigma_{34})(\sigma_{14}-\sigma_{23}) : \\ (\sigma_{12}-\sigma_{34})(\sigma_{14}-\sigma_{23}) : \\ (\sigma_{12}-\sigma_{34})(\sigma_{14}+\sigma_{23}) \end{bmatrix},$$

$$d_\gamma^a \equiv B_{\{14,23\}}B_{\{41,32\}} = \begin{bmatrix} (\sigma_{12}-\sigma_{34})(\sigma_{13}-\sigma_{24}) : \\ (\sigma_{12}+\sigma_{34})(\sigma_{13}-\sigma_{24}) : \\ (\sigma_{12}-\sigma_{34})(\sigma_{13}+\sigma_{24}) \end{bmatrix}$$

and

$$d_\alpha^b \equiv B_{\{21,34\}}B_{\{12,43\}} = \begin{bmatrix} (\sigma_{13}-\sigma_{24})(\sigma_{14}+\sigma_{23}) : \\ (\sigma_{13}+\sigma_{24})(\sigma_{14}-\sigma_{23}) : \\ (\sigma_{13}+\sigma_{24})(\sigma_{14}+\sigma_{23}) \end{bmatrix},$$

$$d_\beta^b \equiv B_{\{31,24\}}B_{\{13,42\}} = \begin{bmatrix} (\sigma_{12}-\sigma_{34})(\sigma_{14}+\sigma_{23}) : \\ (\sigma_{12}+\sigma_{34})(\sigma_{14}+\sigma_{23}) : \\ (\sigma_{12}+\sigma_{34})(\sigma_{14}-\sigma_{23}) \end{bmatrix},$$

$$d_\gamma^b \equiv B_{\{41,23\}}B_{\{14,32\}} = \begin{bmatrix} (\sigma_{12}+\sigma_{34})(\sigma_{13}+\sigma_{24}) : \\ (\sigma_{12}-\sigma_{34})(\sigma_{13}+\sigma_{24}) : \\ (\sigma_{12}+\sigma_{34})(\sigma_{13}-\sigma_{24}) \end{bmatrix}.$$

There is no abuse of notation as the α, β, and γ diagonal points are constructed from the bimedian lines of the α, β, and γ opposite sides respectively.

We start with this first theorem concerned with the collinearity of diagonal points from different diagonal triangles.

Theorem 9 *The diagonal points, d, d^a and d^b are collinear, on six distinct lines, called* ***d-lines****.*

Proof. By computations we see that the following diagonal points from the quadrangles $\square, \square_a$ and $\square_b$ respectively are collinear on the lines

$$\begin{aligned}
&\langle \sigma_{34}-\sigma_{12} : \sigma_{12}+\sigma_{34} : 0\rangle && \text{through } d_\alpha, d_\beta^a, d_\gamma^b,\\
&\langle \sigma_{12}+\sigma_{34} : \sigma_{34}-\sigma_{12} : 0\rangle && \text{through } d_\alpha, d_\gamma^a, d_\beta^b,\\
&\langle \sigma_{13}+\sigma_{24} : 0 : \sigma_{24}-\sigma_{13}\rangle && \text{through } d_\beta, d_\gamma^a, d_\alpha^b,\\
&\langle \sigma_{24}-\sigma_{13} : 0 : \sigma_{13}+\sigma_{24}\rangle && \text{through } d_\beta, d_\alpha^a, d_\gamma^b,\\
&\langle 0 : \sigma_{23}-\sigma_{14} : \sigma_{14}+\sigma_{23}\rangle && \text{through } d_\gamma, d_\alpha^a, d_\beta^b,\\
&\langle 0 : \sigma_{14}+\sigma_{23} : \sigma_{23}-\sigma_{14}\rangle && \text{through } d_\gamma, d_\alpha^a, d_\beta^b. \qquad \square
\end{aligned}$$

It is quite pleasant that the equations of these six lines reduce to something so elementary, in our view.

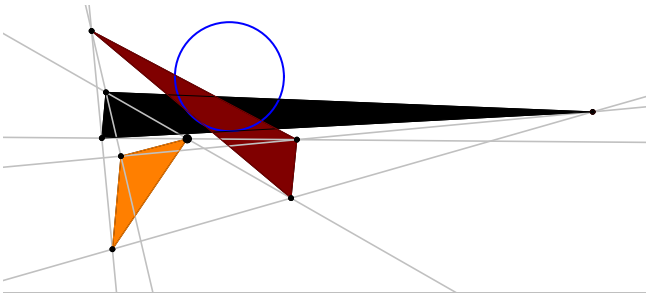


Figure 21: *Collinearity of diagonal triangles*

It is apparent that there is a strong relation between these three triangles, this is emphasized in the next couple of theorems which are concerned with perspectivity.

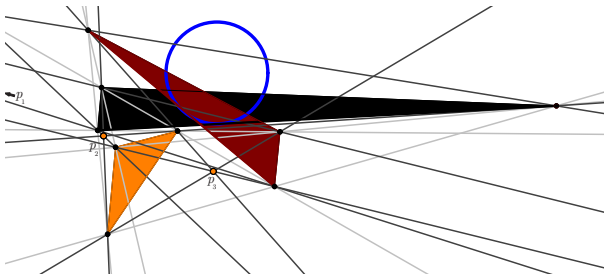


Figure 22: *The diagonal triangles are pairwise perspective.*

Theorem 10 *The three diagonal triangles of the quadrangles $\square, \square_a$ and $\square_b$ are pair-wise perspective.*

Proof. We can see this by computing the lines through corresponding diagonal points, for example

$$\begin{aligned}
d_\alpha d_\alpha^a &= \langle (\sigma_{13}-\sigma_{24})(\sigma_{14}+\sigma_{23}) : (\sigma_{13}+\sigma_{24})(\sigma_{23}-\sigma_{14}) : 0\rangle,\\
d_\beta d_\beta^a &= \langle (\sigma_{12}-\sigma_{34})(\sigma_{14}+\sigma_{23}) : 0 : (\sigma_{12}+\sigma_{34})(\sigma_{23}-\sigma_{14})\rangle,\\
d_\gamma d_\gamma^a &= \langle 0 : (\sigma_{12}-\sigma_{34})(\sigma_{13}+\sigma_{24}) : (\sigma_{12}+\sigma_{34})(\sigma_{24}-\sigma_{13})\rangle
\end{aligned}$$

are all concurrent at the point

$$p_1 \equiv \begin{bmatrix} (\sigma_{12}+\sigma_{34})(\sigma_{13}+\sigma_{24})(\sigma_{14}-\sigma_{23}) : \\ (\sigma_{12}+\sigma_{34})(\sigma_{13}-\sigma_{24})(\sigma_{14}+\sigma_{23}) : \\ (\sigma_{12}-\sigma_{34})(\sigma_{13}+\sigma_{24})(\sigma_{14}+\sigma_{23}) \end{bmatrix}.$$

Similarly the lines $d_\alpha d_\alpha^b$, $d_\beta d_\beta^b$, and $d_\gamma d_\gamma^b$ and the lines $d_\alpha^a d_\alpha^b$, $d_\beta^a d_\beta^b$, and $d_\gamma^a d_\gamma^b$ are concurrent respectively at the points

$$p_2 \equiv \begin{bmatrix} (\sigma_{12}-\sigma_{34})(\sigma_{13}-\sigma_{24})(\sigma_{14}+\sigma_{23}) : \\ (\sigma_{12}-\sigma_{34})(\sigma_{13}+\sigma_{24})(\sigma_{14}-\sigma_{23}) : \\ (\sigma_{12}+\sigma_{34})(\sigma_{13}-\sigma_{24})(\sigma_{14}-\sigma_{23}) \end{bmatrix}$$

and

$$p_3 \equiv \begin{bmatrix} \sigma_{234}-\sigma_{134}-\sigma_{124}+\sigma_{123} : \\ \sigma_{234}-\sigma_{134}+\sigma_{124}-\sigma_{123} : \\ \sigma_{234}+\sigma_{134}-\sigma_{124}-\sigma_{123} \end{bmatrix}. \qquad \square$$

The points that correspond in the different diagonal triangles are the α, β, and γ diagonal points respectively. We also have that these points of perspectivity are collinear, and we can find an attractive equation for this common line.

Theorem 11 *The points of perspectivity p_1, p_2 and p_3 for each pair of diagonal triangles are collinear.*

Proof. The points p_1,p_2 and p_3 lie on the line

$$\begin{aligned} L_1 &\equiv \left\langle \begin{matrix} \left(\sigma_{14}^2-\sigma_{23}^2\right)\left(\sigma_{12}^2-\sigma_{13}^2-\sigma_{24}^2+\sigma_{34}^2\right) : \\ \left(\sigma_{24}^2-\sigma_{13}^2\right)\left(\sigma_{12}^2-\sigma_{14}^2-\sigma_{23}^2+\sigma_{34}^2\right) : \\ \left(\sigma_{12}^2-\sigma_{34}^2\right)\left(\sigma_{13}^2-\sigma_{14}^2-\sigma_{23}^2+\sigma_{24}^2\right) \end{matrix} \right\rangle \\ &= \left\langle \begin{matrix} (A_2A_3-A_1A_4)(A_3A_4-A_2A_4-A_1A_3+A_1A_2) : \\ (A_1A_3-A_2A_4)(A_3A_4-A_2A_3-A_1A_4+A_1A_2) : \\ (A_3A_4-A_1A_2)(A_2A_4-A_2A_3-A_1A_4+A_1A_3) \end{matrix} \right\rangle \\ &= \left\langle \begin{matrix} \left(d^2-f^2\right)(ag+bg+cg-2df) : \\ \left(g^2-d^2\right)(af+bf+cf-2dg) : \\ \left(f^2-g^2\right)(ad+bd+cd-2fg) \end{matrix} \right\rangle. \end{aligned} \qquad \square$$

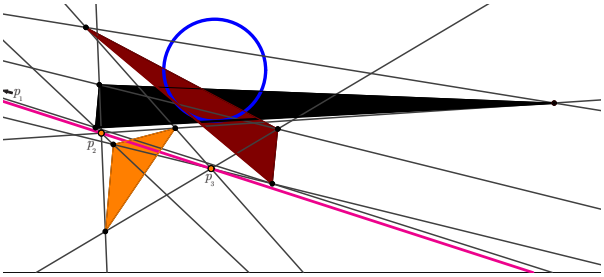


Figure 23: *A common Desargues line*

As discussed in [11], a corollary of Desargues' theorem is that if three triangles are pair-wise perspective and their points of perspectivity are collinear, then the corresponding sides of the three triangles are concurrent and hence the Desargues line is shared. We can verify this in this particular situation as follows

Corollary 2 *The perspective diagonal triangles share the same Desargues line which is*

$$L_2 \equiv \langle df : dg : fg \rangle .$$

Proof. The lines $d_\alpha d_\beta$, $d_\alpha^a d_\beta^a$, and $d_\alpha^b d_\beta^b$ are all concurrent at the point

$$\begin{aligned} &[0 : -\sigma_4+\sigma_3-\sigma_2+\sigma_1 : \sigma_4+\sigma_3-\sigma_2-\sigma_1] \\ &= [0 : f : -d] \end{aligned}$$

while the lines $d_\alpha d_\gamma$, $d_\alpha^a d_\gamma^a$, and $d_\alpha^b d_\gamma^b$ are concurrent at the point

$$\begin{aligned} &[\sigma_4-\sigma_3-\sigma_2+\sigma_1 : 0 : \sigma_4+\sigma_3-\sigma_2-\sigma_1] \\ &= [g : 0 : -d] \end{aligned}$$

and the lines $d_\beta d_\gamma$, $d_\beta^a d_\gamma^a$, and $d_\beta^b d_\gamma^b$ are concurrent at the point

$$\begin{aligned} &[\sigma_4-\sigma_3-\sigma_2+\sigma_1 : \sigma_4-\sigma_3+\sigma_2-\sigma_1 : 0] \\ &= [g : -f : 0]. \end{aligned}$$

These give the same Desargues line

$$L_2 \equiv \langle df : dg : fg \rangle$$

for each pair of perspective triangles. $\square$

The dual of L_2 is the point

$$L_2^\perp = L_2^T \mathbf{B} = \begin{bmatrix} (bf-dg)(cd-fg) : \\ (ag-df)(cd-fg) : \\ (ag-df)(bf-dg) \end{bmatrix}.$$

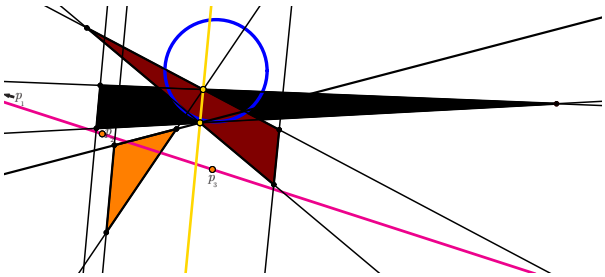


Figure 24: *Shared Desargues line (yellow) of the perspective triangles*

The situation here is richer than that of Desargues' corollary, as the points of the triangles are also collinear in threes.

The final representation of the formulas in the last two theorems deserve a mention. These are only dependent on variables of the general bilinear form and are independent of *sigma values*. This suggests that these lines have other roles to play in the geometry of the hyperbolic quadrangle, independent of centroid considerations. It would also be interesting of course to have synthetic projective arguments for these results that we have described algebraically.

10 Connections with desmic systems and tetrahedra, and final remarks

Both the books of Sommerville [13] and Coolidge [5], after constructing a projective metric, discuss how a side has two midpoints, a three-point system has four centroids, and finally that a four-point system has eight centroids. Their construction of the centroids for a four-point takes place in a three-dimensional space and is referred to as a *desmic system*. A desmic system is concerned with 'strongly' related tetrahedra, where these relations are the same as those described in the Quadrangle centroids subsection of this chapter. Namely each centroid lies on a line joining one point with the center of the other three (g-lines,) and that the produced tetrahedra are pair-wise perspective in four different ways, where the points of perspectivity are exactly the points of the other tetrahedra. They also state that the corresponding planes of perspective tetrahedra of a desmic system intersect at four lines which are coplanar, the four planes (one for each point of perspectivity) are the faces of the third tetrahedra. This last fact is unseen in the planar quadrangle case, as the third dimension is needed for this type of perspectivity. Yet it might be possible to view the planar quadrangle case that we go through above as some sort of projection of this desmic system onto a plane.

A major point of departure with the planar case is that tetrahedra *do not have an analogue of a diagonal triangle*. So the relations we described between the diagonal triangles of the quadrangles are unique and separate from what is described in these books. Furthermore, since classically only the interior points of the absolute were used for basic geometry these eight centroids were not visible, and thus the relations unseen.

The observant reader might have noticed that at no point do we prove that the sets of centroids $\{q_\iota, q_\alpha, q_\beta, q_\gamma\}$ and $\{q_1, q_2, q_3, q_4\}$ are in fact quadrangles. They are quadrangles except when the points of the original quadrangle lie on a hyperbolic circle.

If this happens it means that the subtriangles share a circumcenter, or equivalently they share a circumline. Recall that a circumline is the line through three collinear midpoints of a triangle. By a simple counting exercise, the number of collinear midpoints must be six. This is for the number of collinear midpoints must be more than 3, as different subtriangles have different midpoints, and less than 7 as there are six sides and seven or more collinear midpoints would imply that the points of the quadrangle are collinear. Finally the number of collinear midpoints must be a multiple of 3, giving the only possible choice as 6.

Thus, the three bimedian lines produced from these midpoints coincide, producing three collinear quadrangle centroids. In this case the diagonal triangle degenerates to three collinear points, coinciding with the three collinear quadrangle centroid points. All the remaining theorems still hold but one of the perspective triangles is degenerate, as in Figure 25.

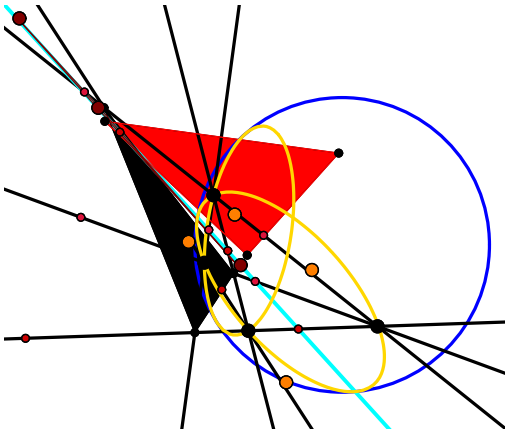


Figure 25: *Three quadrangle centroids are collinear on the blue line.*

If furthermore the points of the quadrangle lie on two circles, then the situation completely degenerates as two sets of six midpoints, and three centroids are collinear. This configuration results in two midpoints coinciding, and furthermore reduces to seven quadrangle centroids. The diagonal theorems in this case do not hold, as one diagonal triangle degenerates into two points, as seen in Figure 26.

Finally, apart from these theorems relating to the quadrangle centroids, there appear to be many other notable relations between subtriangle centroids and some structures concerning the circumlines of the subtriangles. And more generally we can look at other quadrangle structures that are analogous to the triangle centre investigations in [14], [17] and [20] in hyperbolic geometry.

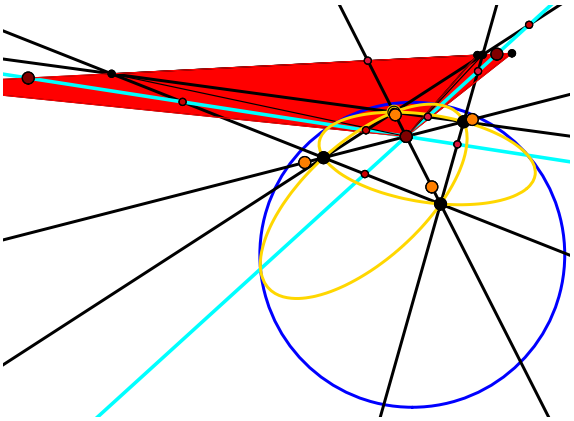


Figure 26: *Two quadrangle centroids coincide at the intersection of the blue lines.*

References

[1] A. ALKHALDI, N. J. WILDBERGER, The Parabola in Universal Hyperbolic Geometry I, *KoG* **17** (2013), 14–42.

[2] A. ALKHALDI, N. J. WILDBERGER, The Parabola in Universal Hyperbolic Geometry II: Canonical points and the Y-conic, *Journal for Geometry and Graphics* **20** (2016), 1–11.

[3] H. BRAUNER, *Geometrie Projektiver Räume I, II*, Bibliographisches Institut, Mannheim, 1976.

[4] H. BUSEMANN, P. J. KELLY, *Projective Geometry and Projective Metrics*, Dover Publications, New York, 2006, originally published by Academic Press, New York, 1956.

[5] J. L. COOLIDGE, *The Elements of Non-Euclidean Geometry*, Merchant Books, 1909.

[6] M. J. GREENBERG, *Euclidean and Non-Euclidean Geometries: Development and History*, 4th ed., W. H. Freeman and Co., San Francisco, 2007.

[7] N. LE, *Four-Fold Symmetry in Universal Triangle Geometry*, Mathematics & Statistics Thesis, Faculty of Science, UNSW, 2015.

[8] N. LE, N. J. WILDBERGER, Incenter Circles, Chromogeometry, and the Omega Triangle, *KoG* **18** (2014), 5–18.

[9] N. LE, N. J. WILDBERGER, Universal Affine Triangle Geometry and Four-fold Incenter Symmetry, *KoG* **16** (2012), 63–80.

[10] A. L. ONISHCHIK, R. SULANKE, *Projective and Cayley-Klein Geometries*, Springer Berlin Heidelberg, New York, 2006.

[11] J. RICHTER-GEBERT, *Perspectives on Projective Geometry: A guided tour through real and complex geometry*, Springer, Heidelberg, 2011.

[12] E. SNAPPER, R. J. TROYER, *Metric Affine Geometry,* Academic Press, New York, 1971.

[13] D. M. Y. SOMMERVILLE, *The Elements of Non-Euclidean Geometry*, G. Bell and Sons, Ltd., London, 1914, reprinted by Dover Publications, New York, 2005.

[14] A. UNGAR, *Hyperbolic Triangle Centers: The Special Relativistic Approach*, FTP, **166**, Springer, 2010.

[15] N. J. WILDBERGER, *Divine Proportions: Rational Trigonometry to Universal Geometry*, Wild Egg Books, Sydney, 2005.

[16] N. J. WILDBERGER, Affine and Projective Universal Geometry `http://arxiv.org/abs/math/0612499v1`.

[17] N. J. WILDBERGER, Universal Hyperbolic Geometry I: Trigonometry, *Geometriae Dedicata* **163** (2013), 215–274.

[18] N. J. WILDBERGER, Universal Hyperbolic Geometry II: A pictorial overview, *KoG* **14** (2010), 3–24.

[19] N. J. WILDBERGER, Universal Hyperbolic Geometry III: First steps in projective triangle geometry, *KoG* **15** (2011), 25–49.

[20] N. J. WILDBERGER, A. ALKHALDI, Universal Hyperbolic Geometry IV: Sydpoints and Twin Circumcircles, *KoG* **16** (2012), 43–62.

[21] R. M. WINGER, *An Introduction to Projective Geometry*, Dover Publications, New York, 1962.

Sebastian Blefari
e-mail: sebastian.eduard.blefari@gmail.com

N J Wildberger
e-mail: n.wildberger@unsw.edu.au

School of Mathematics and Statistics UNSW
Sydney 2052 Australia

Original scientific paper
Accepted 9. 12. 2016.

BORIS ODEHNAL

On Algebraic Minimal Surfaces

On Algebraic Minimal Surfaces

ABSTRACT

We give an overiew on various constructions of algebraic minimal surfaces in Euclidean three-space. Especially low degree examples shall be studied. For that purpose, we use the different representations given by WEIERSTRASS including the so-called Björling formula. An old result by LIE dealing with the evolutes of space curves can also be used to construct minimal surfaces with rational parametrizations. We describe a one-parameter family of rational minimal surfaces which touch orthogonal hyperbolic paraboloids along their curves of constant Gaussian curvature. Furthermore, we find a new class of algebraic and even rationally parametrizable minimal surfaces and call them cycloidal minimal surfaces.

Key words: minimal surface, algebraic surface, rational parametrization, polynomial parametrization, meromorphic function, isotropic curve, Weierstraß-representation, Björling formula, evolute of a spacecurve, curve of constant slope

MSC2010: 53A10, 53A99, 53C42, 49Q05, 14J26, 14Mxx

O algebarskim minimalnim plohama

SAŽETAK

Dajemo pregled različitih konstrukcija algebarskih minimalnih ploha u euklidskom trodimenzionalnom prostoru. Posebice se promatraju primjeri niskog stupnja. U tu svrhu koristimo različite prikaze koje daje WEIERSTRASS, uključujući takozvanu Björlingovu formulu. LIJEV stari rezultat pokazuje da se evolute prostornih krivulja mogu koristiti za konstruiranje minimalnih ploha s racionalnim parametrizacijama. Mi opisujemo jednoparametarsku familiju racionalnih minimalnih ploha koje diraju ortogonalne hiperboličke paraboloide duž njihovih krivulja s konstantnom Gaussovom zakrivljenošću. Štoviše, nalazimo novu klasu algebarskih i čak racionalno parametrizirajućih minimalnih ploha i nazivamo ih cikloidnim minimalnim plohama.

Ključne riječi: minimalna ploha, algebarska ploha, racionalna parametrizacija, polinomialna parametrizacija, meromorfična funkcija, izotropna krivulja, Weierstraßov prikaz, Björlingova formula, evoluta prostorne krivulje, krivulja konstantnog nagiba

1 Introduction

Minimal surfaces have been studied from many different points of view. Boundary value problems, uniqueness results, stability, and topological problems related to minimal surfaces have been and are still topics for investigations. There are only a few results on algebraic minimal surfaces. Most of them were published in the second half of the nine-teenth century, *i.e.*, more or less in the beginning of modern differential geometry. Only a few publications by LIE [30] and WEIERSTRASS [50] give general results on the generation and the properties of algebraic minimal surfaces. This may be due to the fact that computer algebra systems were not available and classical algebraic geometry gained less attention at that time. Many of the computations are hard work even nowadays and synthetic reasoning is somewhat uncertain. Besides some general work on minimal surfaces like [5, 8, 43, 44], there were some isolated results on algebraic minimal surfaces concerned with special tasks: minimal surfaces on certain scrolls [22, 35, 47, 49, 53], minimal surfaces related to congruences of lines [25, 28, 34, 38] minimal surfaces with a given geodesic [23], minimal surfaces of a certain degree, class, or genus (whether real or not) [1, 10, 11, 19, 20, 21, 31, 41, 42, 48], minimal surfaces touching surfaces along special curves [22], minimal surfaces showing special symmetries [14, 15, 16, 17], or minimal surface which allow isometries to special classes of surfaces [4, 6, 18, 52].

The famous algebraic minimal surface by ENNEPER which is of degree 9 and class 6 attracted intensive investigation. Consequently, researchers have found different generations of this surface: as the envelope of the planes of symmetry of all points on the pair of focal parabolas

$$p_1(u) = (\tfrac{4}{3}u, 0, \tfrac{2}{3}u^2 - \tfrac{1}{3}),$$

$$p_2(v) = (0, \tfrac{4}{3}v, \tfrac{1}{3} - \tfrac{2}{3}v^2)$$

or as the unique minimal surface (22) through the rational curve

$$\gamma(t) = \left(t - \tfrac{1}{3}t^3, t^2, 0\right)$$

having γ's normals for its surface normals. Since γ is planar, the surface normals of the uniquely defined minimal surface form a developable surface (to be precise, a plane), and thus, γ is a planar geodesic on ENNEPER's minimal surface. The plane of γ is a plane of symmetry of ENNEPER's surface. This is a manifestation of a more general result by HENNEBERG, see [21, 24, 30, 33]:

Theorem 1. *A minimal surface $\mathcal{M}$ carries a planar and not straight curve c as a geodesic. If $\mathcal{M}$ is algebraic, then the involutes of c have to be algebraic or c is the evolute of a planar algebraic curve.*

We shall make use of this fact later in Sec. 7 when we construct cycloidal minimal surfaces.
A further result due to HENNEBERG (see [21, 24, 30, 33]) is the following

Theorem 2. *Let a minimal surface $\mathcal{M}$ be tangent to a cylinder $\mathcal{Z}$. If $\mathcal{M}$ is algebraic, then the orthogonal crosssection c of $\mathcal{Z}$ is the evolute of an algebraic curve. If c is the evolute of a transcendental curve, then $\mathcal{M}$ is also transcendental.*

However, according to a theorem by RIBAUCOUR, ENNEPER's surface, like many other minimal surfaces, appears as the *central envelope* of isotropic congruences of lines, see [25, 28, 34, 38, 45].
Among the real algebraic minimal surfaces, ENNEPER's surface has lowest possible degree 9. But there are algebraic minimal surfaces that can be found in [12, 13, 21, 30] which are of degree 3 and 4 having the equations

$$\mathcal{G}:\ (x-\mathrm{i}y)^4+3(x^2+y^2+z^2)=0$$

and

$$\mathcal{L}:\ 2(x-\mathrm{i}y)^3-6\mathrm{i}(x-\mathrm{i}y)z-3(x+\mathrm{i}y)=0$$

with respect to a properly chosen Cartesian coordinate system. The surfaces $\mathcal{G}$ and $\mathcal{L}$ have no real equation (polynomial equation with real coefficients exclusively) and do not carry a single real point.
$\mathcal{G}$ is usually called GEISER's *surface* and $\mathcal{L}$ is named after LIE. GEISER's minimal surface is a minimal surface of revolution with an isotropic axis. Obviously, it is of degree 4 and some computation tells us that the equation of its dual surface $\mathcal{G}^\star$, *i.e.*, the surface of its tangent planes has the equation

$$\mathcal{G}^\star:\ 9w_0^2(w_1-\mathrm{i}w_2)^4-(w_1^2+w_2^2+w_3^2)^3=0$$

which is, therefore, of degree 6, and thus, $\mathcal{G}$ is of class 6.
Whereas LIE's surface is of degree 3 and also of class 3 since the implicit equation of the dual surface $\mathcal{L}^\star$ reads

$$\mathcal{L}^\star:\ 27w_0(w_2+\mathrm{i}w_1)^2+9\mathrm{i}(w_1^2+w_2^2)w_3-4\mathrm{i}w_3^3=0.$$

GEISER's surface meets the ideal plane in the same ideal line as LIE's surface does. The ideal line $x-\mathrm{i}y=0$ is a 4-fold line on $\mathcal{G}$ and a 3-fold line on $\mathcal{L}$. It is remarkable that complex (non-real) algebraic minimal surfaces have been undergoing detailed investigations, see, *e.g.*, [1, 10, 12, 13, 48].
In [30], LIE gives a result dealing with the ideal curves of algebraic minimal surfaces:

Theorem 3. *The intersection of an algebraic minimal surface with the plane at infinity consists of finitely many lines.*

Some of the ideal lines on a minimal surface may have higher multiplicities and pairs of complex conjugate lines can also occcur.
For the coordinatization of ideal points and lines we refer to Sec. 2.
The results on degrees, ranks, and classes of real algebraic minimal surfaces differ from the results on complex algebraic minimal surfaces. For real algebraic minimal surfaces we have (see [30])

Theorem 4. *The sum of the degree and class of a* real *algebraic minimal surface is at least 15.*

The two aforementioned examples of complex minimal surfaces obviously show a different behaviour.
It is well-kown (cf. [30, 33]) that 5 is the lowest possible class of a real algebraic minimal surface. HENNEBERG's surface with the parametrization

$$f(u,v)=\begin{pmatrix} \mathrm{c}_{3u}\mathrm{S}_{3v}-3\mathrm{c}_u\mathrm{S}_v \\ \mathrm{s}_{3u}\mathrm{S}_{3v}+3\mathrm{s}_u\mathrm{S}_v \\ 3\mathrm{c}_{2u}\mathrm{C}_{2v} \end{pmatrix} \tag{1}$$

is an example for that, since the implicit equation of its dual surface equals

$$u_0(u_1^2+u_2^2)^2++u_3(u_1^2-u_2^2)(3u_1^2+3u_2^2+2u_3^2)=0. \tag{2}$$

The algebraic degree of HENNEBERG's surface equals 15. ENNEPER's surface is the only known example of a minimal surface where the degree and class sum up to 15: the degree equals 9 (cf. (23)), the class equals 6 (cf. (24)).
LIE gives also results on the class of an algebraic minimal surface:

Theorem 5. *The class of an orientable algebraic minimal surface is always even.*

HENNEBERG's surface is of class 5 and non-orientable. The rational minimal Möbius strip given in [35] is of class 15.
In Sec. 2, we introduce coordinates and define all necessary abbreviations. Then, the different parametrization techniques for minimal surfaces are collected. Proofs for these

can be found in most of the standard monographs on minimal surfaces or differential geometry such as [2, 33, 46]. Sec. 3 is dedicated to ENNEPER's surface and its natural generalizations. In Sec. 4, BOUR's minimal surfaces gain attention. We show different ways to find these minimal surfaces and give estimates on the algebraic degrees of these surfaces. Then, in Sec. 5, RICHMOND's surface appears as one in a one-parameter family. Sec. 6 gives additional and apparently new results on a well-known kind of minimal surface tangent to a hyperbolic paraboloid. Sec. 7 deals with an apparently new class of minimal surfaces. The fact that cycloids (cycloidal curves with cusps) have rational normals and are algebraic as well as their evolutes and involutes are (see [32, 51, 55, 56]), allows us to construct a family of algebraic minimal surfaces that admit even rational parametrizations. We debunk their relations to curves of constant slope on quadrics of revolution.

The reasons for the interest in algebraic and, especially in rational minimal surfaces are manifold: Rational parametrizations can be converted into a geometrically favorable representation, namely into the Bézier representation. Moreover, rational parametrizations can easily be handled with computer algebra systems. This allows the computation of implicit equations of surfaces and their duals and makes them accessible for further study which is then no longer restricted to the purely differential geometric approach. The behaviour at infinity as well as other algebraic properties can be studied.

We have to confess that implicit equations of algebraic minimal surfaces will hardly show up in this paper because they can be really long. The algebraic equation of a d-dimensional algebraic variety of degree D has at most

$$q = \frac{1}{(d+1)!}\prod_{k=1}^{d+1}(D+k)$$

coefficients. In the case of the classical low degree examples by ENNEPER, RICHMOND, HENNEBERG, and BOUR with degrees 9, 12, 15, and 16 we could expect up to 220, 455, 816, and 969 terms provided that no special coordinate system is chosen and that the equations are expanded in full length.

2 Prerequisites

Since we are dealing with minimal surfaces in the Euclidean three-space, Cartesian coordinates (x,y,z) are sufficient. Vectors and matrices are written in bold characters. The canonical innerproduct of two vectors $\mathbf{u},\mathbf{v}\in\mathbb{R}^3$ is denoted by $\langle\mathbf{u},\mathbf{v}\rangle$. The Euclidean length $\|\mathbf{v}\|$ of a vector $\mathbf{v}$ is then given by $\|\mathbf{v}\|=\sqrt{\langle\mathbf{v},\mathbf{v}\rangle}$. The induced crossproduct of two vectors $\mathbf{u},\mathbf{v}\in\mathbb{R}^3$ is the vector $\mathbf{u}\times\mathbf{v}\in\mathbb{R}^3$.

In the following, we shall use the abbreviations

$$\mathrm{c}_x := \cos x,\ \mathrm{s}_x := \sin x,\ldots$$

$$\mathrm{C}_x := \cosh x,\ \mathrm{S}_x := \sinh x,\ldots$$

for the trigonometric and hyperbolic functions whenever there is not enough space for the equations.
Sometimes, we deal with ideal points, lines, and the ideal plane. Then, we shall homogenize the underlying Cartesian coordinates by

$$x\to X_1X_0^{-1},\ y\to X_2X_0^{-1},\ z\to X_3X_0^{-1}.$$

When we compute the intersection of a (minimal) surface with the ideal plane (plane at infinity), then we let $X_0=0$ and obtain the equation of a curve (or, more generally speaking, a *cycle* which is the union of finitely many algebraic curves) in terms of the homogeneous coordinates $(X_1:X_2:X_3)$ in the ideal plane. However, we shall not write this down in detail and define coordinates in the ideal plane by simply setting $X_1=x$, $X_2=y$, and $X_3=z$. It is sufficient to do so, because substituting $X_0=0$ into the homogeneous equation returns all monomials of the highest degree of the inhomogeneous equation.
In the following, we collect some results and representations of minimal surfaces that will be useful for the generation of algebraic minimal surfaces. These representations are well-known and proofs can be found in the literature, see, *e.g.*, [2, 27, 30, 33, 36, 46].

2.1 BJÖRLING's problem

Let $\gamma\colon I\subset\mathbb{R}\to\mathbb{R}^3$ be a smooth curve and let $\nu\colon I\to\mathrm{S}^2$ be a smooth unit vector field along γ with $\langle\gamma',\nu\rangle\equiv 0$, *i.e.*, ν is perpendicular to γ in the entire interval I. Both are considered to have complex continuations. A real parametrization $\mathbf{f}\colon D\subset\mathbb{R}^2\to\mathbb{R}^3$ of the uniquely defined real minimal surface $\mathcal{M}$ through γ with its normals along γ parallel to ν is then given by

$$\mathbf{f}(u,v)=\mathfrak{Re}\left(\gamma(t)-\mathrm{i}\int_{t_0}^{t}\nu(\theta)\times\mathrm{d}\gamma(\theta)\right). \tag{3}$$

We call the pair (γ,ν) a *scroll* and it is the envelope of the one-parameter family of planes $\langle\nu(t),\mathbf{x}-\gamma(t)\rangle=0$. The curve γ shall henceforth be called the *spine curve* of the scroll.
Since γ and ν are considered to have complex continuations, the parameter t in (3) is assumed to be a complex parameter. Subsequent to the integration, t is replaced by $t=u+\mathrm{i}v$ and finally the real part of the vector function in $\mathbb{C}^3$ is extracted. Formula (3) is called *Björling formula*, see [2, 27, 33, 36], and was first published by H.A. SCHWARZ

in [44]. Actually, the Björling formula is just the solution of a problem posed by E.G. BJÖRLING in 1844.

The Björling formula can be a starting point for the construction of algebraic minimal surfaces, but it has a big disadvantage like all other integral formulae: Antiderivatives of rational or algebraic functions may sometimes be not rational or even algebraic.

A remarkable application of the Björling formula (3) may be its application to non planar curves. The following result is due to LIE, see [30]:

Theorem 6. *The minimal surface that touches the evolute $c^\star$ of an algebraic space curve c exactly at the centers of curvature of c is algebraic.*

However, the algebraic degree of the surface generated according Thm. 6 may not only be high, it may even be hard to determine.

As an application of Thm. 6, we can give the following low degree example: We choose the PH-curve (for details and definition see [9])

$$\mathbf{c}(t) = (6t, 6t^2, 4t^3), \quad t \in \mathbb{R}. \tag{4}$$

Its evolute is then parametrized by

$$\mathbf{c}^\star(t) = \begin{pmatrix} -12t^3 \\ 3 - 12t^4 + 6t^2 \\ 16t^3 + 6t \end{pmatrix}, \quad t \in \mathbb{R} \tag{5}$$

and the normals $\nu(t)$ are $\lambda\mathbf{c}_1 = (1, 2t, 2t^2)$ with $\lambda = 1 + 2t^2$. The requirements for the application of the Björling formula are met since $\langle \dot{\mathbf{c}}^\star, \mathbf{c}_1 \rangle = 0$. A real parametrization of the real minimal surface on the scroll $(\gamma, \nu) = (\mathbf{c}^\star, \mathbf{c}_1)$ is found with (3) and reads

$$\begin{aligned}\mathbf{f}(u,v) =& 12 \begin{pmatrix} 4uv(u^2 - v^2) \\ 6u^2v^2 - u^4 - v^4 \\ 0 \end{pmatrix} + \\ &+ 12 \begin{pmatrix} 3uv^2 - u^3 \\ v^3 - 3u^2v \\ \frac{4}{3}u^3 - 4uv^2 \end{pmatrix} + 6 \begin{pmatrix} 2uv \\ u^2 - v^2 - v + \frac{1}{2} \\ u(2v+1) \end{pmatrix}.\end{aligned} \tag{6}$$

Figure 1 shows the minimal surface parametrized by (6) together with the curves $\mathbf{c}$ and $\mathbf{c}^\star$.

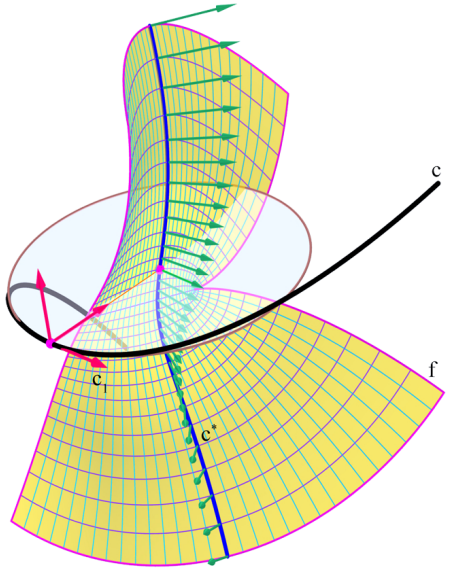


Figure 1: *The minimal surface on the scroll $(\mathbf{c}^\star, \mathbf{c}_1)$ is derived from the evolute $\mathbf{c}^\star$ of a cubic PH-curve $\mathbf{c}$.*

Implicitization shows that the surface (6) is of degree 16 and the intersection with the ideal plane consists of the ideal line of all planes parallel to $x = 0$ with multiplicity 16. Surprisingly, the class of this minimal surface equals 8 as we can see from the implicit equation of the dual surface:

$$\begin{aligned}&3\omega^2\Omega^2 + (4w_0w_2 - 15w_1^2)\omega\Omega^2 - 2\Omega\omega^3 - \omega^4 \\ &+4w_1^2(3w_1^2 - 4w_0 2w_2)\Omega^2 + 4w_1^5(2w_1 + 9w_3)\Omega \\ &+w_1(4w_0w_2(2w_1 + 3w_3) - 9w_1^2(5w_1 - 6w_3))\Omega\omega \\ &+2w_1(w_0w_2(w_1 + 3w_3) - 6w_1^2(w_1 + w_3))\omega^2 \\ &+(39w_1^2 + 18w_1w_3 - 2w_0w_2)\Omega\omega^2 \\ &+(12w_1^2 + 6w_1w_3 - w_0^2 - 2w_0w_2)\omega^3 + w_1^5(w_1 + 6w_3)\omega = 0\end{aligned} \tag{7}$$

where $\omega := w_1^2 + w_2^2$ and $\Omega := w_1^2 + w_2^2 + w_3^2$.
We can summarize this in

Corollary 1. *The minimal surface on the scroll $(\mathbf{c}^\star, \mathbf{c}_1)$ with $\mathbf{c}^\star$ given in (5) (evolute of the polynomial cubic PH-curve $\mathbf{c}$ from (4)) and with $\mathbf{c}_1$ being $\mathbf{c}$'s unit tangent vector field is a rational minimal surface of degree 16 and class 8.*

The cubic curve (4) as well as its evolute (5) are non planar curves. In contrast to that, we can choose the planar PH-cubic (semi-cubi parabola)

$$\gamma(t) = (4t^3, 0, 6t^2 + 3) \tag{8}$$

that lies in the xz-plane. Together with its unit normals

$$\nu(t) = \frac{1}{\sqrt{1+t^2}}(-1, 0, t) \tag{9}$$

a scroll (γ, ν) is defined and (3) yields the isotropic curve

$$\varphi(t) = (4t^3, -4\mathrm{i}\sqrt{(t^2+1)^3}, 6t^2+3). \tag{10}$$

which is subsequently reparametrized by $t = \mathrm{S}_\tau$. Then, $\tau = v + \mathrm{i}u$ (note that the real part equals v). Finally, the extraction of the real part of (10) gives (1). Since the normals ν from (9) along γ from (8) form a developable ruled surface (a plane), γ turns out to be a planar geodesic on HENNEBERG's minimal surface (1). The plane of γ is a plane of symmetry for HENNEBERG's minimal surface, cf. Thm. 1. Figure 2 shows a part of HENNEBERG's minimal surface with the geodesic semi-cubic parabola (8).

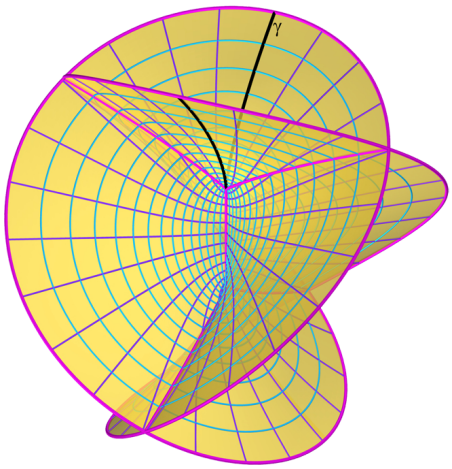


Figure 2: HENNEBERG*'s minimal surface with the geodesic semi-cubic parabola* γ.

A rational parametrization of HENNEBERG's surface can be obtained in two ways. The usual replacement of trigonometric and hyperbolic functions by their well-known rational equivalents delivers a parametrization involving polynomials of degrees higher than necessary. The substitution $\mathrm{S}_v = V$ yields a parametrization of bi-degree $(6,3)$, since $\mathrm{C}_{2v} = 1 + 2\mathrm{S}_v{}^2 = 1 + 2V^2$ and $\mathrm{S}_{3v} = 3\mathrm{S}_v + 4\mathrm{S}_v{}^3 = 3V + 4V^3$.
Implicitization shows that HENNEBERG's surface is of algebraic degree 15.
The dual surface, *i.e.*, the set of tangent planes of HENNEBERG's surface, can be given either in parametric form by

$$\mathbf{f}^\star = \begin{pmatrix} \dfrac{2\mathrm{c}_u}{\mathrm{c}_{2u}\mathrm{S}_{3v} + 3\mathrm{c}_{2u}\mathrm{S}_v} \\ \dfrac{2\mathrm{s}_u}{\mathrm{c}_{2u}\mathrm{S}_{3v} + 3\mathrm{c}_{2u}\mathrm{S}_v} \\ \dfrac{-1}{\mathrm{C}_{2v}\mathrm{c}_{2u} + 2\mathrm{c}_{2u}} \end{pmatrix} \tag{11}$$

or by the implicit equation (2).

2.2 WEIERSTRASS's formulae

2.2.1 The integral formula

There are some equivalent formulae which where first given by WEIERSTRASS. These allow us to compute parametrizations of minimal surfaces by prescribing a pair of meromorphic functions: Let $A: D \subset \mathbb{C} \to \mathbb{C}$ and $B: D \subset \mathbb{C} \to \mathbb{C}$ be meromorphic functions, *i.e.*, they are holomorphic except at countably many points $p_i \in D \subset \mathbb{C}$. From A and B we find a real parametrization of a real minimal surface via

$$\mathbf{f}(u,v) = \mathfrak{Re} \int \begin{pmatrix} A(1-B^2) \\ \mathrm{i}A(1+B^2) \\ 2AB \end{pmatrix} \mathrm{d}w. \tag{12}$$

Again, we assume that $w = u + \mathrm{i}v$ is the complex parameter in the domain D. The extraction of the real part of the complex vector valued function gives the real parametrization of the real minimal surface defined by A and B.
There is an alternative, but equivalent form for (12). Let G and H be two meromorphic functions defined over the same domain $D \subset \mathbb{C}$, then

$$\mathbf{f}(u,v) = \mathfrak{Re} \int \begin{pmatrix} G^2 - H^2 \\ \mathrm{i}(G^2 + H^2) \\ 2GH \end{pmatrix} \mathrm{d}w \tag{13}$$

also yields a real parametrization of a real minimal surface. (13) transforms into (12) by letting $A = G^2$ and $B = HG^{-1}$ provided that $G \not\equiv 0$.
In many textbooks on differential geometry and in a huge amount of publications, a further but equivalent integral representation of minimal surfaces can be found. However, this third version is obtained from (12) by substituting $B(w) = w$ and $A(w)$ is an arbitray meromorphic function. This seems to be a restriction that presumes that $A(w)$ can globally and in a closed form be written as a function $A(B(w))$ depending on $B(w)$.

2.2.2 Recovering the functions A, B

From the parametrization $\mathbf{f}$ of a minimal surface we can recover the meromorphic functions A and B, see [27, 33, 36]: First, we compute $\mathbf{F} := \partial_u \mathbf{f} - \mathrm{i}\partial_v \mathbf{f}$. Then, we use the coordinate functions $\mathbf{F}^i$ of $\mathbf{F}$ and find

$$A = \frac{1}{2}(\mathbf{F}^1 - \mathrm{i}\mathbf{F}^2) \quad \text{and} \quad \mathrm{B} = \frac{\mathbf{F}^3}{2\mathrm{A}}. \tag{14}$$

For example, the generating meromorphic functions of the minimal surface given by (6) are

$$A = 3 - 12\mathrm{i}w \quad \text{and} \quad B = 1 + 2\mathrm{i}w.$$

2.2.3 Integral free representation of minimal surfaces

Let $A(w): D \subset \mathbb{C} \to \mathbb{C}$ be a meromorphic function and let further $A' = \frac{\mathrm{d}A}{\mathrm{d}w}$, $A'' = \frac{\mathrm{d}^2A}{\mathrm{d}w^2}$, and $A''' = \frac{\mathrm{d}^3A}{\mathrm{d}w^3}$ denote its first, second, and third complex derivative. The vector

$$\mathbf{i} = \begin{pmatrix} 1-w^2 \\ \mathrm{i}(1+w^2) \\ 2w \end{pmatrix} \tag{15}$$

is an *isotropic vector* in three-dimensional Euclidean space $\mathbb{R}^3$ since $\langle \mathbf{i}, \mathbf{i} \rangle = 0$. Again, primes $'$ indicate differentiation with respect to the complex variable w. Now, we define

$$\mathbf{j} = A''\mathbf{i} - A'\mathbf{i}' + A\mathbf{i}''. \tag{16}$$

It is elementary to verify that $\langle \mathbf{j}', \mathbf{j}' \rangle = 0$, and thus, $\mathbf{j}'$ is isotropic. Therefore, $\mathbf{f} = \mathfrak{Re}\,\mathbf{j}$ is a real parametrization of a real minimal surface. This parametrization is usually written as

$$\mathbf{f}(u,v) = \mathfrak{Re} \begin{pmatrix} (1-w^2)A'' + 2wA' - 2A \\ \mathrm{i}(1+w^2)A'' - 2\mathrm{i}wA' + 2\mathrm{i}A \\ 2wA'' - 2A' \end{pmatrix} \tag{17}$$

where $A''' \not\equiv 0$ in D, see [2, 27, 33]. In case of a quadratic polynomial A, (17) parametrizes a line. A cubic polynomial A delivers an Enneper surface. The minimal surface adjoint to HENNEBERG's surface is uniquely determined by the geodesic astroid α and its normals. The integral free parametrization of minimal surfaces allows us to state:

Theorem 7. *Each algebraic function $A: D \subset \mathbb{C} \to \mathbb{C}$ with $A''' \not\equiv 0$ (in the entire domain D) yields an algebraic minimal surface parametrized by* (17)*.*

Moreover, it is clear that polynomials $A \in \mathbb{C}[w]$ deliver polynomial parametrization. Further, each rational function $A = P/Q$ with $P, Q \in \mathbb{C}[w]$ and $\gcd(P,Q) = 1$ yields rational parametrization of minimal surfaces. However, just inserting rational or algebraic functions cannot guarantee that the algebraic degree of the resulting minimal surface is low. Sometimes a reparametrization turns a rational parametrization of a minimal surface into a polynomial one.

2.2.4 The associate family

The minimal surface adjoint to HENNEBERG's surface is uniquely determined by the geodesic astroid α and its normals. In any of the above cases, the real parametrization $\mathbf{f}$ of a real minimal surface was found by computing the real part $\mathbf{f} = \mathfrak{Re}\,\varphi(w)$ of some complex vector valued function $\varphi(w)$. The vector valued function $\varphi(w)$ parametrizes an isotropic curve in Euclidean three-space, *i.e.*, a curve with constant slope $\pm\mathrm{i}$. The computation of the real part is equivalent to the addition of the complex conjugate vector function and subsequent multiplication by $\frac{1}{2}$, *i.e.*, $\mathbf{f} = \frac{1}{2}(\varphi + \overline{\varphi}) = \mathfrak{Re}\,\varphi$. This is just the analytical formulation of a fundamental result by LIE (see [27, 30, 33, 36]):

Theorem 8. *Translating an isotropic curve φ (curve of constant slope $\pm\mathrm{i}$) along another isotropic curve ψ sweeps a minimal surface. The minimal surface is real if, and only if, φ and ψ are complex conjugate curves.*

The curve $\varphi(w)$ is an isotropic (minimal) curve of Euclidean geometry. This property is not altered if we multiply $\varphi(w)$ by $\mathrm{e}^{\mathrm{i}\tau}$ prior to the extraction of the real part. The latter multiplication by a complex factor of absolute value 1 is, geometrically speaking, just a rotation of the complex curve. The family of real minimal surfaces given by

$$\mathbf{f}(\tau) = \mathfrak{Re}(\mathrm{e}^{\mathrm{i}\tau}\varphi(w)) = \mathrm{c}_\tau \mathfrak{Re}\,\varphi(w) + \mathrm{s}_\tau \mathfrak{Im}\,\varphi(w) \tag{18}$$

is called the *associate family*. Especially, $\mathbf{f}^\perp := \mathbf{f}(\frac{\pi}{2})$ is called the *adjoint minimal surface* to $\mathbf{f}$. The following theorem is obvious:

Theorem 9. *The family of minimal surfaces associate to an algebraic minimal surface consists only of algebraic minimal surfaces.*

Proof. From (18) we can see that the parametrizations of the minimal surfaces in the associate family are linear combinations of $\mathfrak{Re}\,\varphi(w)$ and $\mathfrak{Im}\,\varphi(w)$ with coefficients c_τ and s_τ. If $\mathbf{f}$ is obtained via (17), then both $\mathbf{f}(0) = \mathfrak{Re}\,\varphi(w)$ and $\mathbf{f}(\frac{\pi}{2}) = \mathfrak{Im}\,\varphi(w)$ are algebraic and so is any of their linear combinations. □

It is elementary to verify that the meromorphic function A from (12) changes to $\mathrm{e}^{\mathrm{i}\tau}A$ and B does not change during the transition from the minimal surface defined by A and B to the members of its associate family.

We shall have a look at the minimal surface adjoint to HENNEBERG's surface (1). A parametrization $\mathbf{f}^\perp$ of this adjoint surface is found by multiplying (10) by $\mathrm{e}^{\mathrm{i}\frac{\pi}{2}} = \mathrm{i}$, reparametrizing by $t = \mathrm{S}_\tau$. Then, $\tau = v + \mathrm{i}u$ and we extract the real part which gives

$$\mathbf{f}^\perp(u,v) = \begin{pmatrix} \mathrm{c}_{3u}\mathrm{S}_{3v} - 3\mathrm{S}_v\mathrm{c}_u \\ \mathrm{s}_{3u}\mathrm{S}_{3v} + 3\mathrm{S}_v\mathrm{s}_u \\ 3\mathrm{C}_{2v}\mathrm{c}_{2u} \end{pmatrix}. \tag{19}$$

The surface (19) has more symmetries than HENNEBERG's surface: It is symetric with respect to the planes

$$x = 0, y = 0, x \pm y = 0, z = 0.$$

The algebraic minimal surface $\mathbf{f}^\perp$ (19) is of degree 26 and a part of it is shown in Fig. 3. Its intersection with the ideal plane is the 18-fold ideal line of all planes parallel to $z = 0$ together with the four-fold pair of ideal lines of complex conjugate isotropic planes.

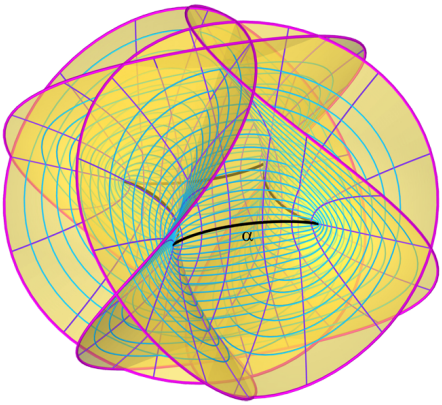


Figure 3: *The minimal surface adjoint to* HENNEBERG*'s surface is uniquely determined by the geodesic astroid* α *and its normals.*

The surface $\mathbf{f}^{\perp}$ intersects the plane $z=0$ along the astroid

$$\alpha(t) = (4c_t{}^3, 4s_t{}^3, 0) \tag{20}$$

that turns out to be a geodesic on the surface $\mathbf{f}^{\perp}$.
On the other hand, α can be taken as the spine curve of the scroll (α, ν) with its unit normals

$$\nu(t) = (-s_t, -c_t, 0). \tag{21}$$

Inserting (20) and (21) into (3), we obtain a parametrization of $\mathbf{f}^{\perp}$ that is slightly different from (19) but equivalent to that. Summarizing this, we can state (a known result, see [22, 23, 33]) in

Theorem 10. *The adjoint minimal surface to* HENNEBERG*'s minimal surface carries a geodesic astroid* α*. The adjoint to* HENNEBERG*'s minimal surface is the uniquely determined minimal surface on the scroll* (α, ν) *with* ν *being* α*'s unit normal vector field.*

It is noteworthy that the astroid α (20) is a hypocycloid. This will be of importance in Sec. 7.

3 ENNEPER's surfaces

There is not just one Enneper surface even if we don't mention equiform copies of the standard form. The well-known example

$$\mathcal{E}_1(u,v) = \begin{pmatrix} -\frac{1}{3}u^3 + uv^2 + u \\ \frac{1}{3}v^3 - u^2v - v \\ u^2 - v^2 \end{pmatrix} \tag{22}$$

with its bi-cubic parametrization is one in a one-parameter family of algebraic minimal surfaces that admit even polynomial parametrizations. It can be found with (13) by letting $G=1$ and $H=w$ or with (17) where $A=\frac{1}{6}z^3$.

The algebraic degree of the classical Enneper surface equals nine since an implicit equation can be given by

$$\begin{aligned}&[9(y^2-x^2)+4z(z^2+3)]^3 - \\ &-27z[9(y^2-x^2)-z(9(x^2+y^2)+8z^2)+8z]^2 = 0.\end{aligned} \tag{23}$$

The class of ENNEPER's surface equals six as can be read off from the implicit equation of its family of tangent planes

$$\begin{aligned}&w_0^2(w_1^2+w_2^2)^2 - 3(w_1^2-w_2^2)^2w_3^2 - 4(w_1^2-w_2^2)^2(w_1^2+w_2^2) + \\ &+2w_0w_3(w_1^2-w_2^2)\cdot(3w_1^2+3w_2^2+2w_3^2) = 0.\end{aligned} \tag{24}$$

ENNEPER's minimal surface is an example of a non-orientable minimal surface with *even* class, cf. Thm. 5.
The term of degree nine in (23) equals z^9 which shows that the ideal line of all planes parallel to $z=0$ comprises the set of ideal points of ENNEPER's surface.
According to LIE [30], the sum of the class and the degree of an algebraic minimal surface is at least 15, and thus, ENNEPER's surface is the confirming example. Its real self-intersection consists of the pair

$$\begin{aligned}s_1 &= (0, \tfrac{3}{8}t(3t^2+8), \tfrac{9}{8}t^2+3), \\ s_2 &= (-\frac{3}{8}t(3t^2+8), 0, -\frac{9}{8}t^2-3)\end{aligned}$$

of polynomial cubic curves (semi-cubic parabolas) in the symmetry planes $x=0$ and $y=0$.
The more general version of ENNEPER's surface is given by

$$\mathcal{E}_n(u,v) = \mathfrak{Re}\begin{pmatrix} w - \dfrac{w^{2n+1}}{2n+1} \\ \mathrm{i}w + \dfrac{\mathrm{i}w^{2n+1}}{2n+1} \\ \dfrac{2w^{n+1}}{n+1} \end{pmatrix} \tag{25}$$

where $n \in \mathbb{N}\setminus\{0\}$ is usually called the *order of the Enneper surface*. These minimal surfaces are obtained from (13) with

$$G(w) = 1 \quad \text{and} \quad \mathrm{H(w)} = \mathrm{w^n}.$$

With $n=1$ we obtain the *classical* minimal surface by ENNEPER parametrized by (22) first given in [8].
Dropping the restriction $n \in \mathbb{N}\setminus\{0\}$, we obtain the plane $x=0$, *i.e.*, a flat minimal surface if $n=0$. The case $n=-1$ is still to be excluded if one is interested in algebraic minimal surfaces. However, the case $n=-1$ yields the catenoid

$$2\mathrm{C}_{\frac{z}{2}} = \sqrt{x^2+y^2}.$$

Surprisingly, the case $n=-2$ yields RICHMOND's surface (31), which will be discussed in Sec. 5. The surface $\mathcal{E}_3(u,v)$ is displayed in Fig. 4.

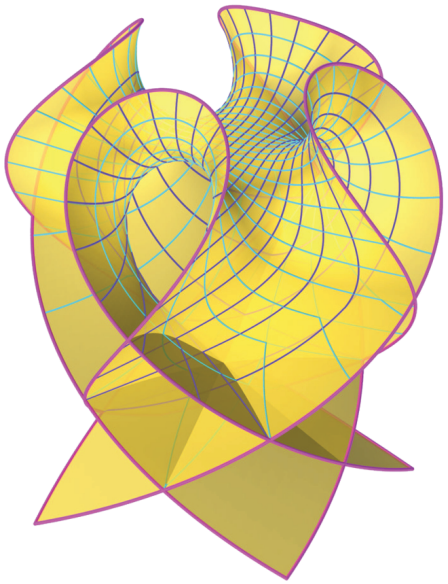


Figure 4: *The Enneper surface of order 3 is of algebraic degree 49, cf. Thm 11.*

We can give an upper bound on the algebraic degree and a precise value for the class of the Enneper surfaces of an arbitrary order:

Theorem 11. *Enneper surfaces of order $n \in \mathbb{N}\setminus\{0\}$ are algebraic minimal surfaces whose degree is at most $(2n+1)^2$ and its class equals $2n(2n+1)$.*

Proof. The polynomial parametrization of an Enneper surface (25) of order n is of bi-degree $(2n+1, 2n+1)$. Elimination of u and v from the coordinate functions means computing resultants with respect to u and v. Thus, the algebraic degree of $\mathcal{E}_n$ is at most $(2n+1)^2$.
In order to show that the class equals $2n(2n+1)$, we use a result by LIE (cf. [30, vol. 1, p. 315]): The rank of the isotropic curve (25) equals $r = 3n+1$ and the multiplicity of the absolute conic as a curve on the tangent developable this particular isotropic curve equals $\mu = n$. According to LIE, the class of the minimal surface generated by the isotropic curve (25) equals $2\mu(r-\mu) = 2n(3n+1-n) = 2n(2n+1)$. □

The computation of the implicit equations of the surfaces $\mathcal{E}_n$ up to $n = 7$ shows that the bound $\deg \mathcal{E}_n = (2n+1)^2$ is sharp at least in these cases.

4 BOUR's surfaces

The minimal surfaces by E. BOUR (see [4]) are characterized by allowing local isometries to surfaces of revolution. Parametrizations of the surfaces in this one-parameter family are obtained from (12) by inserting

$$A(w) = cw^{m-2},\ c \in \mathbb{C}\setminus\{0\}, m \in \mathbb{R}\setminus\{0\} \tag{26}$$

and $B(w) = w$. Alternatively, we can use

$$G = \sqrt{c}w^{\frac{m}{2}-1} \quad \text{and} \quad H = \sqrt{c}w^{\frac{m}{2}}$$

together with (13). With (26) and (12) we arrive at the parametrization

$$\mathcal{B}_m(u,v) = \mathfrak{Re}\, c \cdot \begin{pmatrix} \frac{1}{m-1}w^{m-1} - \frac{1}{m+1}w^{m+1} \\ \frac{\mathrm{i}}{m-1}w^{m-1} + \frac{\mathrm{i}}{m+1}w^{m+1} \\ \frac{2}{m}w^m \end{pmatrix}. \tag{27}$$

We call m the *order of the Bour surface* $\mathcal{B}_m$.
It means no restriction to assume $|c| = 1$, *i.e.*, $c = \mathrm{c}_\tau + \mathrm{i}\mathrm{s}_\tau$ since the multiplication of A by c causes only a scaling of the respective minimal surface with the scaling factor $|c|$. On the other hand, the multiplication with any complex number $c = \mathrm{c}_\tau + \mathrm{i}\mathrm{s}_\tau$ (with $\tau \in S^1$) corresponding to a point on the Euclidean unit circle chooses one certain member of the family of minimal surfaces associate to $\mathcal{B}_m$.
Well-known and non-algebraic minimal surfaces can be found among the surfaces by BOUR: $m = 0$, $c = 1$ lead to the catenoid; the choice $m = 0$, $c = \mathrm{i}$ results in the helicoid

$$2\arctan\frac{x}{y} = z$$

which is adjoint to the catenoid. If $m = \pm 1$ the resulting minimal surfaces are not algebraic independent of c, but they seem to be worth a closer inspection. A part of this non-algebraic minimal surface is displayed in Fig. 5.

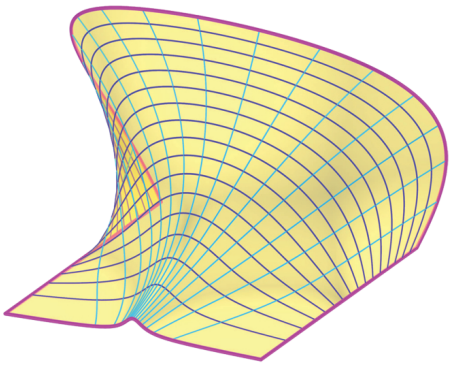


Figure 5: *The non-algebraic minimal surface $\mathcal{B}_{-1}$.*

BOUR's minimal surfaces are algebraic if, and only if, $m \in \mathbb{Q}\setminus\{-1, 0, -1\}$. The following result makes clear that negative m can be excluded from our considerations:

Lemma 1. *For any $m \in \mathbb{Z}\setminus\{-1,0,1\}$ we have $\mathcal{B}(m) = \mathbf{S}\cdot\psi(\mathcal{B}(-m))$ where $\mathbf{S} = \mathrm{diag}(1,-1,-1)$ is the matrix describing the reflection in the x-axis and ψ is the reparametrization*

$$u = \frac{U}{U^2+V^2},\ v = -\frac{V}{U^2+V^2}, \tag{28}$$

or equivalently, $\psi : w = u + \mathrm{i}v \mapsto W^{-1}$ *(with* $W = U + \mathrm{i}V$*) which is the inversion in the Euclidean unit circle in the parameter plane.*

Proof. Let $m < -1$. We observe the changes

$$w^{m-1} \to w^{-n-1}, \quad w^{m+1} \to w^{-n+1}$$

with $m = -n$. Then, we reparametrize with ψ according to (28) and the latter powers of w change again:

$$w^{-n-1} \to (W^{-1})^{-n-1} = W^{n+1},$$
$$w^{-n+1} \to (W^{-1})^{-n+1} = W^{n-1},$$

both with positive n. Thus, the second and third coordinate function change their sign and $\mathbf{S} = \mathrm{diag}(1,-1,-1)$. Finally, changing $U \to u$ and $V \to v$ simplifies the comparison of the parametrizations. □

Especially, the surfaces for $m = 2,3,4,5$ are of relatively low degree. ENNEPER's minimal surface corresponds to $m = \pm 2$ with arbitrary c.

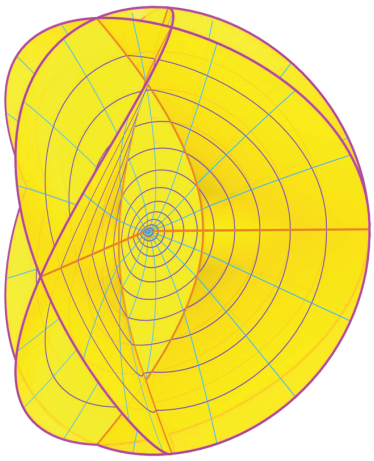


Figure 6: BOUR*'s surface of order 3 is a Bézier minimal surface of algebraic degree 16 and of class 8.*

With $m = 3$ we find a minimal surface of degree 16 and class 8 which is displayed in Fig. 6. The surface has three planes of symmetry: $y = 0$ and $3x^2 = y^2$ whose intersections with the plane $z = 0$ are three straight lines concurrent in the point $(0,0,0)$ which lie entirely in the surface. All three lines turn out to be four-fold lines on the surface. The Bour minimal surface of order 3 meets the ideal plane in the ideal line of all planes parallel to $z = 0$ with multiplicity 16. The planar rational (polynomial) quartic PH-curve (cf. [9])

$$\gamma(t) = \left(-\frac{1}{4}t^4 + \frac{1}{2}t^2, 0, \frac{2}{3}t^3\right)$$

together with its normal vectors can be used to construct a parametrization of this minimal surface with the Björling formula (3). Therefore, γ is a geodesic on the surface.
If now $m = \pm 4$, we obtain an algebraic minimal surface of degree 25 and class 10. The four lines

$$(x^2 - 2xy - y^2)(x^2 + 2xy - y^2) = 0$$

are five-fold lines on this minimal surface. With the Björling formula (3) the two planar and congruent PH-curves

$$\gamma_1 = \left(0, -\frac{1}{5}t^5 + \frac{1}{3}t^3, \frac{1}{2}t^4\right),$$
$$\gamma_2 = \left(-\frac{1}{5}t^5 + \frac{1}{3}t^3, 0, \frac{1}{2}t^4\right)$$

in the planes $x = 0$ and $y = 0$ together with their rational normals also define the Bour minimal surface of order 4. Both curves, γ_1 and γ_2 are planar geodesics on the Bour surface of order 4 and the plane $z = 0$ is a plane of symmetry. Again, the intersection with the ideal plane is a line whose multiplicity equals the algebraic degree of the surface.
The above given examples show that BOUR's minimal surfaces can also be obtained as minimal surfaces on PH-scrolls as a solution to Björling's problem. In a more general version, we have

Theorem 12. *The minimal surfaces on the scroll* (γ, ν) *with*

$$\gamma(t) = \left(\tfrac{-1}{m+1}t^{m+1} + \tfrac{1}{m-1}t^{m-1}, \tfrac{2}{m}t^m, 0\right) \tag{29}$$

where $m \geq 2$ *and*

$$\nu(t) = \frac{1}{1+t^2}(-2t, 1-t^2, 0) \tag{30}$$

are BOUR*'s minimal surfaces of order m up to equiform transformations.*

Proof. We insert (29) and (30) into (3) and arrive at (27). Note that ν from (30) satisfies $\nu = w^{m-2}\gamma'^{\perp}$. □

We can give an upper bound on the algebraic degree and class of BOUR's minimal surfaces of order m in

Theorem 13. *The algebraic degree and the class of* BOUR*'s minimal surface of order m are equal to* $(m+1)^2$ *and* $2(m+1)$ *provided that* $m \geq 2$.

Proof. We use the same arguments as in the proof of Thm. 11. □

Like the generalized Enneper surfaces (25), the Bour surfaces (27) are Bézier minimal surfaces (as long as they are algebraic).

5 RICHMOND's surfaces

The original Richmond surface (as shown in Fig. 7) comes along as one special example in a one-parameter family of minimal surfaces. It has the simple parametrization

$$\mathbf{f}(u,v) = \begin{pmatrix} \frac{1}{3}u^3 - uv^2 + \frac{u}{u^2+v^2} \\ \frac{1}{3}v^3 - u^2v - \frac{v}{u^2+v^2} \\ 2u \end{pmatrix}. \tag{31}$$

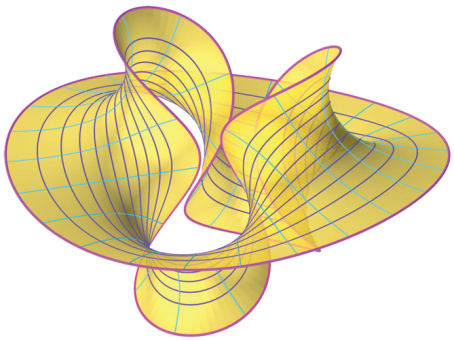


Figure 7: *Richmond's minimal surface of degree 12 and class 12.*

RICHMOND's surface is the only real algebraic minimal surface of degree 12 up to equiform transformations, see [33]. The class of RICHMOND's surface equals 12, not 17 as RICHMOND stated in [39] (This was corrected in [40].)
The minimal surfaces associated to RICHMOND's surface (31) are just similar copies of that surface, see [39].
When using (12) in order to parametrize the surface, we have to insert

$$A(w) = \frac{1}{w^2}, \quad B(z) = w^2.$$

RICHMOND's minimal surface can also be constructed as a minimal surface on a scroll: Use the planar curve

$$\gamma(t) = \left(\frac{1}{3}t^3 + \frac{1}{u}, 0, 2t\right) \tag{32}$$

for the spine curve with unit normals

$$\nu(t) = \frac{1}{1+t^2}\left(-2t, 0, t^2-1\right) \tag{33}$$

along γ and insert both into (3). The unit normal vector field of the curve γ from (32) is not precisely that given by (33) but can be transformed by the reparametrization $t \to \sqrt{t}$ into (33). Note that the plane $y = 0$ that contains γ is a plane of symmetry of RICHMOND's minimal surface and γ is a planar geodesic of the surface.
More generally speaking, associated to the family of curves

$$\gamma_a(t) = \left(t^3 + \frac{a^2}{2t}, 0, 2at\right)$$

with $a \in \mathbb{R} \setminus \{0\}$ and the unit normal vector field

$$\nu_a(t) = \left(\frac{-6at^2}{a^2+9t^4}, 0, \frac{a^2-9t^4}{a^2+9t^4}\right)$$

there is a one-parameter family of rational, and thus, algebraic minimal surfaces of Richmond type whose parametrizations read

$$\mathcal{R}(a,u,v) = \begin{pmatrix} u^3 - 3uv^2 + \frac{1}{12}\frac{a^2u}{u^2+v^2} \\ 3u^2v - v^3 + \frac{1}{12}\frac{a^2v}{u^2+v^2} \\ au \end{pmatrix}.$$

The generalization is straight forward. We choose

$$A(w) = \frac{1}{w^2} \quad \text{and} \quad \mathrm{B(w)} = \mathrm{w^{m+1}} \tag{34}$$

with $m \in \mathbb{N} \setminus \{0\}$ which yields a one-parameter family of minimal surfaces when inserted into (12). We shall call m the *order of the Richmond surface* $\mathcal{R}_m$.
Figure 8 shows two Richmond surfaces: one of order 3, the other one of order 4.
Note that A has a pole of degree 2 at $w = 0$. Especially, the surface with $m = 1$ is given by (31). Again, we observe that replacing m by $-m$ results in the same surface. So it is sufficient to consider only positive m.
It is no surprise that the family of generalized Richmond minimal surfaces contains members of other families. For example $\mathcal{R}_1 = \mathcal{B}_1$ with $c = 1$.
Alternatively, we could use the representation (13) with

$$G(w) = \frac{1}{w}, \; H(z) = w^m.$$

The parametrizations of the generalized Richmond surfaces read

$$\mathcal{R}_m(u,v) = \mathfrak{Re}\begin{pmatrix} -\frac{1}{w} - \frac{w^{2m+1}}{2m+1} \\ -\frac{\mathrm{i}}{w} + \frac{\mathrm{i}w^{2m+1}}{2m+1} \\ -\frac{2w^m}{m} \end{pmatrix} \tag{35}$$

and they make clear that these are algebraic surfaces that admit even a rational parametrization.
We can give an upper bound for the algebraic degree of the generalized Richmond surfaces:

Theorem 14. *The generalized Richmond surfaces of order* $m \in \mathbb{N} \setminus \{0\}$ *are at most of algebraic degree* $2(m+1)(2m+1)$. *The class of the generalized Richmond surfaces equals exactly* $2(m+1)(2m+1)$.

Proof. For the proof of the upper bound of the degree, we use similar arguments as in the proofs of Thm. 11 and Thm. 13.
In order to verify the formula for the class of the generalized Richmond surfaces, we use the results from [30, vol. 1, p. 315] and compute, like in the proof of Thm. 11: $r=3m+2$ and $\mu=m+1$ which yields the class $2\mu(r-\mu)=2(m+1)(3m+2-m-1)=2(m+1)(2m+1)$. □

The regular reparametrization

$$u = r\mathrm{c}_s, \quad v = r\mathrm{s}_s$$

changes (35) to

$$\mathcal{R}_m(r,s)=\begin{pmatrix} -\frac{r^{2m}}{m+1}\mathrm{c}_{(m+1)s}-\frac{1}{(m-1)r}\mathrm{c}_{(m-1)s} \\ -\frac{r^{2m}}{m+1}\mathrm{s}_{(m+1)s}-\frac{1}{(m-1)r}\mathrm{s}_{(m-1)s} \\ 2r\mathrm{c}_s \end{pmatrix} \tag{36}$$

which is not just favorable for plotting the surface. It also enables us to show

Theorem 15. *The Richmond minimal surfaces* (35) *with* $m \in \mathbb{Q}\setminus\{-1,0,1\}$ *carry a one-parameter family of harmonic oscillation curves of order two.*

Proof. Let the first and the second coordinate function be the real and the imaginary part of a complex number and build $w = x + \mathrm{i}y$. Then, apply EULER's formula and find

$$w(s) = -\frac{r^{2m}}{m+1}\mathrm{e}^{\mathrm{i}(m+1)s} - \frac{1}{m-1}\mathrm{e}^{\mathrm{i}(m-1)s}.$$

If $r \in \mathbb{R}\setminus\{0\}$ is fixed, then, according to [55, 56], $w(s)$ is a complex parametrization of an ordinary cycloidal curve. Finally, we observe that the third coordinate function $z = 2r\cos s$ is periodic for any $r \in \mathbb{R}$. Thus, the s-lines on the surface (36), *i.e.*, the curves with fixed r are higher oscillation curves in the sense of [37]. □

By assumption, $m \in \mathbb{Q}$, and thus, the curves are closed.

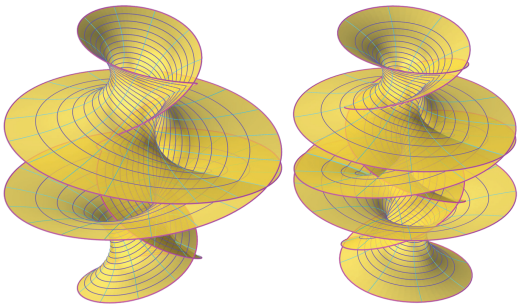


Figure 8: *Minimal surfaces of Richmond type: Left:* $m = 3$ *of algebraic degree 56; right:* $m = 4$ *of algebraic degree 90.*

6 Minimal surfaces tangent to orthogonal hyperbolic paraboloids

We consider the one-parameter family of hyperbolic paraboloids

$$\mathcal{P}: \ (1-b^2)xy = 2bz \tag{37}$$

with $b \in \mathbb{R}\setminus\{-1,0,1\}$ and the cylinder of revolution

$$\mathcal{Z}: \ x^2+y^2 = 1. \tag{38}$$

The cylinder intersects the paraboloids (37) along the rational quartic space curves

$$\gamma(t) = \left(\mathrm{c}_t, \mathrm{s}_t, \frac{1-b^2}{4b}\mathrm{s}_{2t}\right). \tag{39}$$

In the following, we use the abbreviations

$$\beta_1 := 1+b^2, \ \beta_2 := 1-b^2, \ \beta_3 := b^4+6b^2+1.$$

Let now the normal vector field be given by

$$\nu(t) = \mathrm{grad}(P)|_\gamma = \frac{1}{\beta_1}(\beta_2\mathrm{s}_t, \beta_2\mathrm{c}_t, -2b). \tag{40}$$

Then, we insert γ and ν from (39) and (40) into (3) and find the parametrizations of the minimal surfaces in the one-parameter family of minimal surfaces touching the paraboloids (37) along their intersection with $\mathcal{Z}$. From their parametrizations

$$\mathbf{f}(u,v) = \frac{1}{12b\beta_1}\begin{pmatrix} \beta_2^2\mathrm{c}_{3u}\mathrm{S}_{3v}+3\mathrm{c}_u(\beta_3\mathrm{S}_v+4b\beta_1\mathrm{C}_v) \\ -\beta_2^2\mathrm{s}_{3u}\mathrm{S}_{3v}+3\mathrm{s}_u(\beta_3\mathrm{S}_v+4b\beta_1\mathrm{C}_v) \\ 3\beta_2\mathrm{s}_{2u}(\beta_1\mathrm{C}_{2v}+2b\mathrm{S}_{2v}) \end{pmatrix}, \tag{41}$$

we can immediately see that these surfaces admit rational parametrizations of bi-degree (6,6). Figure 9 shows the minimal surface parametrized by (41) together with the hyperbolic paraboloid, the curve γ from (39), and the unit normal vector field ν as given in (40). Moreover, Fig. 9 gives an idea how the minimal surface tangent to a hyperbolic paraboloid deviates from the paraboloid.

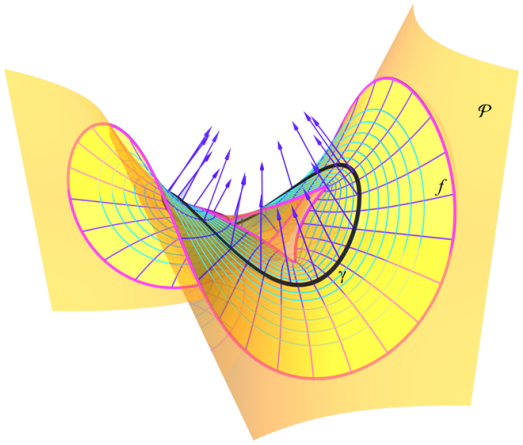


Figure 9: *The minimal surface* (41) *on the scroll* (γ,ν). *The curve* γ *is a curve along which the Gaussian curvature on the hyperbolic paraboloid* $\mathcal{P}$ *is constant.*

The rational representation of these minimal surfaces allows us to compute an implicit equation of each surface in the family. Hereby, we find that all minimal surfaces (41) are algebraic surfaces of degree 30. They all have the cycle $z^{18}(x^2+y^2)^6=0$ in the ideal plane in common.
The curves of constant Gaussian curvature K on the hyperbolic paraboloid P given in (37) lie on cylinders of revolution coaxial with the one in (38). For any $b\in\mathbb{R}\setminus\{-1,0,1\}$ these cylinders have the equation

$$x^2+y^2=\frac{1}{\beta_2\sqrt{-K}}-\frac{4b^2}{\beta_2^2} \tag{42}$$

for any admissible value $K<K_{\max}=-\frac{\beta_2^2}{16b^4}$. From

$$r^2=\frac{1}{\beta_2\sqrt{-K}}-\frac{4b^2}{\beta_2^2}$$

we can determine the cylinder's radius. Conversely, we can choose b such that the radius r corresponds to a certain value K. This gives rise to

Theorem 16. *The minimal surfaces that touch an orthogonal hyperbolic paraboloid* $\mathcal{P}$ *along the curves of constant Gaussian curvature on* $\mathcal{P}$ *are rational (and thus algebraic) minimal surfaces and can be parametrized by* (41). *These minimal surfaces are of degree 30 and of class 10.*
The parameter curves $v=$ const. *are rational (and thus algebraic and closed) oscillation curves of order two,* i.e., *their orthogonal projections onto* $z=0$ *are cycloidal curves of order two and the z-coordinate function is harmonic.*

Proof. From the parametrization (41) we can derive an implicit equation after a rational substitution of the trigonometric and hyperbolic functions. Thus, the rationality is obvious and the degree turns out to be 30. From the parametrization of the set of points (41), we can derive a parametrization of the set of tangent planes. Eliminating the paramters yields a polynomial of degree 10 and so the class of the minimal surface equals 10.
The x- and the y-coordinate can be considered the real and the imaginary part of a complex variable. Thus, for fixed $v\in\mathbb{R}$, we have $w(u)=x+\mathrm{i}y$ which gives a complex representation of the *top view* of the parameter curves:

$$w(u)=\frac{\beta_2^2 S_{3v}}{12b\beta_1}e^{-3\mathrm{i}u}+\frac{\beta_3 S_v+4b\beta_1 C_v}{4b\beta_1}e^{\mathrm{i}u}.$$

Comparing the latter with the formulae given in [55, 56], we can see that these are the path curves of the end points of open two-bar mechanisms. The ratio of the angular velocities of the rotating bars equals $-3:1$ and the lengths of the legs are the absolute values of the coefficients of the exponential functions. The z-coordinate is a multiple of $\sin 2u$, and thus, harmonic. □

Finally, we shall mention that meromorphic functions A,B: $\mathbb{C}\to\mathbb{C}$ in the Weierstraß-representation (12) are

$$A=\frac{(1+b)^2}{8\mathrm{i}b\beta_1}\left(e^{3\mathrm{i}w}(1-b)^2+e^{-\mathrm{i}w}(1+b)^2\right) \tag{43}$$

and the simple function

$$B=\mathrm{i}\frac{1-b}{1+b}e^{-\mathrm{i}w}. \tag{44}$$

Since $b\in\mathbb{R}\setminus\{-1,0,1\}$, the function B can never vanish. The substitution $t=e^{\mathrm{i}w}$ in (43) transforms A into a rational function. Together with B from (44) which is linear anyway, and thus, also rational, we can find the minimal surfaces from Thm. 16 via (12) with rational generators A and B.
The associate minimal surfaces show a surprising behavior:

Theorem 17. *The minimal surfaces associated to* (41) *are congruent to* f. *Traversing the associate family of minimal surfaces means rotating the original one about the z-axis.*

Proof. Derive the parametrization or implicit equation of the surfaces in the associate family. The congruence transformation can easily be read of from the parametrization. □

Consequently, all members of the associate family of the minimal surface f given by (41) have the same algebraic properties.
The minimal surfaces (41) intersect the hyperbolic paraboloid $\mathcal{P}$ with equation (37) in the lines $x=z=0$ and $y=z=0$, each with multiplicity 6 and along the curve γ with multiplicity two (according to the construction).

7 Minimal surfaces with geodesic cycloids

Thm. 1 and Thm. 10 give rise to a generalization of HENNEBERG's adjoint surface which was the minimal surface on a scroll with an astroid (20) for its spine curve. Here, we shall recall that there is a notion of cycloid that shall not be of use here: Frequently, the word cycloid is used for a curve that is generated by rolling a circle on a straight line, see [29, 32, 51]. The minimal surface with this *straight cycloid* as a planar geodesic is known as CATALAN's minimal surface (see [27, 33, 36] and it is not algebraic.
The cycloidal curves that emerge from rolling a circle along another one yields a one-parameter family of rational, and thus, algebraic minimal surfaces. We have

Theorem 18. *Let $r,R \in \mathbb{R}\setminus\{0\}$ be real constants with $R+2r \neq 0$ and $R+r \neq 0$. The minimal surfaces on the scroll (ζ,ν) with $\zeta \subset \pi_3 : z=0$ and $\nu \in \mathrm{S}^1$*

$$\begin{aligned}\zeta(t) &= \begin{pmatrix} (R+r)\mathrm{c}_t + r\mathrm{c}_{\frac{(R+r)t}{r}} \\ (R+r)\mathrm{s}_t + r\mathrm{s}_{\frac{(R+r)t}{r}} \\ 0 \end{pmatrix}, \\ \nu(t) &= \frac{1}{2\mathrm{c}_{\frac{Rt}{2r}}} \begin{pmatrix} -\mathrm{c}_t - \mathrm{c}_{\frac{(R+r)t}{r}} \\ -\mathrm{s}_t - \mathrm{s}_{\frac{(R+r)t}{r}} \\ 0 \end{pmatrix}\end{aligned} \tag{45}$$

can be parametrized by

$$\mathbf{f}(u,v) = \begin{pmatrix} (R+r)\mathrm{c}_u\mathrm{C}_v + r\mathrm{c}_{\frac{(R+r)u}{r}}\mathrm{C}_{\frac{(R+r)v}{r}} \\ (R+r)\mathrm{s}_u\mathrm{C}_v + r\mathrm{s}_{\frac{(R+r)u}{r}}\mathrm{C}_{\frac{(R+r)v}{r}} \\ -\frac{4r(R+r)}{R}\mathrm{c}_{\frac{Ru}{2r}}\mathrm{S}_{\frac{Rv}{2r}} \end{pmatrix}. \tag{46}$$

These minimal surfaces are algebraic, rational, and closed if, and only if, $R,r \in \mathbb{Q}\setminus\{0\}$.
In any case, the cycloid $\zeta \subset \pi_3$ is a geodesic on the minimal surface.
The surfaces with $R,r \in \mathbb{Q}\setminus\{0\}$ contain at least one straight line.

Proof. Insert γ and ν from (45) into (3). This gives (46).
The geodesic property of the cycloidal spine curves is a direct consequence of Thm. 1.
The straight lines are part of the double curves in symmetry planes. □

In the case $R+2r=0$, the cycloid ζ from (45) collapses to a diameter of the circle $(R\mathrm{c}_t, R\mathrm{s}_t, 0)$. If $R+r=0$, the polhodes of ζ are not just congruent, they are identical and no rolling takes place.
The cycloids ζ parametrized by (45) are closed, rational, and thus, algebraic, if, and only if, $r : R \in \mathbb{Q}\setminus\{0\}$. They have cusps of the first kind at

$$\cos\frac{tR}{r} = -1 \iff t = (2k+1)\pi\frac{r}{R},$$

i.e., finitely many if $r : R \in \mathbb{Q}\setminus\{0\}$, provided the admissible choice of r and R. Consequently, the minimal surfaces (46) have branch points exactly at the cusps of the cycloids ζ given by (45).
From the parametrization (46) it is clear that the u-lines (curves with $v=$ const.) on the cycloidal minimal surfaces have a very special shape. We have

Theorem 19. *The u-lines on the cycloidal minimal surfaces given by* (46) *are generalized oscillation curves. Their orthogonal projections onto the planes $z=c$ (with $c \in \mathbb{R}$) are cycloidal curves.*

Proof. A closer look at the first and second coordinate function of the parametrization (46) tells us that, for fixed $v \in \mathbb{R}$, we have the parameterization of cycloidal curves. These curves can also be written in terms of complex coordinates by letting $w(u) = x + \mathrm{i}y$ and applying EULER's formula as

$$w(u) = (R+r)\mathrm{C}_v\mathrm{e}^{\mathrm{i}u} + r\mathrm{C}_{\frac{(R+r)t}{r}}\mathrm{e}^{\mathrm{i}u\frac{R+r}{r}}.$$

Comparing with [56], we find the lengths

$$A_1 = (R+r)\mathrm{C}_v, \;\; A_2 = r\mathrm{C}_{\frac{(R+r)v}{r}}$$

of the legs of a generating two-bar mechanism and the (ratio of the) angular velocities of the bars are

$$\omega_1 : \omega_2 = 1 : \frac{R+r}{r}.$$

From that we can compute the radii of the polhodes of the motion that generates the orthogonal projections of u-lines as path curves, see [55, 56]. □

The meromorphic functions $A,B : \; D \subset \mathbb{C} \to \mathbb{C}$ from (13) can also be given:

Lemma 2. *The cycloidal minimal surfaces can be obtained from the Weierstraß-representation* (13)*. Therein, the meromorphic functions A and B are:*

$$\begin{aligned} A(w) &= -\tfrac{\mathrm{i}}{2}(R+r)\left(\mathrm{e}^{-\mathrm{i}w} + \mathrm{e}^{-\mathrm{i}\frac{R+r}{r}w}\right), \\ A(w)\cdot B(w) &= \mathrm{i}(R+r)\mathrm{c}_{\frac{(R+r)w}{2r}}. \end{aligned} \tag{47}$$

Proof. In order to find A and B from (46), we use (14). □

More ore less surprisingly, there is a connection to the curves of constant slope on quadrics of revolution and the curves $\gamma(u) = \mathbf{f}(u,0)$ on the cycloidal minimal surfaces. The family of minimal surfaces associated to (46) can be given with (18) as

$$\mathbf{f}(u,v,\tau) = \mathrm{c}_\tau\cdot\mathbf{f} + \mathrm{s}_\tau\cdot\mathbf{f}^\perp \tag{48}$$

where $\mathbf{f}$ is the parametrization (46) and $\mathbf{f}^{\perp}$ reads

$$\mathbf{f}^{\perp} = \begin{pmatrix} (R+r)\mathrm{S}_v\mathrm{s}_u + r\mathrm{S}_{\frac{(R+r)v}{r}}\,\mathrm{s}_{\frac{(R+r)u}{r}} \\ -(R+r)\mathrm{S}_v\mathrm{c}_u - r\mathrm{S}_{\frac{(R+r)v}{r}}\,\mathrm{c}_{\frac{(R+r)u}{r}} \\ -\frac{4r(R+r)}{R}\,\mathrm{s}_{\frac{Ru}{2r}}\,\mathrm{C}_{\frac{Rv}{2r}} \end{pmatrix}. \tag{49}$$

The spine curves of the scrolls are obtained by substituting $v = 0$ in (48). These spine curves can be taken as the spine curve γ of a scroll on which, according to the Björling formula (3), minimal surfaces can be errected. Now, we have the following

Theorem 20. *The one-parameter family of curves* $\mathbf{f}(u,0,\tau)$ *with parametrization* (48) *and* (49) *are curves of constant slope on quadrics of revolution. These curves are closed, rational, and thus, algebraic spacecurves provided that* $r : R \in \mathbb{Q}\setminus\{0\}$, $R+2r \neq 0$, *and* $R+r \neq 0$. *The slope angle* σ *is independent of R and r and is related to* τ *(modulo* 2π*) by*

$$\mathrm{c}_\sigma = -\mathrm{s}_\tau \iff \sigma = \tau + \frac{\pi}{2}. \tag{50}$$

Proof. The top views of the curves $\mathbf{b} = f(u,0,\tau)$, *i.e.*, the orthogonal projections of the curves $f(u,0,\tau)$ onto planes parallel to $z = 0$ are cycloids (with cusps). It is well-known (see, *e.g.*, [3, 7]) that the curves of constant slope on quadrics of revolution appear as epi-, hypo-, hyper, and paracycloids in a top view (in the direction of the lead). The case of paraboloids of revolution differs a little bit: In the corresponding top views, we can see the involutes of circles, cf. [26].

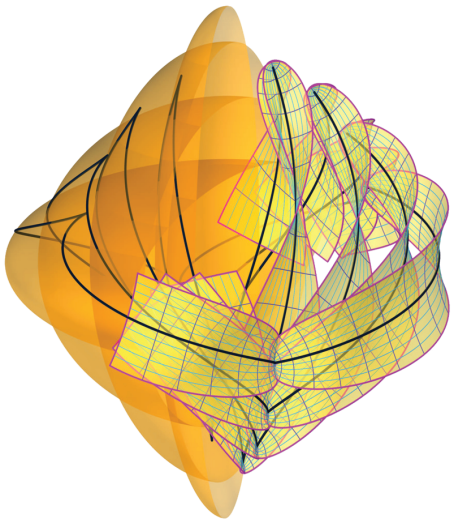


Figure 10: *The spine curves of the cycloidal minimal surfaces are bent smoothly into curves of constant slope on quadrics of revolution.*

We compute $\mathbf{b}' = \frac{\mathrm{d}}{\mathrm{du}}\mathbf{b}$. The lead is given by the unit vector $\mathbf{l} = (0,0,1)$. Now, it is elementary to verify that

$$\mathrm{c}_\sigma = \frac{\langle \mathbf{b}', \mathbf{l}\rangle}{\|\mathbf{b}'\|} = -\mathrm{s}_\tau$$

which makes clear that the slope of the spine curves $\mathbf{b} = f(u,0,\tau)$ is constant and independent of the choice of R and r and (50) is valid. It is easily verified that the coordinate functions of $\mathbf{b}$ satisfy

$$Q : x^2 + y^2 + \frac{k^2R^2z^2}{4r(r+R)} = (2r+R)^2\mathrm{c}_\tau{}^2 \tag{51}$$

with $k = \cot\tau$ which is the equation of quadrics Q of revolution.

The rationality is clear if $r : R \in \mathbb{Q}\setminus\{0\}$ since then $\cos nu$ and $\sin nu$ can be expressed in $\cos u$ and $\sin u$ which can subsequently be replaced with their rational equivalents provided that $(R+r)/r = n$ is an integer. If $(R+r)/r = m/n$ with $\gcd(m,n) = 1$, we reparametrize by letting $u' = ru$, expand $\cos mu, \ldots$ in $\sin u$ and $\cos u$ followed by the rational reparametrization. Since cycloids are closed if $r : R \in \mathbb{Q}\setminus\{0\}$, the curves of constant slope on the quadrics (51) are also closed. □

Figure 10 illustrates the contents of Thm. 20.

We shall note (51) can be the equation of an ellipsoid or a one-sheeted hyperboloid as well. The latter appears if $r < 0$. Two-sheeted hyperboloids will not be described by (51) since then the coefficient of z^2 as well as the right-hand side of (51) have to be negative. This is not possible since the right-hand side is a full square.

On the other hand, the top-views of the curves of constant slope on a two-sheeted hyperboloid of revolution are *paracycloids*, *i.e.*, curves that belong to the class of spiraloids and are transcendental independent of r and R are, cf. [29, 32, 51, 54]. In the case that (51) describes a one-sheeted hyperboloid, k, and thus, the slope of the curves $\mathbf{b}$ is always larger than that of the quadrics' asymptotic cone. Otherwise the curves of constant slope appear as *hypercycloids* in the top-view. These curves are closely related to paracycloids, and like these, they are always transcendental and belong to the class of spiraloids, see [29, 32, 51, 54].

7.1 A cardioid as a geodesic curve

The low degree minimal surfaces of cycloidal type can be found by choosing small values for the radii R and r of the polhodes of the cycloid ζ. The case of an astroid which occurs with $r : R = -1 : 4$ is described in Sec. 2, especially in Thm. 10.

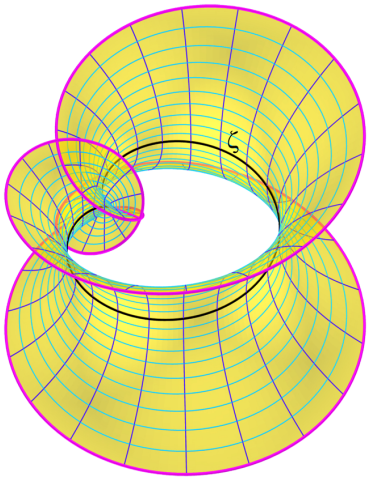


Figure 11: *The cycloidal minimal surface with $R = r = 1$ and its geodesic cardiod ζ.*

Figure 11 shows the algebraic minimal surface along the cardiod ζ. This surface occurs with $r : R = 1 : 1$. The algebraic degree of the *cardioidal minimal surface* is 20 and the class equals 36. The intersection μ of the minimal surface with the plane at infinity has the equation $z^{16}(x^2+y^2)^2 = 0$ which tells us that the ideal line of all planes parallel to $z = 0$ is the only real part of μ (with multiplicity 16). The second factor corresponds to a pair of complex conjugate ideal lines with multiplicity 2.
The x-axis of the underlying Cartesian coordinate frame is a four-fold line on the surface and together with the cardiod ζ and the six-fold isotropic pair of lines through the origin of the underlying coordinate frame it completes the surface's intersection with $z = 0$. A rational parametrization can be achieved by substituting

$$\mathrm{c}_u = \frac{1-U^2}{1+U^2}, \quad \mathrm{s}_u = \frac{2U}{1+U^2} \tag{52}$$

and, surprisingly, with

$$\mathrm{S}_v = V, \quad \mathrm{C}_v = \sqrt{1+V^2} \tag{53}$$

since the hyperbolic functions showing up in the coordinate functions can be exressed in $\sinh v$ exclusively. Thus, the cardioidal minimal surface admits a rational Bézier representation of bi-degree $(8,4)$.
The adjoint surface looks like a *compressed helicoid*, see Fig. 12. Note that this surface cannot be a ruled surface, because the transcendental helicoid is the only ruled minimal surface. It is of algebraic degree 38. The intersection with the ideal plane is the cycle $z^{32}(x^2+y^2)^3$. The surface carries the two eight-fold straight lines $x = z = 0$ and $x = y = 0$.

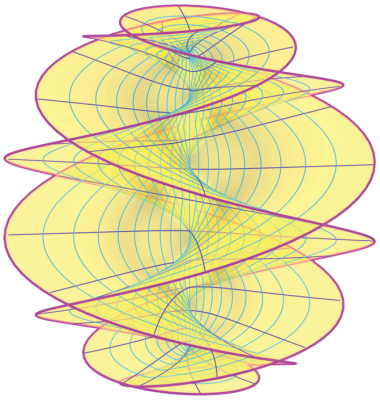


Figure 12: *A* compressed helicoid *as the adjoint to the cardiodal minimal surface.*

7.2 Steiner's hypocycloid

Steiner's hypocycloid appears in geometry in many ways. However, it is also a cycloidal curve and we can obtain it by choosing $R = 3$ and $r = -1$ in (45). The corresponding cycloidal minimal surface (46) turns out to be of algebraic degree 28 and of class 16. From the construction it is clear, that the horizontal cross-section with the plane $z = 0$ consists of the three-cusped hypocycloid. Moreover, the lines of symmetry $y = 0$ and $3x^2 = y^2$ (all three with multiplicity four) are part of the cross-section. Sine $y = z = 0$ annihilates the equation of this minimal surface, the x-axis of the underlying coordinate frame is entirely contained in this minimal surface.

The intersection of the *hypocycloidal minimal surface* with the ideal plane is given by the equation $z^{16}(x^2+y^2)^6 = 0$. Thus, the ideal line of all planes parallel to $z = 0$ is a 16-fold line on this surface. As is the case with any algebraic minimal surface, the ideal curve degenerates completely and splits into a finite number of lines.

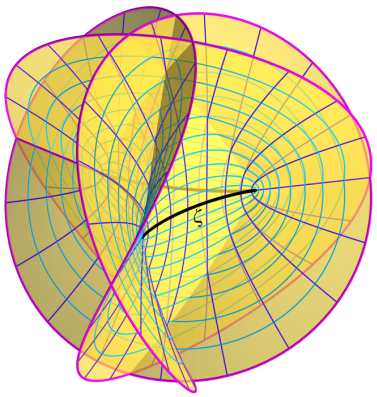


Figure 13: *The minimal surface on a geodesic hypocycloid ζ with three cusps.*

A rational Bézier representation of bi-degree (8,4) can be found by substituting (52) and (53). Figure 13 shows a part of the surface with a geodesic hypocycloid.

7.3 A geodesic nephroid

A final low degree example shall be discussed: We choose $R = 2$ and $r = 1$. This results in a minimal surface with a geodesic nephroid. The surface is of algebraic degree 24 and of class 72. The intersection with the ideal plane is the 18-fold ideal line of all planes parallel to $z = 0$ together with a three-fold pair of complex conjugate lines. Figure 14 shows the minimal surface with a geodesic nephroid.

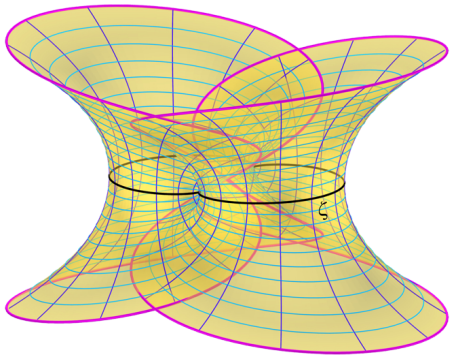


Figure 14: *The rational minimal surface with a geodesic nephroid* ζ.

The *nephroidal minimal surface* admits a rational Bézier representation of bi-degree (6,6) since we have to substitute

$$C_v = \frac{1+V^2}{1-V^2}, \quad S_v = \frac{2V}{1-V^2}.$$

The z- and the y-axis are contained in the surface.

8 Final remarks

The curves of constant slope mentioned in Thm. 20 can also be used as spine curves of scrolls on which minimal surfaces can be errected. Unfortunately, the minimal surface that touch the quadrics of revolution along curves of constant slope are, in general, not algebraic. With Thm. 2 the following theorem is a natural consequence:

Theorem 21. *The minimal surfaces that touch the vertical cylinders (generators parallel to the lead) along the curves of constant slope on quadrics of revolution are algebraic if the curves of constant slope are algebraic too.*

Note that the curves of constant slope on quadrics of revolution are algebraic if they are closed. Thus, the minimal surfaces mentioned in Thm. 21 are algebraic if the spine curves of the scrolls are closed curves of constant slope. Since the normals of all minimal surfaces described in Thm. 18 stay horizontal while the surfaces traverse the associate family, and furthermore, since the vertical cylinders' (horizontal) normals are always orthogonal to the tangents of the curves of constant slope, we can state

Theorem 22. *The algebraic minimal surfaces that touch the vertical cylinders along the curves of constant slopes on quadrics of revolution are precisely the algebraic minimal surfaces mentioned in Thm. 20.*

The algebraic degrees are growing rapidly and there will hardly be some low degree examples among the minimal surfaces described in Thm. 21.

References

[1] G. ANDREWS: *The primitive double minimal surface of the seventh class and its conjugate.* (1923) Michigan Historical Reprint Series, Michigan, 2005.

[2] W. BLASCHKE: *Vorlesungen über Differential Geometrie und geometrische Grundlagen von Einsteins Relativitätstheorie.* Springer, Berlin, 1921.

[3] W. BLASCHKE: *Bemerkungen über allgemeine Schraublinien.* Mh. Math. Phys. **19** (1908), 188–204.

[4] E. BOUR: *Théorie de la déformation des surfaces.* J. École Imp. Polytech. **22** (1862), cahier 39, 1–148.

[5] G. DARBOUX: *Lessons sur la theorie generale des surfaces et le applications geometrique du calcul infinitesimal.* I^re & II^me partie, Gautier Villars, Paris, 1914.

[6] A. DEMOULIN: *Sur les surfaces minima applicables sur des surfaces de révolution ou sur des surfaces spirales.* Darboux Bull. **2**/21 (1897), 244–252.

[7] A. ENNEPER: *Über Loxodromen.* Ber. Sächs. Ges. Wiss. Leipzig, 1902.

[8] A. ENNEPER: *Analytisch geometrische Untersuchungen.* Gött. Nachr. 1868, 258-277, 421-443.

[9] R.T. FAROUKI: *Pythagorean-Hodograph Curves: Algebra and Geometry Inseparable.* Springer, Berlin-Heidelberg, 2008.

[10] A.S. GALE: *On the rank, order and class of algebraic minimum curves.* Yale University, Yale, 1902.

[11] R. GLASER: *Ueber die Minimalflächen.* Inaug. Dissertation, Tübingen, 1891.

[12] C.F. GEISER: *Zur Erzeugung von Minimalflächen durch Scharen von Kurven vorgegebener Art.* Sb. Königl. Preuss. Akad. Wiss. Berlin, phys.-math. Classe, 1904, 677–686.

[13] C.F. GEISER: *Notiz über die algebraischen Minimalflächen.* Math. Ann. **3** (1871), 530–534.

[14] E. GOURSAT: Étude des surfaces qui admettent tous les plans de symétrie dun polyèdre régulier. Ann. de. Sci. Éc. Norm. **3**/IV (1887), 159–200.

[15] E. GOURSAT: Étude des surfaces qui admettent tous les plans de symétrie dun polyèdre régulier. Ann. de. Sci. Éc. Norm. **3**/IV (1887), 241–312.

[16] E. GOURSAT: Étude des surfaces qui admettent tous les plans de symétrie dun polyèdre régulier. Ann. de. Sci. Éc. Norm. **3**/IV (1887), 316–340.

[17] E. GOURSAT: Sur la théorie des surfaces minima. C.R. **CV** (1888), 743–746.

[18] J. HAAG: *Note sur les surfaces minima applicables sur une surface de révolution.* Darb. Bull. **30**/2 (1906), 75–94.

[19] J. HAAG: *Note sur les surfaces (B) algébraiques.* Darb. Bull. **30**/2 (1906), 293–296.

[20] J. HADAMARD: *Sur certaines surfaces minima.* Darb. Bull. **26**/2 (1902), 357-360.

[21] L. HENNEBERG: *Determination of the lowest genus of the algebraic minimum surfaces.* Bri. Ann. **IX**/2 (1878), 54-57.

[22] L. HENNEBERG: Über diejenige Minimalfläche, welche die Neilsche Parabel zur ebenen geodätischen Linie hat. Wolf Z. **XXI** (1876), 17–21.

[23] L. HENNEBERG: *Über solche Minimalflächen, welche eine vorgeschriebene ebene Curve zur geodätischen Linie haben.* Dissertation, Zürich, 1876.

[24] L. HENNEBERG: *Über die Evoluten der ebenen algebraischen Curven.* Wolf Z. **XXI** (1876), 22–23.

[25] J. HOSCHEK: *Liniengeometrie.* Bibliographisches Institut, Zürich, 1971.

[26] M. HUTH: *Kurven konstanter Steigung auf Flächen.* Jb. Realschule zu Stollberg, 1983.

[27] H. KARCHER: *Construction of minimal surfaces.* In: *Surves in Geometry*, Univ. of Tokyo, 1989, Lecture Notes No. 12, SFB 256, Bonn, 1989.

[28] R. KOCH: *Über die Mittelfläche einer isotropen Geradenkongruenz.* J. Geom. **23** (1984), 152–169.

[29] J.D. LAWRENCE: *A Catalog of Special Plane Curves.* Dover Publications, New York, 1972.

[30] S. LIE: *Gesammelte Abhandlungen.* Friedrich Engel, Poul Heegaard, eds., B.G. Teubner, Leipzig, 1922–1960.

[31] R. v. LILIENTHAL: *Über Minimaldoppelflächen.* Dt. Math.-Ver. **17** (1908), 277–291.

[32] G. LORIA: *Spezielle algebraische und transcendente ebene Kurven.* B.G. Teubner, Leipzig, 1911.

[33] J.C.C. NITSCHE: *Vorlesungen über Minimalflächen.* Springer-Verlag, 1975.

[34] B. ODEHNAL: *On rational isotropic congruences of lines.* J. Geom. **81** (2005), 126–138.

[35] B. ODEHNAL: *A rational minimal Möbius strip.* Proc. 17th Int. Conf. on Geometry and Graphics, August 4–8, 2016, Beijing, China, article no. 070.

[36] R. OSSERMAN: *A Survey of Minimal Surfaces.* Dover Publications, New York, 2nd edition, 1986.

[37] H. POTTMANN: *Zur Geometrie höherer Planetenumschwungbewegungen.* Mh. Math. **97** (1984), 141–156.

[38] A. RIBAUCOUR: *Étude des elassoïdes ou surfaces à courbure moyenne nulle.* Brux. Mém. cour. in **XLV**/4, 1882.

[39] H.W. RICHMOND: *On minimal surfaces.* Cambr. Trans. **18** (1900), 324–332.

[40] H.W. RICHMOND: *Ueber Minimalflächen.* Math. Ann. **54** (1901), 323–324.

[41] H.W. RICHMOND: *On the simplest algebraic minimal curves, and the derived real minimal surfaces.* Transact. Cambridge Phil. Soc. **19** (1904), 69–82.

[42] C. SCHILLING: *Die Minimalflächen 5. Klasse mit dem Stereoskopbild eines Modells derselben.* Inaug. Dissertation, Göttingen, 1880.

[43] H.A. SCHWARZ: *Miscellen aus dem Gebiete der Minimalflächen.* Wolf Z. **XIX** (1874), 243–271.

[44] H.A. SCHWARZ: *Gesammelte mathematische Abhandlungen.* Springer, Berlin, 1890.

[45] N.K. STEPHANIDIS: *Minimalfllächen und Strahlensysteme.* Arch. Math. **41** (1983), 544–554.

[46] K. STRUBECKER: *Differentialgeometrie.* Sammlung Göschen, Band 1180/1180a, de Gruyter, Berlin, 1969.

[47] E. STÜBLER: *Untersuchungen über spezielle Minimalflächen.* Crelle J. **80** (1875).

[48] E. STUDY: *Über einige imaginäre Minimalflächen.* Ber. königl. sächs. Ges. Wiss., math.-phys. Kl. **LXIII** (1911), 14–26.

[49] R. STURM: *Rein geometrische Untersuchungen über algebraische Minimalflächen.* J. für Math. **CV** (1889), 101–126.

[50] K. WEIERSTRASS: *Gesammelte Werke.* 7 Bände, Mayer und Müller, Berlin, 1894–1927 (Olms, Hildesheim, 1967).

[51] H. WIELEITNER: *Spezielle Ebene Kurven.* G.J. Göschen'sche Verlagshandlung, Leipzig, 1908.

[52] J.K. WHITTEMORE: *Minimal surfaces applicable to surfaces of revolution.* Ann. of Math. (2) **19**/1 (1917), 1–20.

[53] J.K. WHITTEMORE: *Minimal surfaces containing straigth lines.* Ann. of Math. (2) **22** (1921), 217–225.

[54] E. WÖLFFING: *Über Pseudotrochoiden.* Zeitschr. f. Math. u. Phys. **44** (1899), 139–166.

[55] W. WUNDERLICH: *Ebene Kinematik.* Bibliograph. Inst. Mannheim, 1970.

[56] W. WUNDERLICH: *Höhere Radlinien.* Österr. Ingen. Archiv **1** (1947), 277–296.

Boris Odehnal
e-mail: boris.odehnal@uni-ak.ac.at
University of Applied Arts Vienna
Oskar-Kokoschka-Platz 2, A-1100 Vienna, Austria

Original scientific paper
Accepted 14. 12. 2016.

SZILVIA B.-S. BÉLA
MÁRTA SZILVÁSI-NAGY

Adjusting Curvatures of B-spline Surfaces by Operations on Knot Vectors

Adjusting Curvatures of B-spline Surfaces by Operations on Knot Vectors

ABSTRACT

The knot vectors of a B-spline surface determine the basis functions hereby, together with the control points, the shape of the surface. Knot manipulations and their influence on the shape of curves have been investigated in several papers (see e.g. [4] and [5]). The computations can be made very efficiently, if the basis functions and the vector function of the B-spline surface are represented in matrix form (see [1] and [6]). In our latest work [2] we summarized the knot manipulation techniques and the corresponding computations in matrix form. We also developed an algorithm for a direct knot sliding, how a knot can be repositioned in one step instead of inserting a new knot value, then removing an old one from the knot vector.

In this paper we analyse the effect of varying knot intervals on the Gaussian curvature of a B-spline surface at a given point. We present an algorithm for the deformation of a B-spline surface, so that it should go through a given point with a given Gaussian curvature. The result of this deformation is, that a sphere with a given radius will fit tangential the reshaped surface at the given point with equal Gaussian curvatures. In applications the same situation arises, when a ball-end tool is pushed into a surface during processing.

In our algorithm we use only linear interpolation equations besides the repositioning of knot values, in order to get numerically stable and effective solutions.

Key words: surface representations, geometric algorithms

MSC2010: 65D18, 68D05

Podešavanje zakrivljenosti B-splajn ploha operacijama na čvor vektorima

SAŽETAK

Na ovaj način čvor vektori B-splajn ploha određuju temeljne funkcije zajedno s kontrolnim točkama te oblik plohe. U nekoliko članaka (vidi na primjer [4] i [5]) proučavale su se operacije na čvorovima i njihov utjecaj na obilk krivulja. Izračuni mogu biti izvedeni vrlo efikasno ako su temeljne funkcije i vektor funkcije B-splajn plohe prikazane u matričnom obliku (vidi [1] i [6]). U našem posljednjem radu [2] saželi smo metode operacija na čvorovima i odgovarajućih izračuna u matričnom obliku. Također, razvili smo algoritam za izravno klizanje čvorova, tj. pokazali smo kako čvor može biti premješten u jednom koraku umjesto da uvodimo novu vrijednost čvora, a zatim uklanjanjem starog iz čvor vektora.

U ovom članku analiziramo utjecaj mijenjanja intervala čvorova na Gaussovu zakrivljenost u zadanoj točki B-splajn plohe. Prikazujemo algoritam za deformaciju B-splajn plohe tako da prolazi kroz zadanu točku sa zadanom Gaussovom zakrivljenosšću. Rezultat ove deformacije kaže da kugla zadanog radijusa dira preoblikovanu plohu u zadanoj točki s jednakim Gaussovim zakrivljenostima. Ista situacija događa se u primjenama kad je alat kuglama uguran u plohu tijekom procesa.

S ciljem da postignemo numerički stabilna i efikasna rješenja, osim premještanja vrijednosti čvora, u našem algoritmu koristimo samo jednadžbe linearne interpolacije.

Ključne riječi: prikazi plohe, geometrijski algoritmi

1 Problem solution in the symmetric case

The mathematical formulation of the problem

Our task is to push a given sphere into a B-spline surface by reshaping a part of it around a given common interpolation point such that, the surface and the sphere are in a tangential position and they have equal Gaussian curvatures at this point (Fig 1, Fig 2). We only address the geometrical side of the problem, but not the mechanical aspects. Furthermore, we do not set conditions on area or volume preserving. We just focus on this geometrical design problem.

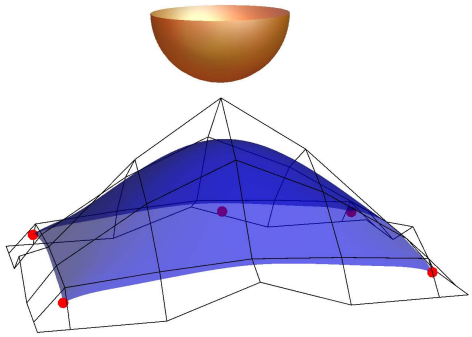


Figure 1: *The sphere will be pushed into the interpolation point given in the middle.*

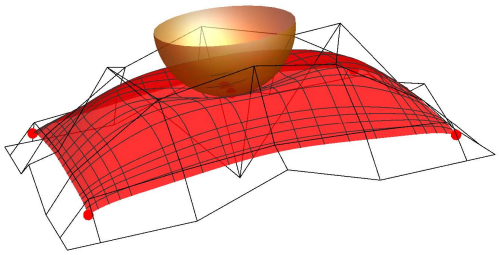


Figure 2: *The deformed surface and the sphere have equal Gaussian curvature at the common point.*

The shape of the surface can be controlled by its control points and by the parametrization of the basis functions, that means, by the knot vectors. The interpolation problem with a prescribed Gaussian curvature leads to quadratic rational expressions of the surface data, but our algorithm avoids nonlinear numerical methods by choosing appropriate variables, interpolation conditions and by applying a simple iteration method.

Let the B-spline surface of degree 3×3 be given with non-uniform periodic knot vectors $(u_1 < u_2 < \cdots < u_{n+4})$ and $(v_1 < v_2 < \cdots < v_{m+4})$, $n,m \geq 4$. The vector function representing the B-spline surface is

$$\mathbf{b}(u,v) = (u^3,u^2,u,1)\cdot\left(\mathbf{N}_\mathbf{u}^4\right)^T\cdot\mathbf{Q}\cdot\mathbf{M}_\mathbf{v}^4\cdot(v^3,v^2,v,1)^T,$$

where $\mathbf{N}_\mathbf{u}^4$ and $\mathbf{M}_\mathbf{v}^4$ are the corresponding coefficient matrices of the basis functions and $\mathbf{Q}$ is the matrix of the control points $\mathbf{q}_{i,j}$, $(i = 1,2,\ldots,n,\ j = 1,2,\ldots,m)$.

We will restrict the computation to a region of 4×4 patches of the B-spline surface because this part is influenced by the second order curvature condition prescribed in the middle of this region. In this case $n,m \geq 7$. The other parts of the surface outside of this region remain unchanged.

The input data of the B-spline surface are the knot vectors (i.e. the parameter values) and the control points. The parameter grid with the actual region is shown in Fig 3, the control net with the generated surface is shown in Fig 4. In these examples the control net and the knot vectors are symmetrical about the midpoint of the actual region, therefore, the generated surface is also symmetrical about this point.

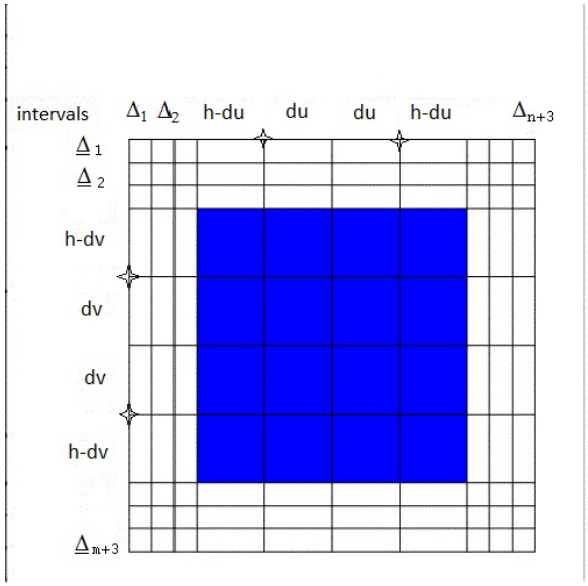


Figure 3: *Parameter domain of the surface:* 4×4 *surface patches are determined by the net of* 11×11 *knot values, (h is constant).*

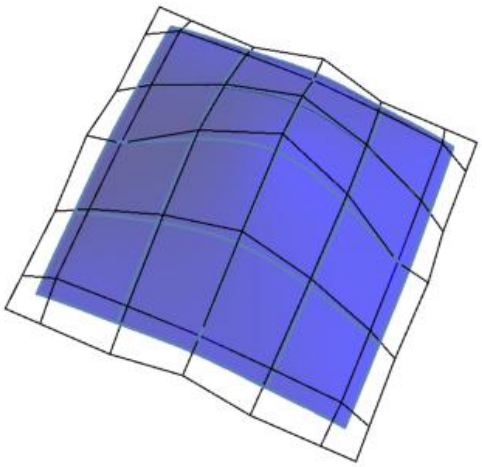


Figure 4: 7×7 *control points and the generated surface.*

First phase of the deformation: interpolation

For the deformation of the surface we prescribe 9 interpolation points and the Gaussian curvature at the interpolation point P in the middle, where the sphere with the corresponding radius will touch the reshaped surface. Four

interpolation points are given in the corners of the B-spline surface. Five interpolation points, namely P and the four corner points are shown in Fig 1. The remaining four interpolation points are determined around P according to the radius of the given sphere symmetrically with respect to P (Fig 5). They are computed on the surface of the sphere. The corresponding parameter values of the required surface are estimated by the relative measurements of the surface patches and the sphere. To the nine interpolation conditions we choose nine variables, which are nine inner control points of the 7×7 control net (Fig 6).

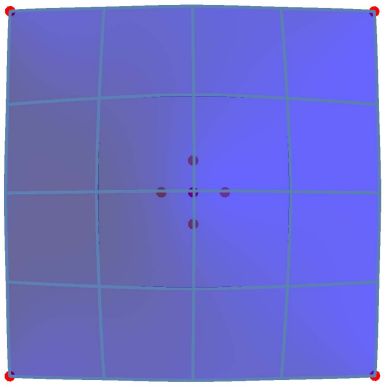


Figure 5: *The nine interpolation points*

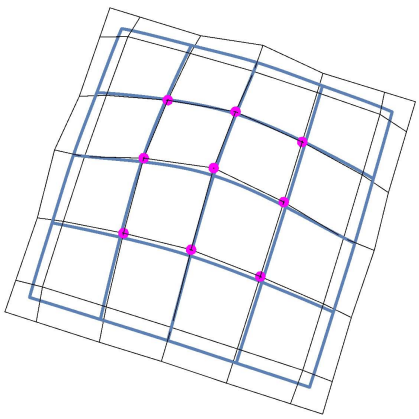


Figure 6: *Nine variable control points in the* 7×7 *control net*

Each of the nine interpolation conditions is a linear vector equation in the nine control points, which are the unknowns of a system of linear equations.

$$\mathbf{p}_i = \mathbf{b}(u_i, v_i),\ i = 1, \ldots, 9$$

The knot vectors are now fixed, and the coefficient matrices are computed accordingly. On the right hand side the vector function $\mathbf{b}(u,v)$ is depending on the nine unknown control points and it is evaluated at the parameter values (u_i, v_i). The unknown nine control points are included in the matrix $\mathbf{Q}$. The pairs of the parameter values (u_i, v_i) belong to the interpolation points, the position vectors of which are denoted by $\mathbf{p}_i$. The solution results in a control net of the B-spline surface interpolating the nine given points. Its Gaussian curvature at the midpoint P is now determined. How can it be equal to the given value?

Second phase of the deformation: adjusting the Gaussian curvature

Now we modify the knot vectors in order to deform the shape of the surface around the interpolation point P. In the knot vectors four knot intervals in the middle of the knot vectors will be changed by repositioning (sliding) the knot values $u_5 \in (u_4, u_6)$, $u_7 \in (u_6, u_8)$ and symmetrically $v_5 \in (v_4, v_6)$, $v_7 \in (v_6, v_8)$, respectively. These knot values are marked in Fig 3. The variables in the knot vectors are du and dv. For smaller du, dv the generated B-spline surface gets nearer to the control net, for larger du, dv it is more flat, and lies farther from the control net.

If we change the knot intervals $du = dv$ and analyse, how the Gaussian curvature of the surface (denoted by κ_G) at the point P changes, we get a monotone scalar function $\kappa_G(du)$. In Fig 7 the corresponding *radius* $= (\sqrt{\kappa_G})^{-1}$ of the sphere is shown depending on $du = dv$.

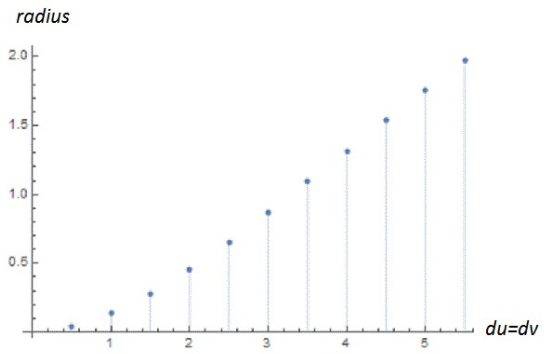


Figure 7: *The radius of the sphere as function of the length of the knot intervals* $du = dv$

The Gaussian curvature is varying in a limited interval while the chosen parameter values are repositioned in the intervals (u_4, u_6), (u_6, u_8) and (v_4, v_6), (v_6, v_8) of a fixed length h, respectively. To a given Gaussian curvature between these limits the corresponding knot intervals $du = dv$ are determined by a simple iteration from this scalar function.

Then the required surface is generated with the new knot vectors. The result of this computation is shown in Fig 2. The point P is an umbilical (a special elliptical) point of the new surface due to the symmetrical data and symmetrical change of the knot vectors. After this deformation the boundary curves of the surface consisting of 4×4 patches do not change, as it is shown in Fig 8.

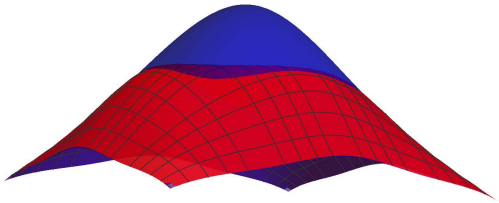


Figure 8: *The original and the deformed surfaces have the same boundary curves.*

2 The asymmetric case

In the non-symmetrical case the deformation presented above leads to an elliptic surface point P, where the main curvature values are different, though the Gaussian curvature is equal to that of the given sphere. This situation is shown in Fig 9 by pushing the sphere a bit into the deformed surface, where the intersection has an elliptical form.

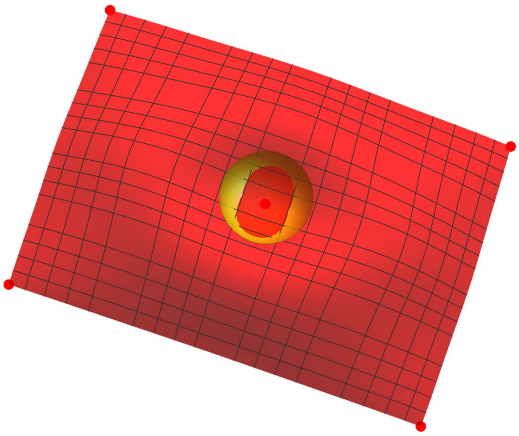


Figure 9: *Elliptical surface point at P with different main curvatures*

Now we carry out a further deformation in order to get a special elliptic, i.e. an umbilical point at P. We repeat the second phase of the deformation by changing the knot intervals, now only in one parameter direction, let us say, on the v-knot vector. For the chosen values of dv within the intervals (v_4, v_6) and (v_6, v_8) we compute the curvatures of the parameter curves at the surface point P. Assuming that the parameter net is orthogonal, the Gaussian curvature is the product of the curvatures of the u- and v-parameter curves. In Fig 10 the dependence of the curvature of the v-parameter curve, denoted by $v-curvature$, on the knot interval dv is shown. Meanwhile the values of the $u-curvatures$ are practically constant (they are not shown). We determine the knot interval dv to the $v-curvature$ which is equal to the $u-curvature$ from this monotone discrete function by simple iteration. The computed surface with this knot vector has an umbilical point at P. That is visualized by the sphere pushed slightly into the surface at the touching point (Fig 11).

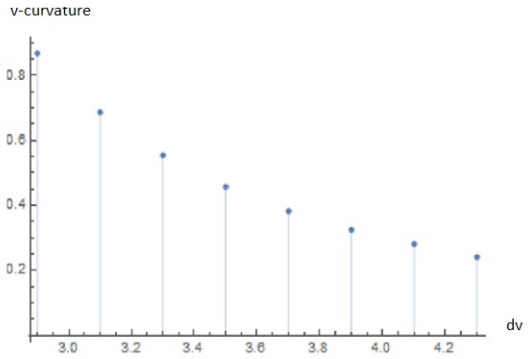


Figure 10: *The values of the $v-curvatures$ at the values of the knot interval dv*

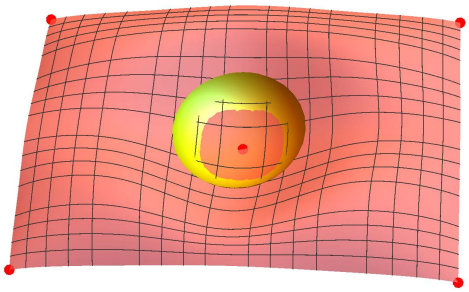


Figure 11: *Umbilical surface point at P with equal main curvatures*

We remark that by this second deformation the Gaussian curvature has been changed slightly. Analysing this variation, an appropriate correction could be carried out in a similar way, repeating the computation for modified values of du and dv, if the difference in the radii of the touching spheres computed from the resulting Gaussian curvatures is not acceptable. We have omitted this step.

We remark also that there are no equal values of u- and v-curvatures on the considered interval, if the surface is strongly asymmetric around the point P. In this case the parameter values of the four inner interpolation points have to be corrected accordingly.

3 The mathematical tools

Matrix form of the basis functions
The B-spline surface in our algorithm is presented in matrix form. This form allows to write the interpolation conditions in the form of explicit vector equations. In [1] we have given a short overview of the published papers about the matrix form of B-spline functions, and we also have given a method for generating the entries of the coefficient matrices of any degree over periodic knot vectors.
In our algorithm we have changed the position of a knot value within a given knot interval in order to analyse the change of the Gaussian curvature at a given surface point. In each step the corresponding coefficient matrices have been generated for the vector function representing the B-spline surface.
In order to represent the basis functions of non-uniform B-splines in matrix form first we describe a reformulation technique of the de Boor-Cox recursion. From this recursion formula we can generate the representation matrix of the basis functions in the not normalized Bernstein basis. Then we can apply a simple transformation from the not normalized Bernstein basis to the polynomial space spanned by the power basis. Thus with the algebraic reformulation of the B-spline recursion we gain the conversion matrices of the B-spline functions to the power basis.
First we present here a simple reformulation of the de Boor-Cox recursion. The basis functions of order k over the knot vector $\{t_1,\dots t_n\}$ are defined by the de Boor-Cox formula as:

$$\begin{aligned} N_i^1(t) &= \begin{cases} 1, & t\in[t_i,t_{i+1}) \\ 0, & \text{otherwise,} \end{cases} \\ N_i^k(t) &= \alpha(t_i,t_{i+k-1};t)\,N_i^{k-1}(t)+\alpha(t_{i+k},t_{i+1};t)\,N_{i+1}^{k-1}(t), \end{aligned}$$

where the function α is defined as

$$\alpha(A,B;t)=\frac{t-A}{B-A} \tag{1}$$

for arbitrary parameters A,B, where $A\neq B$, and for all $t\in[A,B]$.
We can generate the pieces of the basis functions restricted to one knot interval $[t_j,t_{j+1})$ by rewriting the recursion as

$$\begin{aligned} N_i^1(t) &= \begin{cases} 1, & t\in[t_i,t_{i+1}) \\ 0, & \text{otherwise,} \end{cases} \\ N_i^k(t) &= \alpha(t_j,t_{j+1};t)\Big[\alpha(t_i,t_{i+k-1};t_{j+1})\,N_i^{k-1}(t) \\ &+ \alpha(t_{i+k},t_{i+1};t_{j+1})\,N_{i+1}^{k-1}(t)\Big] \\ &+ \alpha(t_{j+1},t_j;t)\Big[\alpha(t_i,t_{i+k-1};t_j)\,N_i^{k-1}(t) \\ &+ \alpha(t_{i+k},t_{i+1};t_j)\,N_{i+1}^{k-1}(t)\Big], \text{ where } t\in[t_j,t_{j+1}). \end{aligned} \tag{2}$$

According to this form we transform all segments of the basis functions from the knot span $[t_j,t_{j+1})$ and represent them over the unit interval as follows:

$$\begin{aligned} N_i^1(t(u)) &= \begin{cases} 1, & i=j \\ 0, & \text{otherwise,} \end{cases} \\ N_i^k(t(u)) &= u\Big[\alpha(t_i,t_{i+k-1};t_{j+1})\,N_i^{k-1}(t(u)) \\ &+ \alpha(t_{i+k},t_{i+1};t_{j+1})\,N_{i+1}^{k-1}(t(u))\Big] \\ &+ (1-u)\Big[\alpha(t_i,t_{i+k-1};t_j)\,N_i^{k-1}(t(u)) \\ &+ \alpha(t_{i+k},t_{i+1};t_j)\,N_{i+1}^{k-1}(t(u))\Big], \end{aligned} \tag{3}$$

where $u\in[0,1), t\in[t_j,t_{j+1})$ and $u(t)=\alpha(t_j,t_{j+1},t)$. Over the knot spans, where $j=k,k+1,\dots,n-k$ the basis functions have k different, non-zero polynomial segments. These segments can be represented by a matrix equation in the not normalized Bernstein basis $\{u^{k-1},u^{k-2}(1-u),\dots,(1-u)^{k-1}\}$ over the unit interval:

$$\begin{pmatrix} N_1^k(t(u)) \\ N_2^k(t(u)) \\ \vdots \\ N_{n-k}^k(t(u)) \end{pmatrix} = \mathbf{C}^k\cdot\begin{pmatrix} u^{k-1} \\ u^{k-2}(1-u) \\ \vdots \\ (1-u)^{k-1} \end{pmatrix}, \quad \begin{matrix} t\in[t_j,t_{j+1}), \\ u\in[0,1), \end{matrix} \tag{4}$$

where $\mathbf{C}^k\in\mathbb{R}^{n-k\times k}$, and it contains the coefficients of $u^m(1-u)^{k-1-m}$ computed recursively from (4). For each $k=2,3,\dots$ this matrix contains several rows, where all elements are zeros. These rows contain the coefficients of the basis functions that are zero over the knot span $[t_j,t_{j+1})$. The non-zero rows for each j (from row $j+1-k$ to row j), where $j\geq k$, give the coefficients of the basis functions with the support containing the interval $[t_j,t_{j+1})$.

If we represent the segments of the basis functions from a given knot span $[t_j,t_{j+1})$ in the matrix equation form (4), then it is easy to transform this representation to a matrix representation in the power basis: $\{u^{k-1},u^{k-2},\dots,u,1\}$. In order to find the transformation matrix of the basis functions to the power basis, it is sufficient to find the transformation matrix $\mathbf{P}^k$ from the not normalized Bernstein basis to the power basis for polynomials of degree $k-1$:

$$\begin{pmatrix} N_1^k(t(u)) \\ N_2^k(t(u)) \\ \vdots \\ N_{n-k}^k(t(u)) \end{pmatrix} = \mathbf{C}^k\cdot\mathbf{P}^k\cdot\begin{pmatrix} u^{k-1} \\ u^{k-2} \\ \vdots \\ 1 \end{pmatrix}, \quad \begin{matrix} t\in[t_j,t_{j+1}), \\ u\in[0,1). \end{matrix} \tag{5}$$

The $\mathbf{P}^k$ matrix is a lower triangular matrix with the entries:

$$\mathbf{P}^k[i,l] = \begin{cases} (-1)^{i-l+1} \cdot \binom{i-1}{l-1}, & l \le i, \\ 0, & \text{otherwise}, \end{cases}$$

where l and $i = 1,\ldots,k$. This matrix can be easily derived according to the Binomial-theorem. A conversion matrix from the not normalized Bernstein basis to the power basis can be also found in the literature [3].

Interpolation and iteration

These mathematical tools are well-known basic methods in solutions of various numerical problems. In our algorithm the linear interpolation problem formulated by a system of linear vector equations has been separated from the non-linear interpolation problem, where a scalar value (i.e. the Gaussian curvature of the surface) has to be interpolated with vector variables. We have solved this problem by simple iteration on a monotone, scalar function by computing discrete values of this function in an appropriate interval. This method is fast and numerical stable. In this way we have avoided non-linear numerical problems.

The computations and the figures have been made by the symbolical algebraic program package Wolfram Mathematica.

4 Remarks

In our algorithm there are some basic assumptions. The first one is that the local deformation of a B-spline surface of degree 3×3 has been computed on a region of 4×4 patches, where the fixed point lies in the middle of this region. Though each control point and each basis function has an effect on four knot intervals, the boundary curves of this part of the surface have not changed under the used interpolation conditions.

The next assumption is that the prescription of the four inner interpolation points are computed from the data of the given sphere. The setting of the parameter values for which the surface interpolates these points has been made on the base of experimental results, as usually in the solutions of many practical problems.

A further simplification in the solution is that we have computed with surfaces on orthogonal parameter grids. In this way we have computed the Gaussian curvature as the product of the curvatures of the u- and v-parameter curves. In this way we have avoided the numerical computation of the main curvatures in each step, because this computation is not essential in our algorithm.

5 Conclusions

We have given an algorithm for the solution of a practical problem: how to press a given sphere into a B-spline surface at a prescribed position. We have shown a solution, when the surface is symmetric around the given interpolation point. Then in the non-symmetric case we have shown a further deformation of the surface in order to transform the elliptical surface point into an umbilical one. In this case the given sphere osculates the deformed surface with equal Gaussian curvatures at the given tangential point.

Acknowledgement

The research of the second author was supported in a cooperation with the Technical University Berlin.

References

[1] SZ. BÉLA, M. SZILVÁSI-NAGY, General matrix representation of B-splines and approximation of B-spline curves and surfaces with third order continuity, *Seventh Hungarian Conference on Computer Graphics and Geometry*, Budapest, 2014, 1–6.

[2] SZ. BÉLA AND M. SZILVÁSI-NAGY, Manipulating end conditions of B-spline curves using knot sliding, *Eighth Hungarian Conference on Computer Graphics and Geometry*, Budapest, 2016, 83–93.

[3] G. FARIN, *Curves and Surfaces for Computer Aided Geometric Design, A Practical Guide*, Academic Press, Second Edition, 1990.

[4] I. JUHÁSZ, M. HOFFMANN, The effect of knot modications on the shape of B-spline curves, *Journal for Geometry and Grapics* **5** (2001), 111–119.

[5] I. JUHÁSZ, M. HOFFMANN, Constrained shape modification on cubic B-spline curves means of knots, *Computer-Aided Design* **36** (2004), 437–445.

[6] M. SZILVÁSI-NAGY, SZ. BÉLA, Stitching B-spline curves symbolically, *KoG* **17** (2013), 3–8.

Szilvia B.-S. Béla
e-mail: belus@math.bme.hu
Márta Szilvási-Nagy
e-mail: szilvasi@math.bme.hu
Department of Geometry, Budapest University of Technology and Economics
Budafoki út 8., H-1521 Budapest, Hungary

Original scientific paper
Accepted 17. 12. 2016.

MIRELA KATIĆ ŽLEPALO

Curves of Foci of Conic Pencils in pseudo-Euclidean Plane

Curves of Foci of Conic Pencils in pseudo-Euclidean Plane

ABSTRACT

In this article, it will be shown that the curve of foci of an order conic pencil in the pseudo-Euclidean plane is generally a bicircular curve of 6th order. In some cases, depending on a position of four base points of the pencil, this curve is of 5th, 4th or 3rd order and in some cases it is even a conic or only a line.

Key words: pseudo-Euclidean plane, conic sections, foci, conic pencil

MSC2010: 51A05; 51M15

Krivulje žarišta u pramenovima konika u pseudo-euklidskoj ravnini

SAŽETAK

U ovom članku pokazat će se da je krivulja žarišta pramena konika u pseudo-euklidskoj ravnini općenito bicirkularna krivulja šestog reda. U nekim slučajevima, u ovisnosti o položaju četiriju temeljnih točaka pramena, krivulja žarišta može biti petog, četvrtog ili trećeg reda, a može biti i konika ili samo pravac.

Ključne riječi: pseudo-euklidska ravnina, konike, žarišta, pramenovi konika

1 Introduction

A *pseudo-Euclidean plane* (*PE-plane*) is a real projective plane where the metric is induced by a real line a and two real points A_1 and A_2 incident with it, see [10]. It is one of nine plane geometries and according to [8], it is parabolic-hyperbolic plane. The affine model of the pseudo-Euclidean plane will be used, where the absolute line a is determined by the equation $x_0 = 0$ and the absolute points A_1, A_2 by the coordinates $(0,1,\pm 1)$, like in [4], [5], [6], [7], [11].
Further on, some basic, well-known definitions are given ([1], [2], [3], [7]).

Definition 1 *Points incident with the absolute line a are called isotropic points.*

Definition 2 *Lines incident with one of the absolute points A_1 or A_2 are called isotropic lines.*

Definition 3 *Foci of a conic are intersection points of its isotropic tangent lines.*

Definition 4 *Circular curve in PE-plane is a curve incident with at least one of two absolute points.*

Definition 5 *A curve k is said to be of (r,t)- type of circularity if the absolute point A_1 is the intersection of the curve k with the absolute line a of multiplicity r, and the absolute point A_2 is the intersection of the curve k with the absolute line a of multiplicity t. The sum $r+t$ is called degree of circularity.*

Definition 6 *The curve of order n is said to be entirely circular if $n = r+t$, i.e. the order of the curve equals the degree of circularity.*

Conics in pseudo-Euclidean plane are divided into ([7], [9]):

- hyperbola intersecting the absolute line in two real and distinct points
- ellipse intersecting the absolute line in a pair of conjugate-imaginary points
- parabola touching the absolute line

- special hyperbola intersecting the absolute line in two real and distinct points out of which one is an absolute point
- special parabola touching the absolute line in an absolute point
- circle intersecting the absolute line in both absolute points.

A conic has four foci which can be real and distinct, conjugate-imaginary, double real or even quadruple real. An ellipse has four real foci. A hyperbola may have four real or four conjugate-imaginary foci. A special hyperbola may have two double real or two double imaginary foci. A parabola has one non-isotropic focus, another one in the isotropic touching point of the parabola and the absolute line, and one focus in each absolute point. A special parabola has double focus in each absolute point. All four foci of a circle are in the same point - quadruple focus (which is also the center of the circle).

2 Curves of foci

A conic is uniquely determined by five of its points. Four points, called *base points* determine infinitely many conics which are called *an order pencil of conics*.
In the affine model of PE-plane, the coordinates of the points are determined with

$$x=\frac{x_1}{x_0}, y=\frac{x_2}{x_0}, \tag{1}$$

the absolute line a with the equation $x_0 = 0$, and the absolute points A_1, A_2 with coordinates $(0,1,\pm 1)$.
A conic is given by equation in homogeneous coordinates

$$a_{00}x_0^2+a_{11}x_1^2+a_{22}x_2^2+2a_{01}x_0x_1+2a_{02}x_0x_2+2a_{12}x_1x_2=0, \tag{2}$$

and in the affine coordinates

$$a_{00}+a_{11}x^2+a_{22}y^2+2a_{01}x+2a_{02}y+2a_{12}xy=0. \tag{3}$$

Some short calculations lead to the conclusions:

- c is a hyperbola iff $a_{12}^2-a_{11}a_{22}>0$,
- c is a parabola iff $a_{12}^2-a_{11}a_{22}=0$,
- c is an ellipse iff $a_{12}^2-a_{11}a_{22}<0$,
- c is a special hyperbola iff $a_{11}+a_{22}+2a_{12}=0$ or $a_{11}+a_{22}-2a_{12}=0$,
- c is a special parabola iff $a_{11}=a_{22}=-a_{12}$ or $a_{11}=a_{22}=a_{12}$,
- c is a circle iff $a_{12}=0$ and $a_{11}=-a_{22}$,

as it is shown in [6].
Let the conic pencil be given with the points $A(0,0)$, $B(1,0)$, $C(0,2)$ and $D(t_1,t_2)$. To avoid special cases, the point D should not be incident with lines AB, BC or AC, i.e. $t_1\neq 0$, $t_2\neq 0$ and $t_2\neq -2t_1+2$.
Substituting the coordinates of points A, B, C and D into (3) yields

$$a_{00}=0,$$
$$a_{00}+2a_{01}+a_{11}=0,$$
$$a_{00}+4a_{02}+4a_{22}=0,$$
$$a_{00}+2a_{01}t_1+a_{11}t_1^2+2a_{02}t_2+2a_{12}t_1t_2+a_{22}t_2^2=0.$$

So, the equation of the conic pencil is

$$a_{11}x^2+\frac{(a_{11}t_1-a_{11}t_1^2+2a_{22}t_2-a_{22}t_2^2)xy}{t_1t_2} +a_{22}y^2-a_{11}x-2a_{22}y=0. \tag{4}$$

Different form of the equation (4) is

$$a_{11}(x^2+\frac{1-t_1}{t_2}xy-x)+a_{22}(y^2+\frac{2-t_2}{t_1}xy-2y)=0.$$

It is obvious that the conic pencil is a linear combination of two degenerate conics of the pencil - the first one consists of lines AC and BD, and the second one of lines AB and CD. Introducing $\lambda=\frac{a_{22}}{a_{11}}$ into equation (4), the following equation for the conic pencil with the base points A, B, C and D is obtained

$$x^2+\frac{2\lambda xy}{t_1}+\frac{xy}{t_2}-\frac{t_1xy}{t_2}-\frac{\lambda t_2xy}{t_1}+\lambda y^2-x-2\lambda y=0, \tag{5}$$

where each conic of the pencil is uniquely defined by parameter $\lambda\in \mathrm{R}\cup\infty$.
In line coordinates, the pencil (5) has the following equation

$$-\lambda^2u^2+\lambda uv-\frac{v^2}{4}+\lambda u-\frac{2\lambda^2u}{t_1}-\frac{\lambda u}{t_2}+\frac{\lambda t_1u}{t_2}+\frac{\lambda^2t_2u}{t_1} +2\lambda v-\frac{\lambda v}{t_1}-\frac{v}{2t_2}+\frac{t_1v}{2t_2}+\frac{\lambda t_2v}{2t_1}+\frac{\lambda}{2}-\frac{\lambda^2}{t_1^2}+\frac{\lambda}{2t_1}-\frac{1}{4t_2^2} +\frac{t_1}{2t_2^2}-\frac{t_1^2}{4t_2^2}+\frac{\lambda}{t_2}-\frac{\lambda}{t_1t_2}+\frac{\lambda^2t_2}{t_1^2}-\frac{\lambda^2t_2^2}{4t_1^2}=0. \tag{6}$$

For an isotropic line $u_0x_0+u_1x_1+u_2x_2=0$ passing through the absolute point $A_1(0,1,1)$ the following must be true

$$u_1 + u_2 = 0, \tag{7}$$

i.e. in affine coordinates

$$v = -u, \tag{8}$$

where $u = \frac{u_1}{u_0}$ and $v = \frac{u_2}{u_0}$. Thus, such line in affine coordinates has the equation $1 + ux - uy = 0$. From that equation follows:

$$u = \frac{1}{y-x}. \tag{9}$$

Analogously for an isotropic line passing through the absolute point $A_2(0,1,-1)$ the following is true

$$v = u \tag{10}$$

and

$$u = \frac{-1}{x+y}. \tag{11}$$

When v in (6) is replaced by using (8), and then u is replaced with (9), the following is obtained

$$\begin{aligned}
&-\frac{1}{(y-x)^2} - \frac{4\lambda}{(y-x)^2} - \frac{4\lambda^2}{(y-x)^2} - \frac{2\lambda}{y-x} + \frac{4\lambda}{t_1(y-x)} \\
&- \frac{8\lambda^2}{t_1(y-x)} + \frac{2}{t_2(y-x)} - \frac{4\lambda}{t_2(y-x)} - \frac{2t_1}{t_2(y-x)} \\
&+ \frac{4\lambda t_1}{t_2(y-x)} - \frac{2\lambda t_2}{t_1(y-x)} + \frac{4\lambda^2 t_2}{t_1(y-x)} - \lambda - \frac{4\lambda^2}{t_1^2} + \frac{2\lambda}{t_1} - \frac{1}{t_2^2} \\
&+ \frac{2t_1}{t_2^2} - \frac{t_1^2}{t_2^2} + \frac{4\lambda}{t_2} - \frac{4\lambda}{t_1 t_2} + \frac{4\lambda^2 t_2}{t_1^2} - \frac{\lambda^2 t_2^2}{t_1^2} = 0.
\end{aligned} \tag{12}$$

The intention is to isolate only those lines passing through the absolute point A_1 out of all lines from the pencil (6). Solving the equation (12) for λ, the following two solutions are obtained

$$\begin{aligned}
\lambda_{1,2} = &-\frac{1}{t_2^2(-2x+t_2x+2y-t_2y+2t_1)^2}[2t_1t_2x^2 - 2t_1^2t_2x^2 - t_1t_2^2x^2 \\
&- t_1^2t_2^2x^2 - 4t_1t_2xy + 4t_1^2t_2xy + 2t_1t_2^2xy + 2t_1^2t_2^2xy + 2t_1t_2y^2 \\
&- 2t_1^2t_2y^2 - t_1t_2^2y^2 - t_1^2t_2^2y^2 - 2t_1^2t_2x + 2t_1^3t_2x + 2t_1t_2^2x - 2t_1^2t_2^2x \\
&- t_1t_2^3x + 2t_1^2t_2y - 2t_1^3t_2y - 2t_1t_2^2y + 2t_1^2t_2^2y + t_1t_2^3y + 2t_1^2t_2^2 \\
&\pm 2\sqrt{-t_1^3t_2^3(-2+2t_1+t_2)(-1+x-y)(x-y)(2+x-y)(t_1-t_2-x+y)}].
\end{aligned} \tag{13}$$

This corresponds to the fact that each isotropic line passing through A_1 is tangent line of two conics of the pencil, and those lines are precisely two lines defined by parameters λ_1 and λ_2. Analogously, separating from the whole pencil (6) only isotropic lines passing through the absolute point A_2 (by substituting v in (6) using (10), and then substituting u using (11)) and solving that equation for λ, the following is obtained

$$\begin{aligned}
\lambda_{3,4} = &-\frac{1}{t_2^2(2t_1-2x+t_2x-2y+t_2y)^2}[2t_1t_2x^2 - 2t_1^2t_2x^2 - t_1t_2^2x^2 \\
&- t_1^2t_2^2x^2 + 4t_1t_2xy - 4t_1^2t_2xy - 2t_1t_2^2xy - 2t_1^2t_2^2xy + 2t_1t_2y^2 \\
&- 2t_1^2t_2y^2 - t_1t_2^2y^2 - t_1^2t_2^2y^2 - 2t_1^2t_2x + 2t_1^3t_2x - 2t_1t_2^2x + 6t_1^2t_2^2x \\
&+ t_1t_2^3x - 2t_1^2t_2y + 2t_1^3t_2y - 2t_1t_2^2y + 6t_1^2t_2^2y + t_1t_2^3y - 2t_1^2t_2^2 \\
&\pm 2\sqrt{-t_1^3t_2^3(-2+2t_1+t_2)(t_1+t_2-x-y)(-2+x+y)(-1+x+y)(x+y)}].
\end{aligned} \tag{14}$$

The solutions are the parameters λ_3 and λ_4 which uniquely determine two conics of the pencil.

Foci of a conic are four intersections of its two tangent lines from the absolute point A_1 with its two tangent lines from the absolute point A_2. Hence, the conics determined by parameters λ_1 and λ_2 must be equal to the conics determined by parameters λ_3 and λ_4, i.e. $\lambda_1 = \lambda_3$ or $\lambda_2 = \lambda_3$ or $\lambda_1 = \lambda_4$ or $\lambda_2 = \lambda_4$ are valid. So, the equation of the curve of foci is obtained by using

$$(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_4)(\lambda_2 - \lambda_4) = 0. \tag{15}$$

Introducing (13) and (14) into (15), the equation of the curve of foci $F(x,y) = 0$ is obtained, where

$$\begin{aligned}
F(x,y) = &F_{60}(t_1,t_2)x^6 + F_{51}(t_1,t_2)x^5y + F_{42}(t_1,t_2)x^4y^2 \\
&+ F_{33}(t_1,t_2)x^3y^3 + F_{24}(t_1,t_2)x^2y^4 + F_{15}(t_1,t_2)xy^5 \\
&+ F_{06}(t_1,t_2)y^6 + F_{50}(t_1,t_2)x^5 + F_{41}(t_1,t_2)x^4y + F_{32}(t_1,t_2)x^3y^2 \\
&+ F_{14}(t_1,t_2)xy^4 + F_{05}(t_1,t_2)y^5 + F_{40}(t_1,t_2)x^4 + F_{31}(t_1,t_2)x^3y \\
&+ F_{22}(t_1,t_2)x^2y^2 + F_{13}(t_1,t_2)xy^3 + F_{04}(t_1,t_2)y^4 \\
&+ F_{30}(t_1,t_2)x^3 + F_{21}(t_1,t_2)x^2y + F_{12}(t_1,t_2)xy^2 + F_{03}(t_1,t_2)y^3 \\
&+ F_{20}(t_1,t_2)x^2 + F_{11}(t_1,t_2)xy + F_{02}(t_1,t_2)y^2 + F_{10}(t_1,t_2)x \\
&+ F_{01}(t_1,t_2)y + F_{00}(t_1,t_2),
\end{aligned} \tag{16}$$

where $F_{ij}(t_1,t_2)$ are polynomials in t_1, t_2 and i, $j = 0,1,2,3,4,5,6$. For example,

$$\begin{aligned}
F_{06} = &16t_1^4t_2^4 + 16t_1^3t_2^5 + 4t_1^2t_2^6 - 64t_1^4t_2^3 - 96t_1^3t_2^4 - 40t_1^2t_2^5 \\
&- 4t_1t_2^6 + 64t_1^4t_2^2 + 192t_1^3t_2^3 + 128t_1^2t_2^4 + 24t_1t_2^5 - 128t_1^3t_2^2 \\
&- 160t_1^2t_2^3 - 48t_1t_2^4 + 64t_1^2t_2^2 + 32t_1t_2^3.
\end{aligned}$$

$F(x,y)$ is a polynomial of degree 6, so it is proved that the curve of foci is of order 6. The next goal is to calculate its intersection points with the absolute line a. In order to do that, since the absolute line has the equation $x_0 = 0$, it is necessary to write the polynomial $F(x,y)$ in homogeneous coordinates. The result is

$$4t_1t_2(-1+t_1)(-2+t_2)(-2+2t_1+t_2)(2t_1+t_2)(x_1-x_2)^2$$
$$(x_1+x_2)^2(-2t_2x_1^2+t_2^2x_1^2-2t_1t_2x_1x_2-t_1x_2^2+t_1^2x_2^2)=0. \quad (17)$$

From the equation (17), it is clear that the absolute points $A_1(0,1,1)$ and $A_2(0,1,-1)$ are double intersections of the curve of foci with the absolute line. It is easy to calculate that those points are not only double intersections, but also double points of the curve of foci. Hence, the curve of foci has type of circularity $(2,2)$. The curve of foci has two more intersections with the absolute line. They are obtained calculating the equation

$$-2t_2x_1^2+t_2^2x_1^2-2t_1t_2x_1x_2-t_1x_2^2+t_1^2x_2^2=0. \quad (18)$$

Those intersections are

$$X_{1,2}(0,\frac{t_1t_2\pm\sqrt{-2t_1t_2+2t_1^2t_2+t_1t_2^2}}{-2t_2+t_2^2},1). \quad (19)$$

It is easy to calculate that those points are precisely the points in which two parabolas of the pencil touch the absolute line. So, if one of two parabolas of the pencil is special parabola, then the curve of foci has type of circularity $(2,3)$ (or $(3,2)$). That is not a case in the Euclidean plane. If both parabolas of the pencil are special parabolas, then the curve of foci has type of circularity $(3,3)$.

Two examples of curves of foci are given in Figures 1 and 2. Six lines connecting four base points of pencils are also shown.

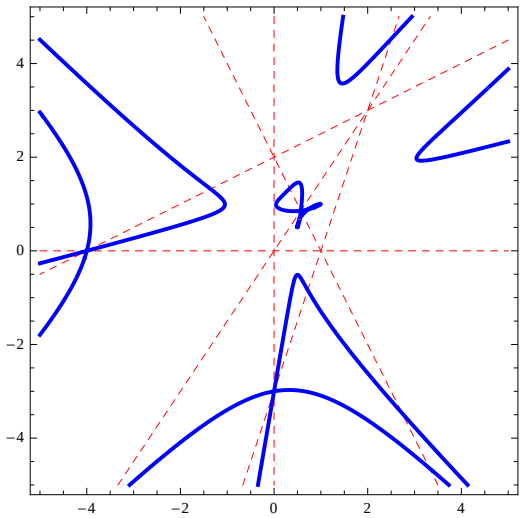


Figure 1: *Curve of foci for* $D(2,3)$

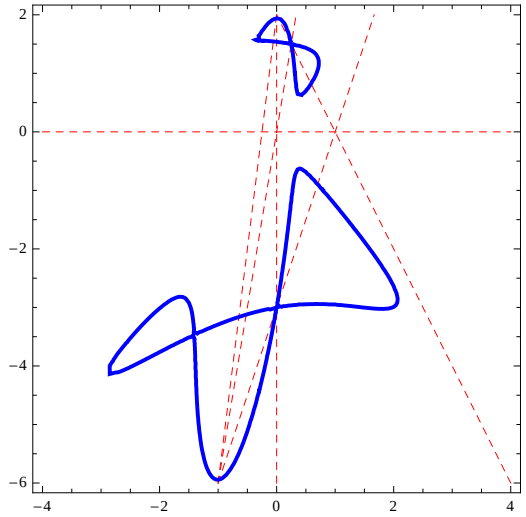


Figure 2: *Curve of foci for* $D(-1,-6)$

The study given above is for the pencil determined by four base points $A(0,0)$, $B(1,0)$, $C(0,2)$ i $D(t_1,t_2)$. In the same way, the investigation may be done more generally, having two base conics

$$c_1(x,y)=a_{00}+a_{11}x^2+a_{22}y^2+2a_{01}x+2a_{02}y+2a_{12}xy,$$
$$c_2(x,y)=b_{00}+b_{11}x^2+b_{22}y^2+2b_{01}x+2b_{02}y+2b_{12}xy.$$

Those two conics intersect in four points and it is assumed that those four points are base points of the pencil. Such conic pencil has the equation

$$c_1(x,y)+\lambda c_2(x,y)=a_{00}+\lambda b_{00}+(a_{11}+\lambda b_{11})x^2$$
$$+(a_{22}+\lambda b_{22})y^2+2(a_{01}+\lambda b_{01})x+2(a_{02}+\lambda b_{02})y$$
$$+2(a_{12}+\lambda b_{12})xy,$$

where $\lambda\in R\bigcup\infty$.
Repeating the process shown in the proof before, the polynomial of degree 6 (which is two long to show it here) is obtained

$$F(x,y)=F_{60}(a_{pr},b_{st})x^6+F_{51}(a_{pr},b_{st})x^5y+F_{42}(a_{pr},b_{st})x^4y^2$$
$$+F_{33}(a_{pr},b_{st})x^3y^3+F_{24}(a_{pr},b_{st})x^2y^4+F_{15}(a_{pr},b_{st})xy^5$$
$$+F_{06}(a_{pr},b_{st})y^6+F_{50}(a_{pr},b_{st})x^5+F_{41}(a_{pr},b_{st})x^4y$$
$$+F_{32}(a_{pr},b_{st})x^3y^2+F_{23}(a_{pr},b_{st})x^2y^3+F_{14}(a_{pr},b_{st})xy^4$$
$$+F_{05}(a_{pr},b_{st})y^5+F_{40}(a_{pr},b_{st})x^4+F_{31}(a_{pr},b_{st})x^3y$$
$$+F_{22}(a_{pr},b_{st})x^2y^2+F_{13}(a_{pr},b_{st})xy^3+F_{04}(a_{pr},b_{st})y^4$$
$$+F_{30}(a_{pr},b_{st})x^3+F_{21}(a_{pr},b_{st})x^2y+F_{12}(a_{pr},b_{st})xy^2$$
$$+F_{03}(a_{pr},b_{st})y^3+F_{20}(a_{pr},b_{st})x^2+F_{11}(a_{pr},b_{st})xy$$
$$+F_{02}(a_{pr},b_{st})y^2+F_{10}(a_{pr},b_{st})x+F_{01}(a_{pr},b_{st})y$$
$$+F_{00}(a_{pr},b_{st}),$$

where F_{ij} are polynomials in a_{pr}, b_{st}, i, $j = 0,1,2,3,4,5,6$ and pr, $st = 00,11,22,01,02,12$. For example,

$$\begin{aligned}F_{06} &= a_{02}a_{11}^2a_{12}b_{02}b_{12}b_{22}^2 - a_{01}a_{11}a_{12}^2b_{02}b_{12}b_{22}^2\\ &- a_{02}^2a_{11}a_{12}b_{11}b_{12}b_{22}^2 - a_{01}a_{02}a_{11}a_{22}b_{11}b_{12}b_{22}^2\\ &+ 2a_{01}^2a_{12}a_{22}b_{11}b_{12}b_{22}^2 - a_{01}a_{02}a_{11}a_{12}b_{12}^2b_{22}^2\\ &+ a_{01}^2a_{12}^2b_{12}^2b_{22}^2 - a_{02}a_{11}^2a_{12}b_{01}b_{22}^3 + a_{01}a_{11}a_{12}^2b_{01}b_{22}^3\\ &+ a_{01}a_{02}a_{11}a_{12}b_{11}b_{22}^3 - a_{01}^2a_{12}^2b_{11}b_{22}^3.\end{aligned}$$

Therefore, the curve of foci is of order 6.

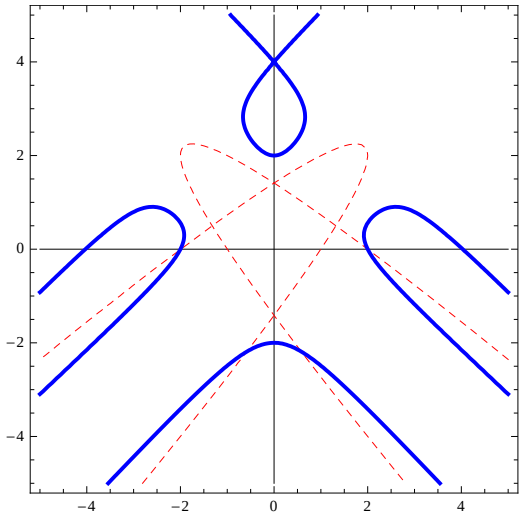


Figure 3: *Entirely circular curve of foci for a pencil containing two special parabolas*

Another example is shown in Figure 3. This is the example for entirely circular curve of foci containing two special parabolas

$$ps_1(x,y) = x^2 - 2xy + y^2 + x - 2,$$
$$ps_2(x,y) = x^2 + 2xy + y^2 - x - 2,$$

which are both shown in Figure 3 together with the curve of foci.
Hence, the following theorems are proved.

Theorem 1 *A curve of foci of all conics of an order conic pencil in PE-plane is a curve of order 6 with the type of circularity* $(2,2)$.

Theorem 2 *If one of two parabolas in the order conic pencil is a special parabola, the curve of foci has the type of circularity* $(3,2)$ *(or* $(2,3)$*).*

Theorem 3 *If both parabolas of the pencil are special parabolas, the curve of foci has the type of circularity* $(3,3)$*, i.e. it is entirely circular.*

3 Curves of foci of order less than 6 - examples

Depending on a position of four base points of the pencil, the curve of foci can be of order less than 6.

Example 1 *For base points* $A(0,0)$, $B(1,0)$, $C(0,2)$ *and* $D(1,3)$*, the line CD is isotropic. The curve of foci is of order 5,*

$$\begin{aligned}F(x,y) &= -1620x^4y + 3240x^3y^2 + 1620x^2y^3 - 3240xy^4\\ &+ 2025x^4 - 9072x^3y + 1620x^2y^2 + 10368xy^3 + 1620y^4\\ &+ 6480x^3 - 5832x^2y - 12960xy^2 - 5184y^3 + 1944x^2\\ &+ 10368xy + 5184y^2 - 5184x - 2592y + 1296.\end{aligned}$$

Example 2 *For base points* $A(0,0)$, $B(1,0)$, $C(0,2)$, $D(1,-2)$*, intersections of the lines AC and BD as well as of the lines AD and BC are isotropic points, i.e. the lines AC and BD as well as the lines AD and BC are parallel. The curve of foci is of order 4,*

$$\begin{aligned}F(x,y) &= 2048x^3y + 5120x^2y^2 + 2048xy^3 - 3072x^2y - 5120xy^2\\ &- 1024y^3 + 1024x^2 + 2048xy + 2304y^2 - 1024x - 512y + 256.\end{aligned}$$

Example 3 *For base points A(0,0), B(1,1), C(0,3), D(1,2), the line AB is isotropic line through the absolute point* A_1*, the line CD is isotropic line through the absolute point* A_2 *and the lines AC and BD are parallel (i.e. they intersect in an isotropic point). The curve of foci is of order 3.*

$$F(x,y) = 16x^3 - 16xy^2 - 60x^2 + 48xy + 8y^2 + 36x - 24y - 9.$$

Example 4 *For base points* $A(0,0)$, $B(1,0)$, $C(0,1,t_1)$ *and* $D = A_2(0,1,-1)$ *where C and D are written in homogeneous coordinates, the curve of foci is*

$$F(x,y) = (2t_1x^2 + 2t_1^2x^2 + y^2 - t_1^2y^2 - 2t_1x - 2t_1^2x + t_1^2)^2,$$

i.e. the curve of foci is a conic. For $0 < t_1 < 1$ *it is an ellipse, and for* $t_1 < 0$ *and* $t_1 > 1$ *it is a hyperbola.*

Example 5 *For base points* $A(0,0)$, $B(1,1)$, $C(0,1,t_1)$ *and* $D = A_2(0,1,-1)$*, the line AB is isotropic and the curve of foci is a special hyperbola which does not exist in the Euclidean plane.*

Example 6 *For base points* $A(0,0)$, $B(1,0)$, $C = A_1(0,1,1)$ *and* $D = A_2(0,1,-1)$*, all conics in the pencil are circles and the curve of foci is a line* $x = \frac{1}{2}$.

4 Conclusion

In this article, it is proved that the curve of foci of a conic pencil in PE-plane is generally a curve of order 6 with the type of circularity $(2,2)$. It is also shown that, depending on the type of parabolas in the pencil, the curve of foci may have the type of circularity $(2,3)$ or $(3,2)$ which is not possible in the Euclidean plane. If both parabolas in the pencil are special parabolas, the curve of foci is entirely circular, i.e. its type of circularity is $(3,3)$. Some examples are shown that the curve of foci may be of order less than 6, but there are some more cases which do not happen in the Euclidean plane. They are possible themes for the next article.

References

[1] R. CESAREC, *Analitička geometrija linearnog i kvadratnog područja - I.Analitička geometrija u ravnini*, Školska knjiga, Zagreb, 1957.

[2] H. S. M. COXETER, *Non-Euclidean Geometry*, 6th ed., Mathematical Association of America, Washington D. C., 1998.

[3] L. HEFFTER, C. KOEHLER, *Lehrbuch der analytischen Geometrie*, Druck und Verlag von B. G. Teubner, Leipzig/Berlin, 1905.

[4] E. JURKIN, N. KOVAČEVIĆ, Entirely Circular Quartics in the Pseudo-Euclidean Plane, *Acta Math. Hungar.* **134**/4 (2012), 571–582.

[5] M. KATIĆ ŽLEPALO, E. JURKIN, Circular cubics and quartics obtained as pedal curves of conics in pseudo-Euclidean plane, *15th International Conference on Geometry and Graphics*, Montreal, 2012.

[6] N. KOVAČEVIĆ, E. JURKIN, Circular Cubics and Quartics in Pseudo-Euclidean Plane Obtained by Inversion, *Math. Pann.* **22**/2 (2011), 199–218.

[7] N. KOVAČEVIĆ, V. SZIROVICZA, Inversion in Minkowskischer Geometrie, *Math. Pann.* **21**/1 (2010), 89–113.

[8] N. M. MAKAROVA, *Proektivnie merooopredelenija ploskosti*, Uchenye zazapiski Moskovskogo pedagogicheskogo instituta, Moskva, 1965.

[9] N. V. REVERUK, *Krivie vtorogo porjadka v psevdoevklidovoi geometrii*, Uchenye zazapiski Moskovskogo pedagogicheskogo instituta 253, Moskva, 1969.

[10] H. SACHS, *Ebene Isotrope Geometrie*, Wieweg, Braunschweig/Wiesbaden, 1987.

[11] A. SLIEPČEVIĆ, M. KATIĆ ŽLEPALO, Pedal Curves of conics in pseudo-Euclidean plane, *Math. Pann.* **23**/1 (2012), 75–84.

Mirela Katić Žlepalo
e-mail: mkatic@tvz.hr
University of Applied Sciences,
Avenija Većeslava Holjevca 15, Zagreb, Croatia

Stručni rad
Prihvaćeno 16. 9. 2016.

IVANČICA MIROŠEVIĆ

Algoritam k-sredina

k-means Algorithm

ABSTRACT

In this paper, k-means algorithm is presented. It is a heuristic algorithm for solving NP-hard optimisation problem of classifying a given data into clusters, with a number of clusters fixed apriori. The algorithm is simple and it's convergence is fast, what makes it widely used, despite its tendency of stoping in a local minimum and inability of recognizing clusters not separated by hyper-planes.

The method of the first variation as a tool for escaping from a local minimum is also presented in the paper.

Key words: k-means algorithm, clustering, first variation method

MSC2010: 65K10

Algoritam k-sredina

SAŽETAK

U članku je objašnjen algoritam k-sredina (k-means algorithm), heuristika koja rješava NP teški optimizacijski problem razvrstavanja podataka (točaka) u skupine (klastere) s unaprijed zadanim brojem skupina. Zbog jednostavnosti i brzine konvergencije, algoritam je u širokoj primjeni, unatoč tendenciji zapinjanja u lokalnom minimumu, te nemogućnosti prepoznavanja skupina koje nisu razdvojive hiperravninama.

U članku je također objašnjena i metoda prve varijacije, heuristika lokalnog traženja kojom algoritam "izvlačimo" iz lokalnog minimuma.

Ključne riječi: algoritam k-sredina, klasteriranje, metoda prve varijacije

1 Uvod

Osnovna ideja algoritma k-sredina je određivanje predstavnika k skupina, i pridruživanje svake točke skupini s najbližim predstavnikom tako da zbroj kvadrata udaljenosti točaka od predstavnika skupina kojima pripadaju bude minimalan. Drugim riječima, algoritam k-sredina generira skupine s minimalnom totalnom varijancom (najkompaktnije moguće skupine).

Nedostatak algoritma je u tome što se na izlazu dobiva samo stabilno, a ne nužno i optimalno rješenje. Odnosno, rješenje bitno ovisi o početnoj k-torci predstavnika skupina. Drugi nedostatak algoritma k-sredina je u tome što može prepoznati samo skupine odvojive hiperravninama.

2 Algoritam k-sredina

Neka je zadan skup točaka $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ u n dimenzionalnom Euklidskom prostoru $\mathbb{R}^n$. Cilj algoritma k-sredina je pronaći, za unaprijed zadani broj $k \geq 2$, optimalnu k-particiju skupa X, $\pi = \{C_1, C_2, ..., C_k\}$, odnosno razmjestiti m točaka skupa X u k skupina (podskupova, klastera) $C_1, C_2, ..., C_k$. Svakoj skupini C_i, $i = 1, \ldots, k$, pridružena je točka

$$\mathbf{c}_i = \frac{1}{|C_i|} \sum_{\mathbf{x} \in C_i} \mathbf{x} \tag{1}$$

koju nazivamo predstavnikom skupine (predstavnik skupine ne mora biti element skupa X). Ovdje je $|C_i|$ kardinalnost od C_i, tj. broj točaka koje pripadaju skupini C_i. Algoritam k-sredina se zasniva na jednostavnoj činjenici da je optimalni izbor predstavnika skupine središte same skupine.

Svakoj particiji $\pi = \{C_1, C_2, ..., C_k\}$ skupa X pridružena je vrijednost ciljne funkcije

$$J = J(\pi) = \sum_{i=1}^{k} \sum_{x \in C_i} ||\mathbf{x} - \mathbf{c}_i||_2^2,$$

gdje su $\mathbf{c}_i$, $i = 1, \ldots, k$, definirani s (1). Algoritam k-sredina u nizu iteracija nastoji minimizirati vrijednost ciljne funkcije, odnosno pronaći particiju kojoj je pridružena najmanja vrijednost ciljne funkcije.

Ovdje je prikazana klasična verzija algoritma k-sredina koja koristi Euklidsku metriku na $\mathbb{R}^n$

$$d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||_2 = \left[\sum_{i=1}^{n} (x_i - y_i)^2 \right]^{\frac{1}{2}}, \ \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{2}$$

Inače, algoritam za particioniranje može se lako prilagoditi tako da koristi kosinusnu sličnost među vektorima kao

metriku,

$$d(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||_2||\mathbf{y}||_2}, \ \mathbf{x},\mathbf{y} \in \mathbb{R}^n, \tag{3}$$

što može biti prikladniji izbor za particioniranje podataka od Euklidske metrike (vektori se prethodno normaliziraju, pa se ovakav algoritam k-sredina naziva sfernim algoritmom k-sredina).

Prije pokretanja algoritma k-sredina potrebno je na neki način zadati početne vrijednosti $\{\mathbf{c}_i\}$ predstavnika skupina, na primjer nasumičnim izborom k točaka iz zadanog skupa točaka X.

Algoritam k-sredina je iterativni algoritam koji ponavlja dva osnovna koraka dok se ne zadovolji neki kriterij konvergencije. U prvom koraku ("*pridruživanje*") svaka točka pridružuje se najbližem predstavniku (određuje se particija). U drugom koraku ("*prepravljanje*") skup predstavnika se prepravlja - za nove predstavnike se uzimaju središta skupina definiranih u koraku "pridruživanje". Algoritam se zaustavlja kada se vrijednost ciljne funkcije prestane smanjivati.

Algoritam 1 *Algoritam k-sredina*

1. ***Inicijalizacija.***
   Zadaj početni skup predstavnika $\{\mathbf{c}_i^0\}_{i=1}^k$. *Postavi brojač* $l = 0$;

2. ***Pridruživanje.***
   Za svaku točku $\mathbf{x} \in X \subset \mathbb{R}^n$ *odredi* $sk(\mathbf{x}) \in \{1,\ldots,k\}$ *(redni broj skupine točke* $\mathbf{x}$*) takav da je*

   $$||\mathbf{c}^l_{sk(\mathbf{x})} - \mathbf{x}||_2 = \min_{j\in\{1,2,\ldots,k\}} ||\mathbf{c}^l_j - \mathbf{x}||_2.$$

   Definiraj skupine

   $$C_i^{(l+1)} = \{\mathbf{x} : sk(\mathbf{x}) = i\}, \ i = 1,\ldots,k.$$

3. ***Prepravljanje.***
   Izračunaj predstavnike novih skupina definiranih u 2. koraku:

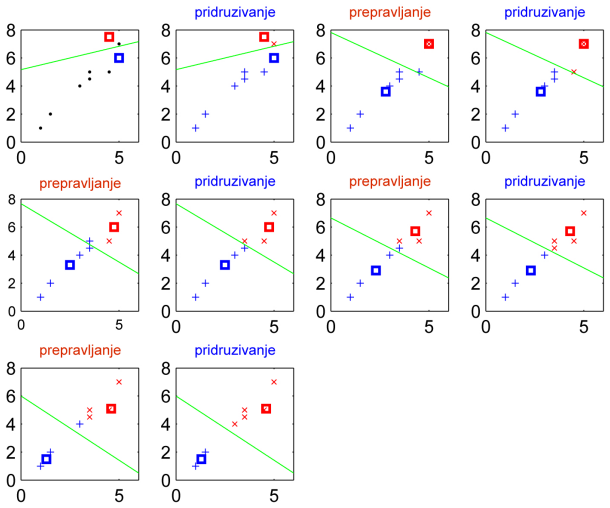
   $$\mathbf{c}_i^{l+1} = \frac{1}{n_i} \sum_{\mathbf{x}\in C_i^{(l+1)}} \mathbf{x},$$

   gdje je $n_i = |C_i^{(l+1)}|$.
   Izračunaj vrijednost ciljne funkcije

   $$J^{(l+1)} = \sum_{i=1}^k \sum_{x\in C_i^{(l+1)}} ||\mathbf{x} - \mathbf{c}_i^{l+1}||_2^2;$$

   Stavi $l = l+1$; *Ponavljaj korake 2. i 3. sve dok se vrijednost J ne prestane smanjivati.*

Primjer 1 Na Slici 1 dan je primjer skupa od sedam točaka $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_7\} = \{(1,1),(1.5,2),(3,4),(3.5,4.5),(3.5,5),(4.5,5),(5,7)\}$, s inicijalnim skupom predstavnika $\{\mathbf{c}_1,\mathbf{c}_2\} = \{(4.5,7.5),(5,6)\}$ za kojeg algoritam k-sredina u pet iteracija daje optimalnu particiju $\pi = \{C_1,C_2\}$, $C_1 = \{\mathbf{x}_1,\mathbf{x}_2\}$ i $C_2 = \{\mathbf{x}_3,\ldots,\mathbf{x}_7\}$.



Slika 1: *Algoritam k-sredina pronalazi optimalnu particiju za zadane inicijalne predstavnike u 5 iteracija. Crvenim znakom 'x' označene su točke prve skupine, a plavim znakom '+' točke druge skupine. Kvadrati označavaju predstavnike skupina.*

Teorem 1 ([3]) *Vrijednost J ciljne funkcije algoritma k-sredina monotono se smanjuje.*

Dokaz. Označimo s $J^{(l)}$ vrijednost ciljne funkcije u l-tom ponavljanju. Vrijedi:

$$J^{(l)} = \sum_{i=1}^k \sum_{x\in C_i^{(l)}} ||\mathbf{x} - \mathbf{c}_i^l||_2^2 \geq \sum_{i=1}^k \sum_{x\in C_i^{(l)}} ||\mathbf{x} - \mathbf{c}^l_{sk(\mathbf{x})}||_2^2 =$$

$$= \sum_{i=1}^k \sum_{x\in C_i^{(l+1)}} ||\mathbf{x} - \mathbf{c}^l_{sk(\mathbf{x})}||_2^2 \geq \sum_{i=1}^k \sum_{x\in C_i^{(l+1)}} ||\mathbf{x} - \mathbf{c}_i^{l+1}||_2^2 = J^{(l+1)}.$$

Druga nejednakost slijedi iz činjenice da vektor središta skupine minimizira kvadratnu devijaciju. □

Algoritam se zaustavlja ako je $J^{(l)} = J^{(l+1)}$.

Primijetimo da su rezultat algoritma k-sredina skupovi (klasteri) odijeljeni hiperravninama. Na primjer, pogledajmo slučaj za $k = 2$:

$\mathbf{x} \in C_1$ akko $||\mathbf{x}-\mathbf{c}_1||_2^2 \leq ||\mathbf{x}-\mathbf{c}_2||_2^2$.

Skup točaka koje su jednako udaljene od $\mathbf{c}_1$ i $\mathbf{c}_2$ dijeli točke na ovaj način. A taj skup točaka je hiperravnina okomita na $\mathbf{c}_1 - \mathbf{c}_2$ koja prolazi polovištem kojeg određuje $\mathbf{c}_1 - \mathbf{c}_2$:

$$(\mathbf{x}-\mathbf{c}_1)^T(\mathbf{x}-\mathbf{c}_1) = (\mathbf{x}-\mathbf{c}_2)^T(\mathbf{x}-\mathbf{c}_2)$$
$$\Longrightarrow 2(\mathbf{c}_1-\mathbf{c}_2)^T\mathbf{x} + \left(\mathbf{c}_2^T\mathbf{c}_2 - \mathbf{c}_1^T\mathbf{c}_1\right) = 0.$$

Općenito, algoritam k-sredina razdjeljuje točke skupom hiperravnina. Rezultirajuća particija prostora odgovara tzv. Voronojevom dijagramu skupa predstavnika.

Definicija 1 *Neka je zadan skup* $S = \{\mathbf{p}_i \in \mathbb{R}^n, i = 1,\ldots,q\}$ *za neki* $q \in \mathbb{N}$. *Voronojeva ćelija (engl. Voronoi cell) pridružena točki* p_i *je skup*

$$V(\mathbf{p}_i) = \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x}-\mathbf{p}_i||_2 \leq ||\mathbf{x}-\mathbf{p}_j||_2, \forall j = 1,\ldots,q\}.$$

Voronojev dijagram skupa S je unija Voronojevih ćelija, odnosno

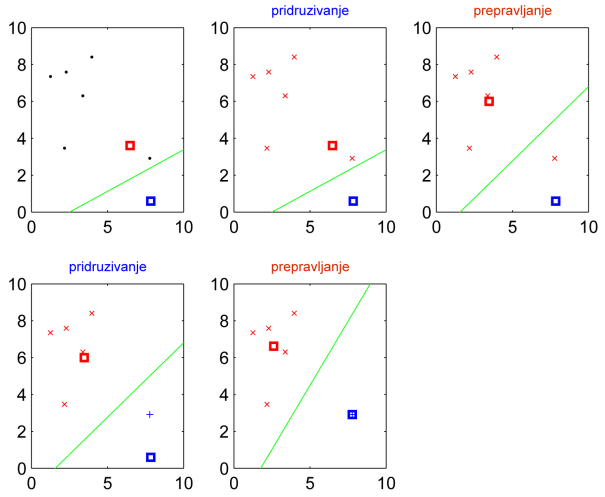
$$V(S) = \bigcup_{i\in\{1,\ldots,q\}} V(\mathbf{p}_i).$$

Primjer 2 Na Slici 1 zelenim crtama naznačeni su pravci kojima algoritam k-sredina razdjeljuje zadane točke na dvije skupine. Dvije rezultirajuće poluravnine predstavljaju dvije Voronojeve ćelije pridružene predstavnicima skupina.

Prilikom izvršavanja algoritma k-sredina može se dogoditi da neki predstavnik ostane bez pridruženih mu točaka (ako su sve točke bliže ostalim predstavnicima). Sama heuristika k-sredina ne govori o tome kako postupiti u situaciji kada skupina u nekom koraku ostane prazna. Uobičajene strategije u praksi su:

- premiještanje predstavnika ispražnjene skupine slučajnim izborom ili proizvoljno
- uzimanje točke koja je najudaljenija od svog predstavnika za novu jednočlanu skupinu
- zadržavanje predstavnika (ostavlja se mogućnost da skupina ponovo primi točke);
- brisanje predstavnika skupine koja je ostala prazna.

Primjer 3 Slika 2 prikazuje primjer particioniranja algoritmom k-sredina u kojem jedan predstavnik ostaje bez pridruženih točaka. Sam predstavnik je zadržan, i već u sljedećem koraku točke su razdijeljene na dvije skupine. Primijetimo da bi u ovom primjeru pomicanje predstavnika u točku najudaljeniju od pridruženog predstavnika rezultiralo većim brojem iteracija.



Slika 2: *Primjer inicijalizacije algoritma k-sredina koja rezultira praznom skupinom. Predstavnik je zadržan, i u sljedećem koraku skupina prima točke.*

3 Dvojnost ciljne funkcije algoritma k-sredina

Pokazat ćemo da algoritam k-sredina smanjivanjem vrijednosti ciljne funkcije automatski povećava prosječnu kvadratnu udaljenost među parovima predstavnika skupina, a smanjuje prosječni zbroj kvadrata razlika udaljenosti parova točaka unutar skupina.

Za $\mathbf{x} \in \mathbb{R}^n$ vrijedi

$$\mathrm{tr}\left(\sum \mathbf{x}\mathbf{x}^T\right) = \sum \mathbf{x}^T\mathbf{x},$$

gdje je

$$\mathrm{tr}A = \sum_{i=1}^{n} a_{ii} \quad \text{za} \quad A = [a_{ij}] \in \mathbb{R}^{n\times n}.$$

Stoga se ciljna funkcija algoritma k-sredina može shvatiti kao $\mathrm{tr}\,(S_W)$, gdje je

$$S_W = \sum_{i=1}^{k} \sum_{\mathbf{x}\in C_i} (\mathbf{x}-\mathbf{c}_i)(\mathbf{x}-\mathbf{c}_i)^T.$$

Matrica S_W naziva se matricom rasipanja unutar skupina (engl. within-class scatter matrix).

Matrica ukupnog rasipanja je dana izrazom

$$S_T = \sum_{\mathbf{x}\in X} (\mathbf{x}-\mathbf{c})(\mathbf{x}-\mathbf{c})^T,$$

gdje je $\mathbf{c}$ srednja vrijednost cijelog skupa podataka. Može se dokazati da je $S_T = S_W + S_B$, gdje je S_B matrica rasipanja između skupina definirana izrazom

$$S_B = \sum_{i=1}^{k} |C_i| (\mathbf{c}_i-\mathbf{c})(\mathbf{c}_i-\mathbf{c})^T.$$

Propozicija 1 *Vrijedi*

$S_T = S_W + S_B$.

Dokaz. Definirajmo za proizvoljni vektor $\mathbf{a}$

$$\begin{aligned} S_T(\mathbf{a}) &= \sum_{\mathbf{x}} (\mathbf{x}-\mathbf{a})(\mathbf{x}-\mathbf{a})^T \\ &= \sum_i \sum_{\mathbf{x}\in C_i} (\mathbf{x}-\mathbf{a}+\mathbf{c}_i-\mathbf{c}_i)(\mathbf{x}-\mathbf{a}+\mathbf{c}_i-\mathbf{c}_i)^T \\ &= \sum_i \sum_{\mathbf{x}\in C_i} ((\mathbf{x}-\mathbf{c}_i)+(\mathbf{c}_i-\mathbf{a}))\left((\mathbf{x}-\mathbf{c}_i)^T+(\mathbf{c}_i-\mathbf{a})^T\right) \\ &= \sum_i \sum_{\mathbf{x}\in C_i} (\mathbf{x}-\mathbf{c}_i)(\mathbf{x}-\mathbf{c}_i)^T + \sum_i \sum_{\mathbf{x}\in C_i} (\mathbf{x}-\mathbf{c}_i)(\mathbf{c}_i-\mathbf{a})^T \\ &\quad + \sum_i \sum_{\mathbf{x}\in C_i} (\mathbf{c}_i-\mathbf{a})(\mathbf{x}-\mathbf{c}_i)^T + \sum_i \sum_{\mathbf{x}\in C_i} (\mathbf{c}_i-\mathbf{a})(\mathbf{c}_i-\mathbf{a})^T \end{aligned}$$

Vrijedi

$$\begin{aligned} S_T(\mathbf{a}) &= \sum_i \sum_{\mathbf{x}\in C_i} (\mathbf{x}-\mathbf{c}_i)(\mathbf{x}-\mathbf{c}_i)^T + \sum_i |C_i|(\mathbf{c}_i-\mathbf{a})(\mathbf{c}_i-\mathbf{a})^T \\ &= S_W + S_B(\mathbf{a}), \end{aligned}$$

jer je $\sum_i \sum_{\mathbf{x}\in C_i} (\mathbf{c}_i-\mathbf{a})(\mathbf{x}-\mathbf{c}_i)^T = 0 = \sum_i \sum_{\mathbf{x}\in C_i} (\mathbf{x}-\mathbf{c}_i)(\mathbf{c}_i-\mathbf{a})^T$.

Prethodna relacija je točna za proizvoljni vektor $\mathbf{a}$, pa je točna i za $\mathbf{a} = \mathbf{c}$, što dokazuje tvrdnju. □

Možemo, također, pokazati da je

$$\begin{aligned} S_T &= \frac{1}{2m} \sum_{i=1}^{m} \sum_{j=1}^{m} (\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^T \\ S_W &= \sum_{i=1}^{k} \frac{1}{2|C_i|} \sum_{\mathbf{x}\in C_i} \sum_{\mathbf{y}\in C_i} (\mathbf{x}-\mathbf{y})(\mathbf{x}-\mathbf{y})^T \\ S_B &= \frac{1}{2m} \sum_{i=1}^{k} \sum_{j=1}^{k} |C_i||C_j| (\mathbf{c}_i-\mathbf{c}_j)(\mathbf{c}_i-\mathbf{c}_j)^T. \end{aligned}$$

Budući da je $\operatorname{tr}(S_W)$ ciljna funkcija algoritma k-sredina, on nastoji minimizirati prosječni zbroj kvadrata razlika parova točaka unutar skupina.

Prema tome, budući da je $\operatorname{tr}(S_T) = \operatorname{tr}(S_W) + \operatorname{tr}(S_B)$, a $\operatorname{tr}(S_T)$ je fiksan, ciljnu funkciju algoritma k-sredina možemo promatrati kao

$$\begin{aligned} \max \operatorname{tr}(S_B) &= \max \sum_i |C_i| \cdot ||\mathbf{c}_i-\mathbf{c}||_2^2 \\ &= \max \frac{1}{2m} \sum_i \sum_j |C_i||C_j| \cdot ||\mathbf{c}_i-\mathbf{c}_j||_2^2, \end{aligned}$$

što znači da algoritam k-sredina pokušava maksimizirati prosječnu kvadratnu udaljenost među parovima predstavnika skupina.

4 O međama broja iteracija algoritma k-sredina u jednodimenzionalnom modelu

Računsku složenost algoritma k-sredina po jednoj iteraciji na m-članom skupu $X \subset \mathbb{R}^n$ možemo razložiti po koracima:

- U koraku "*pridruživanje*" algoritam k-sredina mk puta računa udaljenosti, za što mu treba $3mkn$ osnovnih operacija (zbrajanja, množenja ili uspoređivanja). Za pronalaženje najbližeg predstavnika treba mu mk operacija. Prema tome, složenost pridruživanja je $O(mkn)$.

- U koraku "*prepravljanje*" algoritam k-sredina izvršava mn zbrajanja i kn dijeljenja. Složenost prepravljanja je $O(mn)$ $(k \leq m)$.

- Za računanje vrijednosti ciljne funkcije algoritmu k-sredina treba m zbrajanja udaljenosti izračunatih u prvom koraku, pa je složenost računanja vrijednosti ciljne funkcije $O(m)$.

Prema tome, složenost algoritma k-sredina je $O(mknt)$, gdje je t broj iteracija. Sam broj iteracija može varirati od nekoliko do nekoliko tisuća, ovisno o broju i distribuciji točaka, te o broju skupina, i još uvijek nije u potpunosti teoretski razjašnjen broj iteracija koje algoritam k-sredina može izvršiti u najgorem slučaju. U [4] je dokazano da različitih Voronojevih dijagrama zadanih s k centara, koji m-člani skup $X \in \mathbb{R}^n$ dijele na k podskupova, ima najviše $O(m^{kn})$, što predstavlja i trivijalnu gornju među za broj iteracija algoritma k-sredina. Međutim, činjenica da u uobičajenim aplikacijama k može biti veličine nekoliko stotina, i da algoritam k-sredina u praksi relativno brzo konvergira, čini ovu među beznačajnom.

5 Prva varijacija - algoritam za profinjenje rezultata algoritma *k*-sredina

Veliki nedostatak algoritma *k*-sredina je u tome što može "zapeti" u lokalnom minimumu i rezultirati lošom particijom. Navest ćemo primjer kada loša inicijalizacija algoritma rezultira particijom koja očigledno nije optimalna. Prva varijacija je metoda lokalnog traženja kojom se omogućava izbjegavanje lokalnog minimuma.

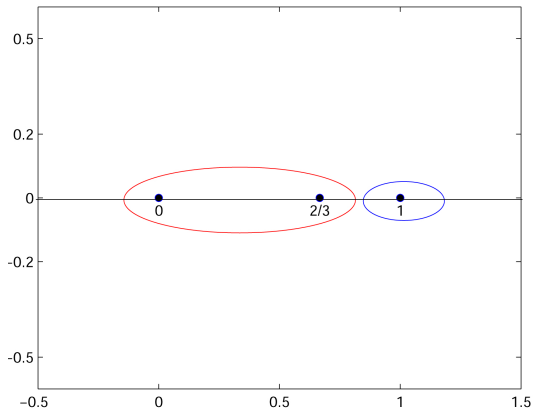
Inače, u praksi se pokazalo da se nedostatci algoritma *k*-sredina najčešće očituju kod malih skupova točaka (do otprilike 200 točaka) [2].

Primjer 4 Neka je zadan skup $X = \{x_1, x_2, x_3\}$, $x_1 = 0$, $x_2 = \frac{2}{3}$, $x_3 = 1$ i inicijalni par predstavnika $c_1^0 = \frac{1}{3}$, $c_2^0 = 1$ (Slika 3). To znači da je u početnoj particiji $C_1^0 = \{x_1, x_2\}$ i $C_2^0 = \{x_3\}$ i vrijedi

$$\begin{aligned} J\left(\{C_1^0, C_2^0\}\right) &= \sum_{i=1}^{2} \sum_{x \in C_i^0} ||\mathbf{x} - \mathbf{c}_i^0||_2^2 \\ &= ||\mathbf{x}_1 - \mathbf{c}_1^0||_2^2 + ||\mathbf{x}_2 - \mathbf{c}_1^0||_2^2 + ||\mathbf{x}_3 - \mathbf{c}_2^0||_2^2 \\ &= \left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + 0 = \frac{2}{9}. \end{aligned}$$

Algortam *k*-sredina neće promijeniti ovakvu početnu particiju jer je $||x_2 - c_1^0||_2^2 = ||x_2 - c_2^0||_2^2$, iako za particiju $\{C_1^1, C_2^1\}$, $C_1^1 = \{x_1\}$ i $C_2^1 = \{x_2, x_3\}$, odnosno za $c_1^1 = 0$, $c_2^1 = \frac{5}{6}$, vrijedi

$$\begin{aligned} J\left(\{C_1^1, C_2^1\}\right) &= \sum_{i=1}^{2} \sum_{x \in C_i^1} ||\mathbf{x} - \mathbf{c}_i^1||_2^2 \\ &= ||\mathbf{x}_1 - \mathbf{c}_1^1||_2^2 + ||\mathbf{x}_2 - \mathbf{c}_2^1||_2^2 + ||\mathbf{x}_3 - \mathbf{c}_2^1||_2^2 \\ &= 0 + \left(\frac{1}{6}\right)^2 + \left(\frac{1}{6}\right)^2 = \frac{1}{18}. \end{aligned}$$



Slika 3: *Primjer početne particije koju algoritam k-sredina ne mijenja, iako nije optimalna.*

Problem zaustavljanja algoritma *k*-sredina u lokalnom minimumu može se djelomično riješiti algoritmom prve varijacije, primjerom heuristike lokalnog traženja.

Definicija 2 *Prva varijacija particije* $\pi = \{C_1, \ldots, C_k\}$ *skupa X je particija* $\pi' = \{C_1', \ldots, C_k'\}$*, koja se dobije pomicanjem jedne točke* $\mathbf{x} \in X$ *iz skupine* $C_i \in \pi$ *u skupinu* $C_j \in \pi$*. Skup svih prvih varijacija particije* $\pi = \{C_1, \ldots, C_k\}$ *označavamo s* $\mathcal{V}(\pi)$*.*

Među svim elementima skupa $\mathcal{V}(\pi)$ tražimo particiju s najmanjom vrijednošću ciljne funkcije.

Definicija 3 *Particija* π^* *je prva varijacija particije* π *skupa X takva da je za svaku prvu varijaciju* π' *skupa X*

$$J(\pi^*) \leq J(\pi').$$

Particija π^* *zove se sljedeća prva varijacija.*

Algoritam prve varijacije generira niz particija $\pi^{(l)} = \{C_1^{(l)}, \ldots, C_k^{(l)}\}$, $l \geq 0$, takav da je $\pi^{(l+1)} = \pi^{(l)*}$, $l = 0, 1, \ldots$

Promotrimo razliku između iteracije algoritma *k*-sredina i iteracije algoritma prve varijacije. Neka je zadana bipartcija $\pi = \{Z, Y\}$ skupa $X \subset \mathbb{R}^n$, pri čemu je $Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ i $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$. Želimo utvrditi treba li jedan vektor, npr. $\mathbf{z}_n$, premjestiti iz Z u Y. Definirajmo potencijalne nove skupine sa

$$Z^- = \{\mathbf{z}_1, \ldots, \mathbf{z}_{n-1}\} \quad \text{i} \quad Y^+ = \{\mathbf{y}_1, \ldots, \mathbf{y}_m, \mathbf{z}_n\}.$$

Algoritam *k*-sredina provjerava vrijednost

$$\Delta_{km} = ||\mathbf{z}_n - \mathbf{c}(Y)||_2^2 - ||\mathbf{z}_n - \mathbf{c}(Z)||_2^2. \tag{4}$$

Ako je $\Delta_{km} < 0$, algoritam *k*-sredina pomiče $\mathbf{z}_n$ iz Z u Y. Inače $\mathbf{z}_n$ ostaje u Z.

Algoritam prve varijacije, međutim, provjerava stvarnu promjenu u vrijednosti ciljne funkcije.

$$\begin{aligned}
\Delta_{pv} &= \sum_{i=1}^{m} ||\mathbf{y}_i - \mathbf{c}\left(Y^+\right)||_2^2 + ||\mathbf{z}_n - \mathbf{c}\left(Y^+\right)||_2^2 \\
&\quad + \sum_{i=1}^{n} ||\mathbf{z}_i - \mathbf{c}\left(Z^-\right)||_2^2 - ||\mathbf{z}_n - \mathbf{c}\left(Z^-\right)||_2^2 \\
&\quad - \sum_{i=1}^{m} ||\mathbf{y}_i - \mathbf{c}\left(Y\right)||_2^2 - \sum_{i=1}^{n} ||\mathbf{z}_i - \mathbf{c}\left(Z\right)||_2^2 \\
&= m||\mathbf{c}(Y) - \mathbf{c}\left(Y^+\right)||_2^2 + n||\mathbf{c}(Z) - \mathbf{c}\left(Z^-\right)||_2^2 \\
&\quad + ||\mathbf{z}_n - \mathbf{c}\left(Y^+\right)||_2^2 - ||\mathbf{z}_n - \mathbf{c}\left(Z^-\right)||_2^2 \\
&= m||\frac{\sum_{i=1}^{m}\mathbf{y}_i}{m} - \frac{\sum_{i=1}^{m}\mathbf{y}_i + \mathbf{z}_n}{m+1}||_2^2 + n||\frac{\sum_{i=1}^{n}\mathbf{z}_i}{n} - \frac{\sum_{i=1}^{n}\mathbf{z}_i - \mathbf{z}_n}{n-1}||_2^2 \\
&\quad + ||\mathbf{z}_n - \frac{\sum_{i=1}^{m}\mathbf{y}_i + \mathbf{z}_n}{m+1}||_2^2 - ||\mathbf{z}_n - \frac{\sum_{i=1}^{n}\mathbf{z}_i - \mathbf{z}_n}{n-1}||_2^2 \\
&= m||\frac{\sum_{i=1}^{m}\mathbf{y}_i - m\mathbf{z}_n}{m(m+1)}||_2^2 + n||\frac{n\mathbf{z}_n - \sum_{i=1}^{n}\mathbf{z}_i}{n(n-1)}||_2^2 \\
&\quad + ||\frac{m\mathbf{z}_n - \sum_{i=1}^{m}\mathbf{y}_i}{m+1}||_2^2 - ||\frac{n\mathbf{z}_n - \sum_{i=1}^{n}\mathbf{z}_i}{n-1}||_2^2 \\
&= \left[\frac{m}{(m+1)^2} + \frac{m^2}{(m+1)^2}\right]||\mathbf{z}_n - \mathbf{c}(Y)||_2^2 \\
&\quad + \left[\frac{n}{(n-1)^2} - \frac{n^2}{(n-1)^2}\right]||\mathbf{z}_n - \mathbf{c}(Z)||_2^2 \\
&= \frac{m}{m+1}||\mathbf{z}_n - \mathbf{c}(Y)||_2^2 - \frac{n}{n-1}||\mathbf{z}_n - \mathbf{c}(Z)||_2^2 \qquad (5)
\end{aligned}$$

Razlika između (4) i (5),

$$\Delta_{km} - \Delta_{pv} = \frac{1}{m+1}||\mathbf{z}_n - \mathbf{c}(Y)||_2^2 + \frac{1}{n-1}||\mathbf{z}_n - \mathbf{c}(Z)||_2^2 \geq 0$$

je zanemariva kada su skupine Z i Y velike. Međutim, $\Delta_{km} - \Delta_{pv}$ može postati bitna kod malih skupina.

Primjer 5 Za $\mathbf{z}_n = \mathbf{x}_2$ iz primjera 4 je $\Delta_{km} = 0$, a $\Delta_{pv} = -\frac{3}{18}$ i to je razlog zašto algoritam k-sredina propušta optimalnu particiju $\{C_1^1, C_2^1\}$.

Algoritam 2 *Algoritam prve varijacije*

1. *Zadaj početnu particiju* $\pi^{(0)} = \{C_1^{(0)}, \ldots, C_k^{(0)}\}$. *Postavi brojač iteracija* $l = 0$.
2. *Generiraj sljedeću prvu varijaciju* $\pi^{(l)*}$.
   Ako je $J\left(\pi^{(l)*}\right) - J\left(\pi^{(l)}\right) < 0$, *postavi* $\pi^{(l+1)} = \pi^{(l)*}$, *povećaj* l *za* 1, *i vrati se na korak* 2.
3. *Stani.*

Prilikom računanja sljedeće prve varijacije algoritam izvršava $3mkn$ operacija za računanje udaljenosti, $3mk$ operacija za računanje vrijednosti ciljne funkcije za sve prve varijacije particije, te mk operacija za određivanje sljedeće prve varijacije, pa je složenost jedne iteracije jednaka složenosti iteracije algoritma k-sredina $(O(mkn))$. Međutim, promjene vrijednosti ciljne funkcije su u svakoj iteraciji jako male jer se pomiče samo jedna točka, dok algoritam k-sredina daje značajnija poboljšanja po iteraciji. Stoga pogledajmo kombinaciju algoritma k-sredina i algoritma prve varijacije:

Algoritam 3 *Algoritam k-sredina poboljšan algoritmom prve varijacije*

1. *Zadaj početnu particiju* $\pi^{(0)} = \{C_1^{(0)}, \ldots, C_k^{(0)}\}$. *Postavi brojač iteracija* $l = 0$.
2. *Generiraj sljedeću particiju* $\pi^{(l)'}$ *algoritmom k-sredina.*
   Ako je $J\left(\pi^{(l)'}\right) - J\left(\pi^{(l)}\right) < 0$, *postavi* $\pi^{(l+1)} = \pi^{(l)'}$, *povećaj* l *za* 1, *i vrati se na korak* 2.
3. *Generiraj sljedeću prvu varijaciju* $\pi^{(l)*}$.
   Ako je $J\left(\pi^{(l)*}\right) - J\left(\pi^{(l)}\right) < 0$, *postavi* $\pi^{(l+1)} = \pi^{(l)*}$, *povećaj* l *za* 1, *i vrati se na korak* 2.
4. *Stani.*

Algoritam 3 alternira između dviju faza:

(a) algoritma k-sredina

(b) algoritma prve varijacije.

U trećem koraku pomiče se samo jedna točka iz jedne skupine u drugu ako to pomicanje rezultira smanjujem vrijednost ciljne funkcije. Niz koraka prve varijacije omogućava izbjegavanje lokalnog minimuma, nakon čega nove iteracije algoritma k-sredina mogu nastaviti brže smanjivati vrijednost ciljne funkcije. Ova ping-pong strategija daje algoritam za profinjenje skupina, koji često poboljšava sam algoritam k-sredina, a računski nije prezahtjevan (složenost je još uvijek $O(mkn)$). Naime, udaljenosti točaka od svih predstavnika potrebne za generiranje sljedeće prve varijacije u trećem koraku već su izračunate u drugom koraku, pa se ne moraju ponovo računati.

Međutim, ni ovako poboljšan algoritam ne daje uvijek optimalno rješenje.

U ovom radu je korištena MatLabova *kmeans* funkcija. Ona se zasniva na dvofaznom iterativnom algoritmu čija

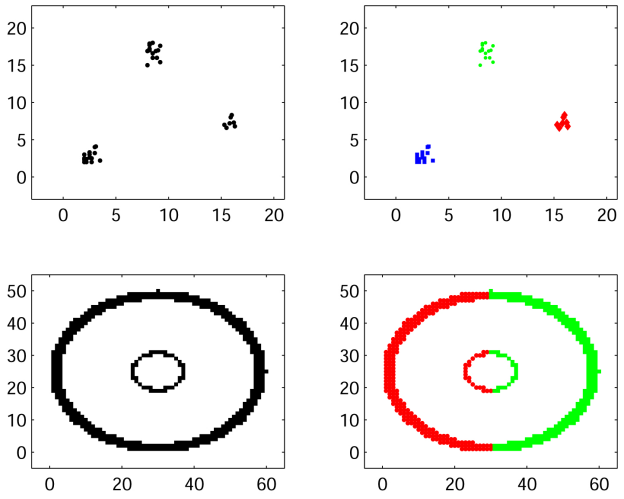
prva faza odgovara klasičnom algoritmu *k*-sredina. U drugoj fazi se pojedinačno premještaju sve točke čije premještanje rezultira smanjenjem ciljne funkcije (zbroja udaljenosti točaka od predstavnika skupina), nakon čega se ponovo računa skup predstavnika. Svaka iteracija se u drugoj fazi sastoji od jednog prolaska kroz sve zadane točke. Ovo je varijanta algoritma prve varijacije. Pritom se samo u prvoj iteraciji prve faze računaju sve udaljenosti točaka od predstavnika. U ostalim iteracijama udaljenosti se računaju samo za točke odnosno predstavnike koji su se pomicali. Budući da se broj pomaknutih točaka relativno mali nakon prvih par iteracija, ovim se bitno smanjuje vrijeme izvršavanja algoritma.

MatLabova *kmeans* funkcija dopušta unos vlastitog inicijalnog skupa predstavnika, ili ga sama računa na osnovi ulaznog skupa točaka na jedan od tri ponuđena načina:

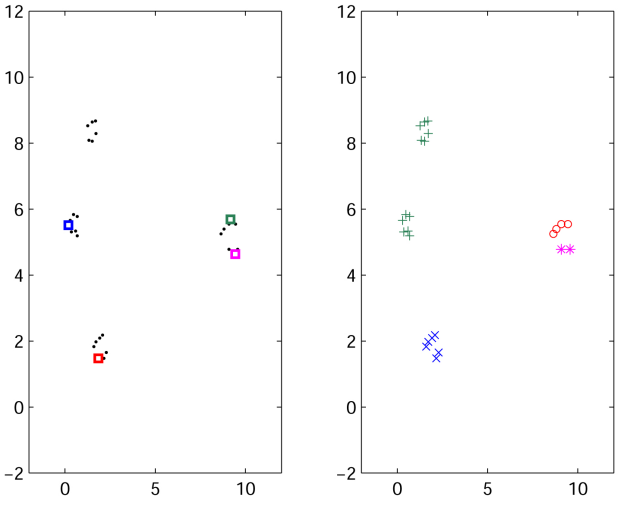
- "sample" - nasumičnim izborom bira *k* predstavnika iz ulaznog skupa točaka (uobičajeni način)
- "uniform" - nasumičnim izborom bira *k* uniformno distribuiranih predstavnika iz ulaznog skupa točaka
- "cluster" - particionira 10% nasumce izabranih točaka i dobivene predstavnike uzima za inicijalne predstavnike cijelog skupa točaka. U uvodnom particioniranju koristi "sample" za inicijalizaciju.

Primjer 6 Na Slici 4 su dva primjera particioniranja algoritmom *k*-sredina. U prvom slučaju algoritam je u 100 pokretanja sa "sample" inicijalizacijom 70 puta dao očiglednu optimalnu particiju. U primjeru s dva koncentrična prstena optimalne skupine nisu linearno odvojive, pa algoritam *k*-sredina daje najbolje što može - dijeli svaki prsten na dva dijela.

Primjer 7 Na Slici 5 (lijevo) je primjer skupa točaka koje su prirodno podijeljene na četiri skupine, te početni skup predstavnika. U ovom slučaju ni algoritam *k*-sredina ni algoritam prve varijacije ne daju optimalnu particiju.



Slika 4: *Rezultat particioniranja točaka MatLabovom* kmeans *funkcijom. Bojom je istaknuta pripadnost točaka skupinama.*



Slika 5: *Primjer inicijalizacije skupa predstavnika (kvadratići) za koju i algoritam k-sredina i algoritam prve varijacije daju lošu particiju. Lijevo su neparticionirane, a desno particionirane točke. Bojom i znakom je istaknuta pripadnost točaka skupinama.*

Literatura

[1] K. Alsabti, S. Ranka, V. Singh, An Efficient K-Means Clustering Algorithm, *IPPS/SPDP Workshop on High Performance Data Mining*, Orlando, Florida, 1998.

[2] I. Dhillon, Y. Guan, J. Kogan, Iterative Clustering of High Dimensional Text Data Augmented by Local Search, *Proceedings of the 2nd IEEE International Conference on Data Mining*, 131–138, Maebishi, Japan, 2002.

[3] I. S. DHILLON, D. S. MODHA, Concept Decompositions for Large Sparse Text Data using Clustering, *Machine Learning*, 42/1, 143–175, 2001.

[4] M. INABA, N. KATOH, H. IMAI, Applications of Weighted Voronoi Diagrams and Randomization to Variance-Based k-Clustering, *Proceedings of the 10th ACM Symposium on Computational Geometry*, 332–339, ACM, 1994.

[5] T. KANUNGO, D. M. MOUNT, N. S. NETANYAHU, C. D. PIATKO, R. SILVERMAN, A. Y. WU, A Local Search Approximation Algorithm for *k*-Means Clustering, *Proceedings of the 18th Annual ACM Symposium on Computational Geometry*, 10–18, 2002.

[6] S. HAR-PELED, B. SADRI, On Lloyd's *k*-means Method, *ACM-SIAM Symposium on Discrete Algorithms*, 2005.

Ivančica Mirošević
ivancica.mirosevic@fesb.hr

Fakultet elektrotehnike, strojarstva i brodogradnje
Sveučilište u Splitu
Ruđera Boškovića 32, 21000 Split

Professional paper
Accepted 20. 12. 2016.

MATE GLAURDIĆ
JELENA BEBAN-BRKIĆ
DRAŽEN TUTIĆ

Graph Colouring and its Application within Cartography

Graph Colouring and its Application within Cartography

ABSTRACT

The problem of colouring geographical political maps has historically been associated with the theory of graph colouring. In the middle of the 19th century the following question was posed: how many colours are needed to colour a map in a way that countries sharing a border are coloured differently. The solution has been reached by linking maps and graphs. It took more than a century to prove that 4 colours are sufficient to create a map in which neighbouring countries have different colours.

Key words: graph, graph colouring, map, map colouring, the four colour theorem

MSC2010: 05C15, 05C90, 86A30, 68R10

O problemu bojanja grafova s primjenom u kartografiji

SAŽETAK

Problem bojanja geografskih političkih karata povijesno je vezan uz teoriju bojanja grafova. Polovicom 19. stoljeća nametnulo se pitanje koliko je boja potrebno da bi se dana geografska karta obojila tako da zemlje koje graniče budu obojane različitim bojama. Do rješenja se došlo povezivanjem karata i grafova. Bilo je potrebno više od jednog stoljeća kako bi se dokazalo da su četiri boje dovoljne za obojiti (geografsku) kartu na takav način da susjedna područja (države) imaju različitu boju.

Ključne riječi: graf, bojanje grafa, karta, bojanje karte, teorem o 4 boje

1 Introduction

The problem of colouring geographical political maps has historically been associated with the theory of graph colouring. In the middle of the 19th century the following question was posed: how many colours are needed to colour a map in a way that countries sharing a border are coloured differently. The solution has been reached by linking maps and graphs. It took more than a century to prove that 4 colours are sufficient to create a map in which neighbouring countries have different colours.

In graph theory, graph colouring is a special case of graph labelling. It is about assigning a colour to graph elements: vertices, edges, regions, with certain restrictions.

With this paper we would like to assess the elements of the theory of graph colouring with an emphasis on its application on practical problems in the field of cartography.

A mathematical basis for map colouring will be given along with the chronology of proving *The Four Colour Theorem*. In addition, world political map will be shown, to determine the minimum number of colours needed to colour a map properly in practice.

2 Elements of Graph Theory

Graph Theory is a special branch of combinatorics closely related to applied mathematics, optimization theory and computer sciences. The simplest and most frequently applied combinatorical structure is a graph, and exactly the simplicity of this structure provides easy transfer and modelling of practical problems in graph terms, as well as the application of known proved theoretical concepts, algorithms and abstract ideas to particular graphs.

2.1 Historical overview of Graph Theory

Graph Theory has rather precise historical aspects. The first paper in the Graph theory was the article "Solutio problematis ad geometriam situs pertinentis", i.e. "The Solution of a Problem Relating to the Geometry of Position" by a well known Swiss mathematician Leonhard Euler (1707-1783) from the year 1736. In this paper, "the Königsberg Bridge Problem" was defined and solved. The Prussian city Königsberg, now Kaliningrad (Russia) occupies both banks of the river Pregolya. The river divides the

city into four territories, two river islands and two coastal areas mutually connected with seven bridges.

The inhabitants of Königsberg sought to solve the issue that troubled them for many years: *Is it possible to walk around the city crossing each of the seven bridges once and only once and to finish the walk through the city at the starting point?*

Leonhard Euler eliminated all features of the terrain except the land masses and bridges presenting it by means of a graph (Fig. 1); the points represent the coastal parts (B and C) and the islands (A and D), while the bridges are presented as graph edges, i.e. as connections of points. His answer to the question mentioned above was clear: such walk is not possible if each part of mainland is not connected with the other parts with an even number of bridges (see Chapter 2.2).

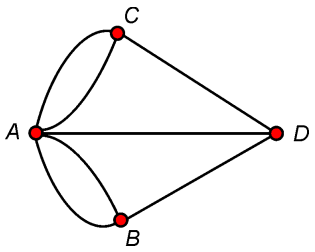


Figure 1: *The Königsberg bridges problem presented by a graph*

Although the origins of the Graph Theory date back as far as the 18th century, it started to develop in the second half of the 20th century. The first book dealing with the topic of Graph Theory was written in 1936 by the Hungarian mathematician D. König, and it is considered to be the beginning of the development of Graph Theory as a separate mathematical discipline. König unified and systematised the earlier results offering the list of 110 published papers where the term graph had appeared explicitly. Among the authors of these papers are famous names like G. Kirchhoff (1824 - 1887) and A. Cayley (1821 -1895). Ever since, graph has become a generally accepted term [6].

Greater development of research in the field of Graph Theory and its applications started in the 60-ties of the twentieth century and has been continuing parallel with the development of information technologies up to the present day.

2.2 Graph Theory basic concepts and definitions

Definition 1 *A graph G consists of a finite non-empty set $V = V(G)$ whose elements are called **vertices**, **points** or **nodes** of G and a finite set $E = E(G)$ of unordered pairs of distinct vertices called **edges** of G.*

Such a graph we denote $G(V,E)$ when emphasizing the two parts of G, (Fig. 2).

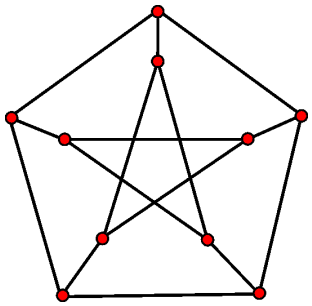


Figure 2: *Example of a simple graph - The Petersen graph*

Definition 2 *An edge $e = u,v$ is said to **join** the vertices u and v, and is usually abbreviated to $e = uv$. In such a case, u and v are called **endpoints** and they are said to be **adjacent**. Further, vertices u and v are said to be **incident** on e and vice versa, the edge e is said to be **incident** on each of its endpoints u and v. Similarly, two distinct edges e and f are adjacent if they have a vertex in common.*

Remark 1 *If two or more edges connect the same endpoints, we call them **multiple edges**. An edge is called a **loop** if its endpoints are the same vertex. The former definition of a graph permits neither multiple edges nor loops. In some texts the term **simple graph** refers to the graph without multiple edges and loops while the one permitting them is called a **multigraph**. Most often it does not matter whether we deal with a simple graph or a multigraf, and if necessary, will be specially emphasized.*

Definition 3 *If the vertex set of a graph G can be split into two disjoint sets X and Y so that each edge of G joins a vertex of X and a vertex of Y, then G is said to be **bipartite**. A **complete bipartite graph** is a bipartite graph in which each vertex in X is joined to each vertex in Y by just one edge. If r is the number if vertices in X and s is the number of vertices in Y we denote this graph $K_{r,s}$, (Fig. 3).*

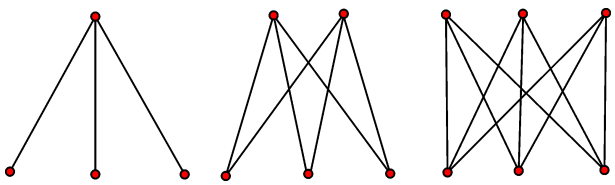


Figure 3: *Complete bipartite graphs $K_{1,3}$, $K_{2,3}$, $K_{3,3}$*

Definition 4 *For the two disjoint graphs $G = (V(G),E(G))$ and $H = (V(H),E(H))$, their union $G \cup H$ is defined by $G \cup H = (V(G) \cup V(H), E(G) \cup E(H))$.*

Definition 5 *A graph G is **connected** if it cannot be represented as the union of two graphs. Otherwise, it is **disconnected**. Any disconnected graph can be represented as the union of connected graphs called **connected components** of G. A graph is said to be **finite** if it has a finite number of vertices and a finite number of edges, otherwise it is **infinite**.*

Graphs within this article shall be finite.

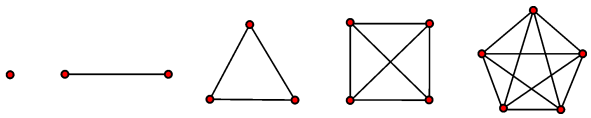
Definition 6 *The **degree** of a vertex v in G, written deg (v), is equal to the number of edges in G incident with v. It shall be taken conventionally that a loop contributes 2 to the degree of v. A vertex of degree zero is called an **isolated vertex** and a vertex of degree 1 is an **end-vertex**.*

Definition 7 *Consider the graphs $G = (V(G), E(G))$ and $H = (V(H), E(H))$. H is a **subgraph** of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. A subgraph H of G is said to be a **spanning subgraph** of G if $V(H) = V(G)$.*

Subgraphs are often obtained from a given graph by deleting its vertices and edges. Specifically, if v is a vertex in G, $G - v$ is a subgraph of G obtained by deleting v and all edges incident with v. Similarly, if e is an edge in G, $G - e$ is a subgraph of G obtained by deleting e from G. However, it is easily seen that contracting an edge of a graph does not give a subgraph. **Contracting** an edge e from G means removing it and identifying its ends u and v so that the resulting vertex is incident with those edges that were originally incident with u or v. Such a graph is denoted by $G|e$.

Definition 8 *A graph G is said to be **complete** if every vertex in G is adjacent to every other vertex in G. A complete graph with n vertices is denoted by K_n, (Fig. 4).*

$K_n : s$ is used to denote a complete graph with $|V| = n$ and $|E| = s$. It is easy to check that K_n has $s = \frac{n(n-1)}{2}$ edges.



$K_1 : 0$ $K_2 : 1$ $K_3 : 3$ $K_4 : 6$ $K_5 : 10$

Figure 4: *Some complete graphs*

Definition 9 *A **walk** in a graph G is an alternating sequence of vertices and edges of the form $v_0, e_1, v_1, e_2, \ldots, e_k, v_k$, where each edge e_i contains the vertices v_{i-1} and v_i, $1 \leq i \leq k$. In a simple graph a walk is determined by a sequence $v_0, v_1, \ldots, v_k$, of vertices; v_0 being the **initial vertex** and v_k the **final vertex**. We say a walk is **from** v_0 **to** v_k, or **connects** v_0 **to** v_k. A walk is **closed** if the initial and final vertices are identified. The number k of edges in a walk is called its **length**. A **trail** is a walk such that all of the edges are distinct. A **path** is a walk such that all of the vertices and edges are distinct. A **circuit** is a closed trail, while a **cycle** is a closed path.*

Definition 10 *A connected graph G is called **Eulerian** if there exists a closed trail containing every edge of G. Such a trail is called an **Eulerian trail**. A non-Eulerian graph G is **semi-Eulerian** if there exists a trail containing every edge of G.*

Let us now observe the theorem that solves the problem of the Königsberg bridges.

Theorem 1 (Euler, 1736) *A connected graph is Eulerian if and only if each vertex has even degree.*

For the proof see e.g. [20].

Considering now a graph given in Figure 1 in the light of the above theorem, we conclude that the closed trail that meets the required conditions does not exist.

From the proof of Theorem 1 arises,

Corollary 1 *Any connected graph with two odd vertices is semi-Eulerian. A trail may begin at either odd vertex and will end at the other odd vertex.*

3 Graph colouring

Definition 11 *Consider a graph G. A **(vertex) colouring** of G is an assignment of colours to the vertices of G such that adjacent vertices have different colours. It is a mapping $c : V(G) \to S$. The elements of S are called **colours**. If $|S| = k$, we say that c is a **k-colouring**. A colouring is **proper** if adjacent vertices have different colours. A graph is **k-colourable** if it has a proper k - colouring.*

Each graph with n vertices is n - colourable, since each vertex may be coloured with a different colour. Consequently, the question is: what is the minimum necessary number of colours to colour the graph properly.

Definition 12 *If a graph G is k-colourable, but not $(k-1)$-colourable, it is said that G is **k-chromatic**. The minimum number of colours needed to colour G is called the **chromatic number** of G and is denoted by $\chi(G)$, $\chi(G) \leq |V|$.*

3.1 Planar graphs and maps

3.1.1 About planar graphs

Although graphs are usually presented two-dimensionally, i.e. in a plane, on paper or screen, it should be noted that each graph can always be presented in three-dimensional Euclidean space without the edges being crossed. The proof of this property is simple and can be found for example in [14]. In this section we deal with requirements needed for a graph drawn in a plane to have the specified property.

Definition 13 *A graph is said to be **planar** if it can be drawn in a plane so that its edges do not cross, (Fig. 5).*

From the above definition it can be deduced that each subgraph of a planar graph is planar, and that each graph with a nonplanar subgraph is nonplanaran.

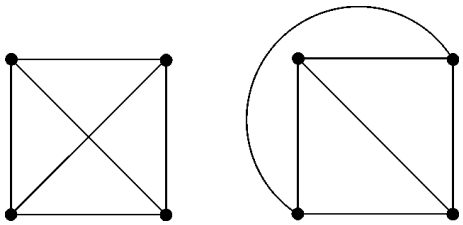


Figure 5: *The complete graph K_4 is a planar graph and a map*

Definition 14 *A **map** is a connected planar graph where all vertices have degree at least 3. A map divides the plane into a number of **regions** or **faces** (one of them infinite). The term **degree** of a region, written deg(r), refers to the length of the cycle that surrounds it. Regions are said to be **adjacent** if they share an edge, not just a point.*

Definition 15 *Graphs G and H are said to be **isomorphic** $(G \approx H)$ if there is a one-to-one correspondence between their vertices and their edges so that adjacent vertices are maped in adjacent ones.*

Definition 16 *Two graphs are said to be **homeomorphic** if they are isomorphic or one from another can be obtained by removing or inserting vertices of degree 2.*

Note that homeomorphism preserves planarity, i.e. inserting vertices of degree 2 does not have impact on planarity.

3.1.2 Some results related to planar graphs

Some Euler's results and their consequences are listed below. Here we state some of the proofs, however, the other proofs can be found in e.g. [14], [20].

1. In any map K the sum of degrees of all regions equals to twice the number of edges in K.

2. "Euler's formula": Let $G = (V,E)$ be a connected planar graph with $v = |V|$, $e = |E|$, and let r denotes the number of its regions. Then, $v - e + r = 2$.

3. If $G = (V,E)$ is a simple connected planar graph with the v vertices, $e \geq 3$ edges and r regions, then $3r \leq 2e$ and $e \leq 3v - 6$. If G does not contain triangles, i.e. the degree of each region is at least 4, then $e \leq 2v - 4$.

4. The graph K_5 is not planar.
   Proof: Indeed, in K_5 we have $v = 5$ and $e = 10$, hence $3v - 6 = 9 < e = 10$, which is in contradiction with the previous result.

5. The graph $K_{3,3}$ is not planar.
   Proof: As $K_{3,3}$ does not contain triangles, to be planar $e \leq 2v - 4$ must be fulfilled, which in this case leads to a contradiction, i.e. $9 \leq 2 \cdot 6 - 4 = 8$.

6. Every simple planar graph contains a vertex of degree lower than 6.

The following important result gives a necessary and sufficient condition for a graph to be planar.

Theorem 2 (Kazimierz Kuratowski, 1930) *A graph is planar if and only if it contains no subgraph homeomorphic to K_5 or $K_{3,3}$.*

For example, graphs given in Fig. 6 are not planar.

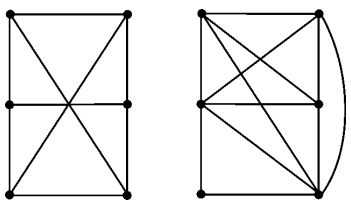


Figure 6: *Examples of non-planar graphs*

Indeed, graph (a) is graph $K_{3,3}$, (Fig. 7):

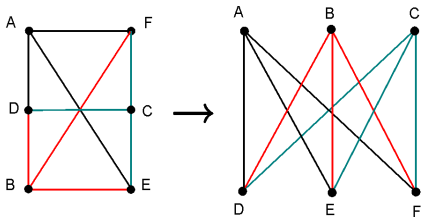


Figure 7: *Graph $K_{3,3}$*

while graph (b) is homeomorphic to graph K_5, (Fig. 8):

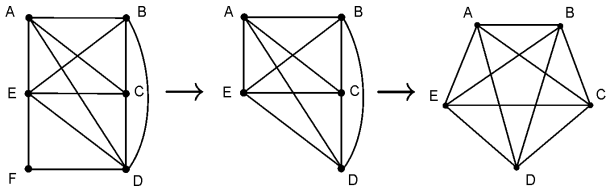


Figure 8: *A graph homeomorphic to graph K_5*

3.1.3 Every planar graph is 6-colourable

The claim will be proved by induction on the number of vertices of a graph.

For the base case we take the claim every graph with at most 6 vertices is 6-colourable. Let G be a simple planar graph with n vertices, and let all the simple planar graphs with $n-1$ vertices be 6-colourable. We consider graph G has a vertex of degree at most 5, say v.$G-v$ is a graph with $n-1$ vertices and as such is 6-colourable. As v has five neighbours, simply colouring it with the remaining colour out of 6, a proper 6-colouring of G is obtained.

3.1.4 Every planar graph is 5-colourable (Kempe, Heawood)

For the proof, we use again induction on the number of vertices of a graph.

The result holds trivially if G has one vertex. Let us assume G is a simple planar graph with n vertices, and let all the simple planar graphs with $n-1$ vertices be 5-colourable. We take in account that within G there is a vertex v of degree at most 5. $G-v$ is a graph with $n-1$ vertices and by the induction hypothesis, is 5-colourable. Now we have to assign a colour to v.

The claim of the theorem holds for $\deg(v) = 5$ since in that case it would be sufficient to colour v with the one remainin colour.

Hence, we may assume and v has five neighbours coloured differently. For being all mutually adjacent to each other would mean K_5 is a subgraph of G, being in contradiction with the assumption that G is planar. Therefore, at least one pair of vertices is not connected. Let v_1 and v_3 be the vertices in question. Contracting the edges vv_1 and vv_3, we get a graph with $n-2$ vertices being 5-colourable by the induction hypothesis. After performing the colouring, we invert the process, i.e. we stretch the contracted edges. Since v_1 and v_3 are not adjacent it causes no problem if being of the same colour. As for the neighbours of v one needs now 4 colours, we simply colour v with the one remaining colour.

3.2 Dual graph of a map

Definition 17 ***Map colouring*** *is the act of assigning different colours to different regions (faces) of a map in a way that no two adjacent regions (regions with a boundary line in common) have the same colour. We now define a map to be* ***k-colourable*** *if its faces can be coloured with k colours.*

Similarity between the above definition and the one defining graph colouring is obvious. In order to show the colouring of a map to be equivalent to the vertex colouring we need a concept of the **dual map**, also known as **geometrical dual**.

Creating a dual map for a given map K is based on a correspondence (dualism) reflected in a following way:

$$\text{region} \leftrightarrow \text{graph vertex}$$
$$\text{adjacent regions} \leftrightarrow \text{adjacent vertices}$$
$$\text{map colouring} \leftrightarrow \text{colouring graph vertices}$$

The procedure is as follows: A point within each region of a map K needs to be selected. If two regions are adjacent, points need to be connected with a curve. These curves can be drawn so that they do not intersect. The result is a new map K^*, called the **dual** of K (Fig. 9). Any colouring of the regions of a map K correspond to vertex colouring of the dual K^*.

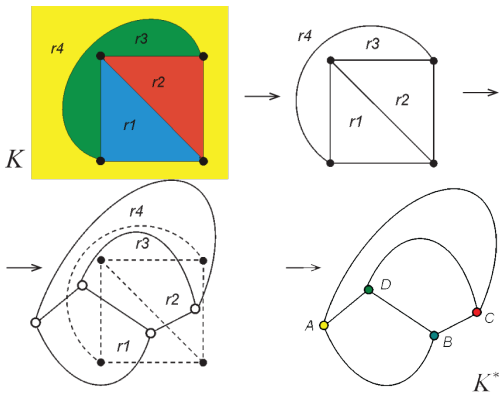


Figure 9: *From map K to its dual K^**

It is easy to see that K being planar and connected entails its geometric dual K^* to be planar and connected graph as well. Even more is fulfilled:

Theorem 3 *Let K be a planar connected graph with v vertices, e bridges and r regions, and let its dual K^* has v^* vertices, e^* bridges and r^* regions. Then, $v^* = r$, $e = e^*$, $r^* = v$.*

Proof: $v^* = r$ follows at once from the definition of a dual graph. As there is a bijection between the edges of K and the edges of K^*, we have $e = e^*$, and as K^* is planar and connected, applying Euler's formula one gets $r^* = 2 - v^* + e^* = v$ □

As mentioned before, any colouring of the regions of a map K correspond to vertex colouring of the dual K^* (Fig. 9). As a concequence, we have the following result.

Theorem 4 *A map K is region (face) k-colourable if and only if the planar graph of its geometrical dual K^* is vertex k-colourable.*

For the proof see [14].

And finally,

Theorem 5 *The four-colour theorem for maps is equivalent to the four-colour theorem for planar graphs.*

For the proof see [20].

4 The four colour theorem

The four colour problem was defined as Francis Guthrie (1831-1899), the student of the University in London in 1852 was given the task to colour the map of English counties with as few colours as possible. He concluded that 4 colours were sufficient to complete the task with the counties sharing a common border being coloured with different colours. He wanted to find out whether each map in a plane or on a sphere can be coloured with 4 colours at the most with the neighbouring countries being coloured with various colours. It implies the fact that each country presents one coherent area. This question shall initiate a great number of attempts to find the answer by mathematicians and laypersons, which shall last for more than a century making this theorem one of the issues remaining unproven for the longest period of time. The main "tool" that the mathematicians will use in solving this problem will be the Graph Theory.

4.1 Historical overview

4.1.1 Francis Guthrie first noticed the problem

Although August Möbius, a German mathematician and astronomer mentioned the four colour problem in one of his lectures held in 1840, it is considered that Francis Guhrie first posed the problem.

Francis Guthrie was a versatile person who was active in many areas. He was a very efficient barrister, acknowledged botanist (two plants were named after him: *Guthriea capensis* and *Erica Guthriei*), but first of all an excellent mathematician. However, since he could not find the solution to the four colour problem, he sent his notes with his brother Frederick to their mutual professor Augustus De Morgan. Augustus De Morgan (1806-1871) was a prominent English mathematician, a professor at the University in London who was very intrigued by this problem. Since he did not know the answer, he wrote a letter on October 23, 1852 to his colleague and friend, Sir William R. Hamilton in Dublin where he presented the statement and gave an example showing that four colours suffice. He wrote in his letter as follows, (Fig. 10):

"A student of mine asked me to give him a reason for a fact which I did not know was fact - and do not yet. He says that if a figure be anyhow divided and the compartments differently coloured so that figures with any portion of common boundary line are differently coloured – four colours may be wanted. I cannot find an example where five colours are needed. If you retort with some very simple case, I think I must do as the Sphynx did…" [23]

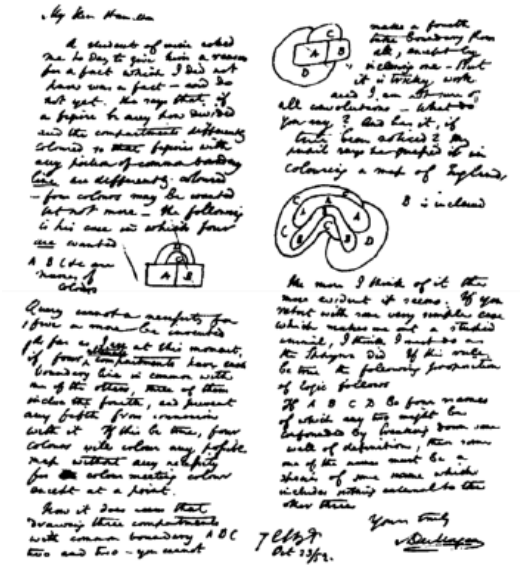


Figure 10: *Display of the original letter [5]*

Sir Hamilton, however, was not interested. Therefore, De Morgan published the problem in 1860 in a literary journal *Athenaeum*. The American mathematician, philosopher and logician Charles Sanders Pierce learned about the problem probably from the journal, and tried to solve it. Although it was said that he managed to solve it, the proof has never been published.

4.1.2 Arthur Cayley refreshes the problem

After the year 1860, in the period of about 20 years, the mathematicians almost completely stopped to be interested in the four colour problem until the British Arthur Cayley (1821 - 1895) "revived" the problem in 1878 at the meeting of the London Mathematical Society. He was namely concerned if anyone of the participants at the meeting managed to find a solution of this problem. Cayley was a mathematician and barrister, as well as a professor at the University in Cambridge. He was the youngest person that was elected a professor at the university in the 19. century. In 1879, he published an article in the journal *Proceedings of the Royal Geographical Society*. In this article, he admitted that he could not prove the statement in spite of the efforts made, but he came to some important conclusions:

– it is sufficient to observe only the maps where exactly three countries meet in each node, so called cubic maps. Namely, if more than three countries meet in some node, then a small circular "patch" is put on that node, the map thus obtained is coloured, and then the patch simply removed, (Fig. 11);

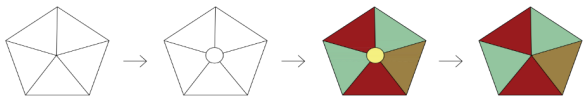


Figure 11: *Putting and removal of a "patch" when colouring a map where five countries meet in a node*

– if the four colour theorem was true, then map colouring could be performed in such a way that all countries located along the map edge are coloured with three colours at the most;

– if an arbitrary map consisting of n countries is already coloured with four colours and if we add one more country to this map, then a new map consisting of n+1 countries can be coloured with four colours.

The previous conclusion inspired Cayley to consider whether the problem could be solved by using the method of mathematical induction.

Hence, if a country is added to a map and correctly coloured, it would refer to proving the induction step: presuming that all maps with n counties can be coloured with 4 colours, it would mean also that all maps with $n+1$ counties can be coloured with 4 colours.

Consequently, the Theorem would thus be proved. However, there are too many combinations and ways in which one country can be added to some map. It is also a problem to attribute a colour to a new country. In some situations, it is trivial. However, there are cases when the colour needs to be changed for a large number of coloured countries in order to colour a new country correctly. For $n = 1, 2, 3$ and 4, the statement is trivial. Then, it could be derived from the statement for $n = 4$ that the Theorem is valid also when $n = 5$, if it is valid for all maps with 5 countries, it would be valid also for all maps with 6 countries, etc. Thus, the statement would be valid generally for all maps. It was very difficult to find a method to enlarge a map from n to $n+1$ countries that would be generally valid [7]. This is why Cayley decided to try to solve a problem by *contradiction*.

The basic idea when proving by contradiction is to assume that the statement we want to prove, say A, is false, i. e. $\neg A$ is true, and then show that this assumption leads to falsehood. Analogously, if a statement $\neg(A \Rightarrow B)$ leads to contradiction, it follows that $A \Rightarrow B$ is true.

It is first presumed that there are maps that cannot be coloured with 4 colours. A map with the smallest number of countries is selected that can be coloured with 5 or more colours. Such map is defined as the **minimal counterexample**. Then, the following statement is valid: the minimal counterexample cannot be coloured with four colours, but any map with fewer countries can be coloured with four colours. Hence, in order to prove the four colour theorem, it is necessary to prove that the **minimal counterexample does not exist**.

The next figure shows that the minimal counterexample does not contain a country that has only two neighbours. The following procedures is applied: if one edge is removed, a map with one country less is obtained that can be coloured with 4 colours at the most. The removed edge is then brought back to the map. The country with two coloured neighbours can be coloured with one out of two remaining colours (Fig. 12).

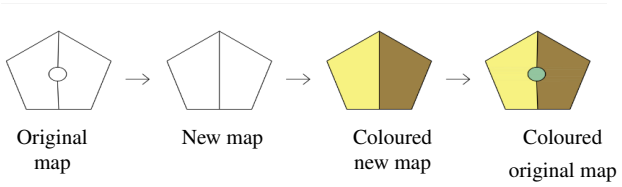


Figure 12: *A procedure applied for a country that has only two neighbours*

Similar proof procedure will also be applied with the country having three neighbours. One edge is removed and 3 countries are obtained out of 4 countries. Such map can

be coloured with 4 colours. Three countries are coloured with three various colours, the removed edge is brought back, and the fourth country is coloured with the remaining fourth colour, (Fig. 13).

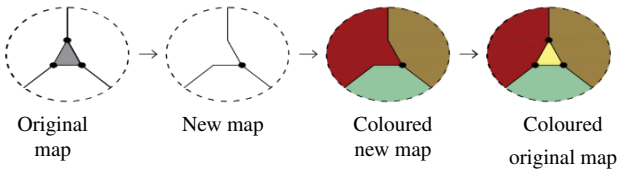


Figure 13: *Colouring countries having three neighbours*

However, there is a problem when applying the method of removing and restoring to the countries with 4, 5 or more neighbors, (Fig. 14 and Fig. 15).

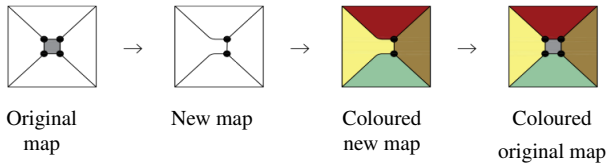


Figure 14: *The case when a map contains a square*

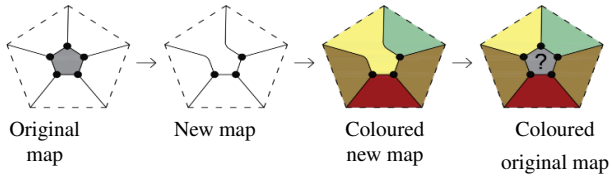


Figure 15: *The case when a map contains a pentagon*

4.1.3 Maps and Euler polyhedra

Leonhard Euler played an important role in proving the four colour conjecture with his findings and research that will later on be used by mathematicians. Dealing with regular polyhedra, he made an important discovery, namely a formula ("Euler formula") that states that:

number of faces − number of edges + number of verteces = 2.

The formula has many applications and can be generalized in various ways with one of them being used to handle planar graphs, i.e. maps (Chapter 3.1.2).

The connection between a map and a polyhedron is achieved by projecting a polyhedron from one point to a plane (Fig. 16). The faces of polyhedron in a plane projection represent countries/regions where one face is observed as the exterior of the projection, and the edges are actually boundary lines of countries. On the other hand, every cubic map, i.e. the map on which exactly three countries meet in each vertex, can be drawn onto a sphere and then, it can be imagined that it presents some polyhedron. In this case, the problem of colouring a spherical map is identical to the problem of colouring a map in the plane [22].

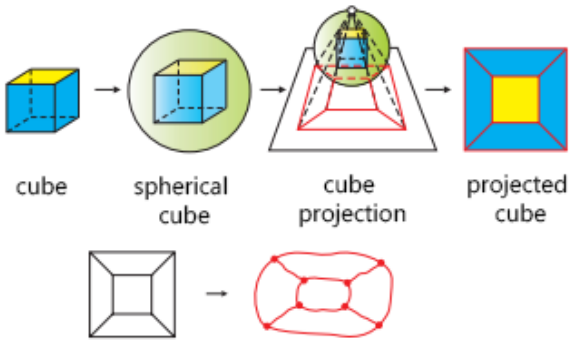


Figure 16: *Projection of a polyhedron from a point into the plane [22]*

A direct consequence of Euler's formula is the so called *enumeration* formula [7]. Using this formula one can count the regions, edges and vertices of a map that has r_2 regions with exactly two neighbors, r_3 regions with exactly 3 neighbors, r_4 regions with exactly 4 neighbors, etc.

Euler used the enumeration formula for proving the"only 5 neighbors" theorem, i.e. that *every cubic map has at least one region with five or fewer neighbors*. In addition, *if a map does not contain any biangle or a triangle and not a single square, it must contain at least 12 pentagons*. Similarly, it can be concluded the following: *if a cubic map consists entirely of pentagons and hexagons, then it must have exactly 12 pentagons* (for the proofs see [23]). Although Arthur Cayley had failed to prove that the minimum counterexamples do not exsist, his idea proved useful because it was used to prove somewhat weaker claim, the six – colour theorem.

4.2 Kempe "solves" the problem

Sir Alfred Bray Kempe (1849-1922) was also a barrister and mathematician. He finished his studies at the Trinity College, Cambridge where he attended the lectures of Arthur Cayley. He was also present at the meeting of the London Mathematical Society where Arthur Cayley spoke about the problem. He succeeded to apprehend the issues of 4 colour theorem, and a year later he published an article in the *American Journal of Mathematics* where he claimed to have managed to solve the problem successfully. The procedure of Kempe's method of colouring any map can be presented in all of the following six steps:

1. Find a country on the map that has 5 or less neighbors (it exists according to Theorem of "only 5 neighbors");

2. Cover the country in question with a piece of blank paper of similar shape, just a little bigger;

3. Extend all borders that touch the "patch" so that they meet at one point on the paper - as if the selected country has been reduced to one point (with this procedure the number of the countries on the map is reduced by 1), (Fig. 17);

Figure 17: *Reducing the number of countries on a map*

4. Repeat the three previous steps until the initial map is reduced to a map with exactly one country;

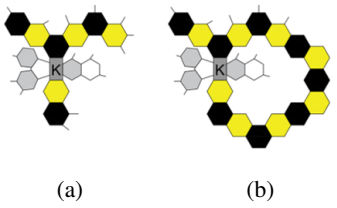
5. Colour the only remaining country with any of the four given colors;

6. Reverse the upper process: remove "patches" all the way back until you get the initial map and colour every "restored" country with different color from the neighbor along the way [8].

Now we face the problem that Cayley couldn't solve, i.e. how to color the country which has 4 or 5 neighbors. Kempe has solved this problem by using the method of chains.

4.2.1 Method of Kempe chain

Kempe assumed that the country K that needs to be coloured has a square form, i.e. borders with 4 countries. He then selected two countries that do not share borders with each other. Fig. 18 presents the country K and two not neighbouring countries that share their border with the country K. They are coloured in black and yellow colour. Now, on both of them, we continue with a line of black-yellow coloured countries. These lines can be connected in such a way that they make a closed circle that is then called a *chain*. Two cases may occur when coloring a map by means of this method, both shown in Figure 18.



(a) (b)

Figure 18: *Two possible outcomes when using the Kempe chain method*

Fig. 18 a) presents the first case of colouring the country K. It can be seen on the figure that the black neighbour of K is not connected with the yellow neighbour of K. Then it is possible to re-colur the black neighbour of the country K, e.g. with yellow colour. Black colour remains then available for the country K so that the map can be in accordance with the theorem. This procedure is shown on Fig. 19.

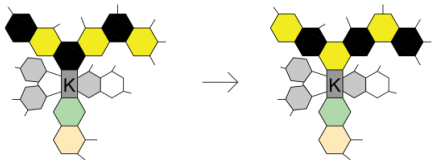


Figure 19: *Colour replacement in the line*

In the case presented on Fig. 18 b), the previous procedure of colour replacement shall not be successful. However, the chain of black and yellow countries makes a loop that starts and ends in the country K. Two other neighbours can be seen on Fig. 20: blue and yellow. It is not possible to join the chains of these two neighbours because they are interrupted by a black-yellow loop. The method is therefore applied similarly as in the first case: if a blue neighbour changes the colour into yellow, and the colours of the entire blue-yellow branch are replaced, the country K can be coloured with the remaining blue colour.

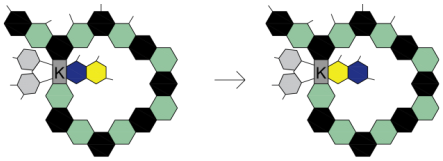


Figure 20: *Replacement of yellow and blue*

It is herewith proved that **no minimal counterexample contains "a square"**. It is namely sufficient to have 4 colours for the "square". It would be necessary to prove furthermore that the minimal counterexample does not contain a pentagon. The pentagon is surrounds by 5 countries that are already coloured with 4 colours. Kempe

solved this problem by selecting two neighbours of P that do not touch each other: in this case, the yellow and red neighbour "above" and "below" P as presented on Fig. 21.

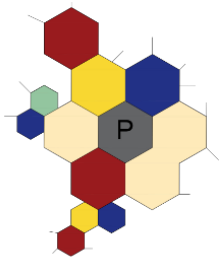


Figure 21: *Pentagon P and its neighbours*

If the above yellow-red line is not connected to the red-yellow line below, then the colours of the neighbours of the country P can be replaced, hence the yellow neighbour of P becomes red. Thus, yellow colour is left as a possible colour for P, as it is presented on Fig. 22.

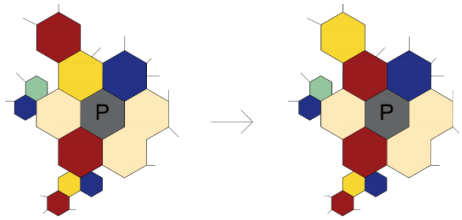


Figure 22: *Yellow neighbour of P becomes red*

If the "above" yellow-red line on Fig. 21 is connected with the red-yellow line "below" the country P, then the blue neighbour of the country P can be observed, as well as red-blue and blue-red lines. Such example is presented on Fig. 23.

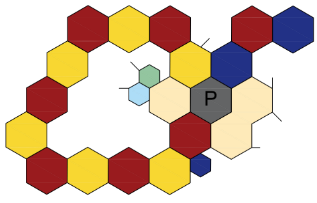


Figure 23: *The red-yellow chain*

Consider the situation given on Fig. 22. If the blue-red line "above" was not connected with the red-blue line "below", the blue neighbour of the country P can be coloured with red colour, and all countries in the blue-red line can be re-coloured. In this way, another red-blue line is obtained. Thus, only blue colour is left for the country P as presented on Fig. 24.

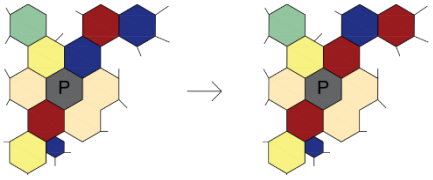


Figure 24: *Blue neighbour of P becomes red*

However, if the chains are linked, then there are two loops together with the previous one. Such situation is presented on the next figure.

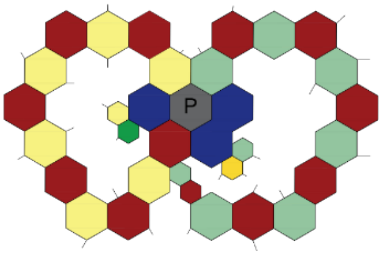


Figure 25: *The blue-yellow and the blue-green chain*

It can be seen on Fig. 25 that blue-yellow line on the left side of the country P cannot be connected with the blue-yellow line on the right side of the country P. The colours of the blue-yellow line on the right side can then be replaced. The blue-green line on the left side cannot be connected with the blue-green line on the right side, hence, the blue-green line on the left side can be re-coloured. If the lines are re-coloured simultaneously, the country P will have the neighbours in yellow, red and green colour, and it can be coloured with blue colour, as presented on Fig. 26.

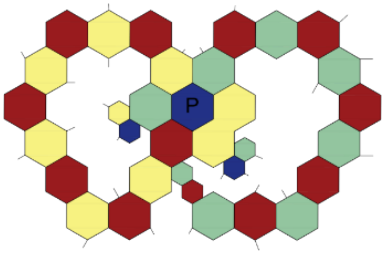


Figure 26: *P is coloured blue*

Hereby, the procedure of colouring the map to which the country *pentagon* has been brought back is completed. Using the above described procedure of colouring the map that has a country with five neighbours, Kempe found a proof that the minimal counterexample **does not contain a pentagon**. It is, however, in contradiction with the "five neighbours" theorem according to which every cubic map contains at least one country with five or less neighbours.

According to Kempe, the four colour theorem would thus be proved. After publishing the article, Kempe was recognized as the person who proved the theorem. His article was published in the *American Journal of Mathematics*.

4.2.2 The flaw in Kempe's proof

It took 11 years to spot the mistake in Kempe's proof. In 1890, Percy John Heawood (1861 - 1955), a professor of mathematics from Durham denied Kempe's theory. After that, Kempe's proof became the most famous inaccurate proof in the history of mathematics.

In his article *Map-colour Theorem*, published in the *Quarterly Journal of Mathematics* in 1890, Heawood explained that the error occurred at the end of the proof in determining the colour of the country with pentagon form. Heawood proved that *there were maps on which it is impossible to make changes of colours in two different chains simultaneously*.

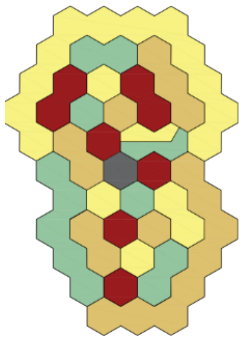


Figure 27: *Map with 25 countries*

On Fig. 27, each of the twenty five countries is coloured with one of the four colours: red, golden, yellow or green, except the central pentagon *P*. It was proved that this map can be coloured only with four colours.

The application of Kempe's methods in determining the pentagon *P* provides the re-colouring of two neighbours of the pentagon *P*.

Each of these two changes is allowed if it is done separately. The problem occurs if it is attempted to make the changes simultaneously.

Two neighbouring countries marked with the letter *A* (yellow colour) and *B* (golden colour) become red as it is presented on Fig. 28. The basic principle of 4 colour problem implying that the neighbouring countries should be coloured with various colours is hereby undermined. Hence, one came to the conclusion that Kempe's method of proof was wrong.

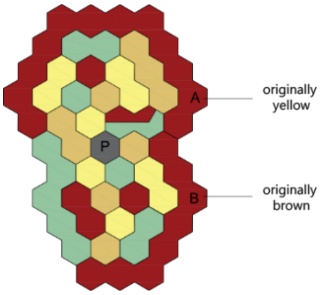


Figure 28: *Contradiction*

In 1891, Kempe admitted publicly that he was wrong. However, he managed to notice the mistake in the proof Heawood did not know how to correct it. In his second paper, Heawood approached the problem using a number theory, but even this attempt to prove the four colour theorem ended unsuccessfully. Using Kempe's ideas, Heawood was able to prove the Five colour theorem. Although the Five colour theorem was weaker than the Four colour theorem, it still represents one more step that will be needed to prove the initial problem.

After his proof had been denied, Kempe approached the problem in somewhat different way; in each country on a map, he highlighted one point (e.g. capital city) and then connected the points representing the neighbouring countries with lines. The new structure matched the structure of a graph. The problem of determining the colours of individual countries was reduced to assigning it to points, but in such a way that the neighbouring points were named differently. The importance of this idea is related to the fact that in such a way the problem of map colouring was transferred into graph theory (Fig. 9). Based on this idea and with the help of computers, the Four colour theorem will finally be proved.

4.3 Heesch, Appel and Haken finally solve the problem

The conjecture on four colours is articulated so simply that it was presumed someone would find an elegant and simple solution one day. However, something completely different happened.

In 1904, a new idea about the proof of this conjecture occurred. This approach started with the search for *unavoidable sets*. Before defining this term, it is necessary to define the terms triangulation and configuration. A plane graph is a *triangulation* if it is connected and every region is a triangle. A *configuration* is a part of triangulation included inside the area (map). The unavoidable set is then defined as a set of configurations with the property that any triangulation must contain one of the configurations in the set. Unavoidable set is actually a set of countries out of

which at least one country must be located on every map. Since it is sufficient to observe only cubic maps, Fig. 29 presents some countries that a cubic map must contain.

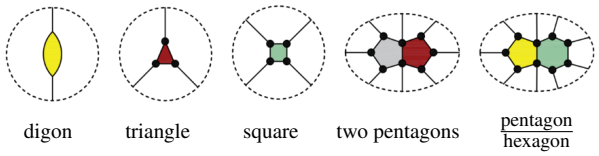


Figure 29: *Unavoidable set for a cubic map*

It can be deduced from the figure that if a map does not contain digon (a country with two neighbours), triangle or rectangle, it must contain either two connected pentagons or connected pentagon and hexagon.

Among the mathematicians who started to search for unavoidable sets, George David Birkhoff (1884 - 1944) made a distinguished contribution in this area. Birkhoff was the first to introduce the concept of *reducibility*. A configuration is reducible if it cannot be contained in a triangulation of the smallest graph which cannot be 4-coloured. It means that reducible configuration is a set of countries that cannot appear in the minimal counterexample. Minimal counterexample is, as it has been defined earlier, a map that contains the smallest number of countries and can be coloured with 5 or more colours. From this follows that the minimal counterexample cannot be coloured with 4 colours, but every map with smaller number of countries can. Hence, in order to prove the Four colour theorem, it is necessary to prove that the minimal counterexample does not exist.

The research and search for unavoidable sets and reducible configurations were developing separately until German mathematician Heinrich Heesch (1906 - 1995) unified those 1960. His goal was namely to find an *unavoidable set of reducible configurations*. If a set is unavoidable, then each map must contain at least one of the configurations from that set, and since every configuration is reducible, it cannot be contained in a minimal counterexample. It would thus be proved that there are no minimal counterexamples, and consequently, the 4 colour theorem would be proved. He therefore developed an algorithm naming it *D-Reduction* that he adapted to computer methods (programming) [24]. This algorithm is used to prove that every graph contains a subgraph from a specific set, i.e. that every map contains some map from an unavoidable set.

Heesch presumed that he would have needed to observe a set of about 8900 configuration. However, certain problems appeared in his approach, as for example the inability to test the reductions of some configurations, mostly because of a large number of vertices within some rings, i. e. configurations that "wrap arround and meet themselves" [17].

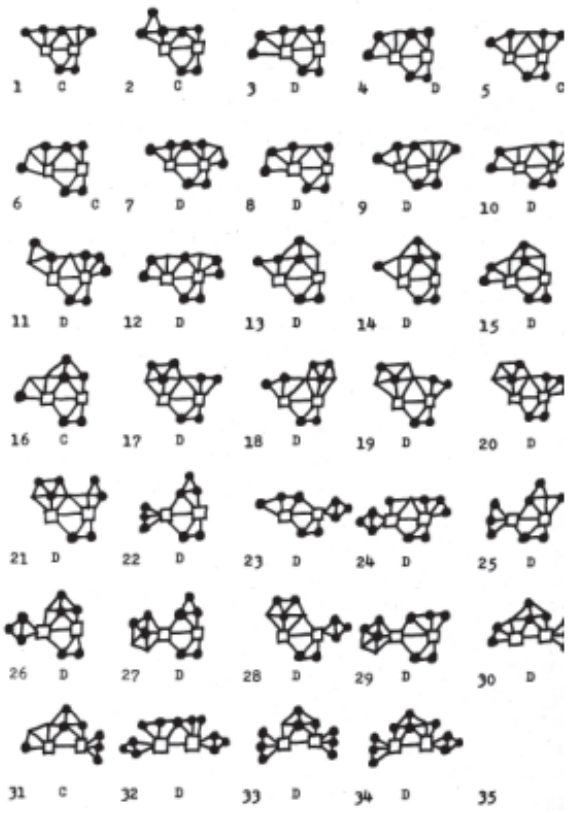


Figure 30: *A set of configurations*

In 1972, Wolfgang Haken, a student of mathematics, physics and philosophy continued after a short collaboration with Heesch to work with a programmer and mathematician Kenneth Appel on upgrading of Heesch's idea. They were focused on the improvement of Heesch algorithm. After two years, John Koch joined them, and the three of them succeeded together to create the programme to be used in searching for unavoidable sets of reducible configurations. Unlike Heesch, they manage to reduce the number of ring vertices from 18 to 14 avoiding thus the complications and simplifying the counting.

Using the programme for searching an unavoidable set of reducible configurations, Appel and Haken, both from the University of Illinois, managed to prove the assumption in 1976. Since there are too many possible configurations, the proof could not be carried out without computer assistance. The usage of computer in proving this problem caused numerous discussions and disapprovals.

Still, Haken and Appel published the proof on July 22, 1976 that was based on the construction of the unavoidable set of 1936 reducible configurations, and in 1977 all three of them published the proof in *Illinois Journal of Mathematics* with the unavoidable set of 1482 reducible configurations. The proof was published in two parts, and the

text was accompanied by the material on microfilm with 450 pages of various diagrams and detailed explanations. However, Ulrich Schmidt found an error in the programme in 1981 that was soon corrected.

Regardless of the difficulties, the four colour problem gained the status of a theorem for the second time. The following comment illustrates the opinion of a great number of mathematicians at that time:

"Good mathematical proof is like a poem - this is a telephone directory"

Appel and Haken published therefore in 1986 an article where they described their methods in details strongly defending the proof and rejecting any doubt, and three years later they also published a book titled *Every Planar Map is Four Colourable*.

Due to the complicated nature of part of the proof that can't be checked without computer assistance, Paul Seymour, Neil Robertson, Daniel Sanders and Robin Thomas decided to simplify the proof and to eliminate all doubts. However, they gave up soon after they had started to study it. They decided to develop their own proof based on the ideas of Appel and Hesch.

As they [17], wrote in the article, the concept of the proof itself is identical to the concept of Appel and Haken. They tested the set of 633 configurations proving that each of them is reducible. Furthermore, they proved that at least one of the 633 configurations appears in a planar graph with 6 vertices (minimal counterexsample). They have thus proved the unavoidability. This part shows the largest difference between their proof and the one made by Appel and Haken. In order to prove the unavoidability, they used the method of discharging that unlike with Appel and Haken has 32 rules as related to 300+.

The article itself and their proof were presented at the International Congress of Mathematicians in Zürich in 1994 where they finally proved that Appel and Haken were right. The proof itself was also upgraded and improved and contained 633 configurations instead of 1482.

Fig. 31 presents 17 out of 633 configurations that were used in this proof. When drawing the configurations, they used Heesch's method of marking. The forms of vertices present the degree of a vertex. Black circle represents the vertex of degree 5, the point (shown on the figure without a symbol) has the vertex degree 6, empty circle represents the vertex of the degree 7, the triangle the vertex of degree 9, and the pentagon represents the vertex of the degree 10.

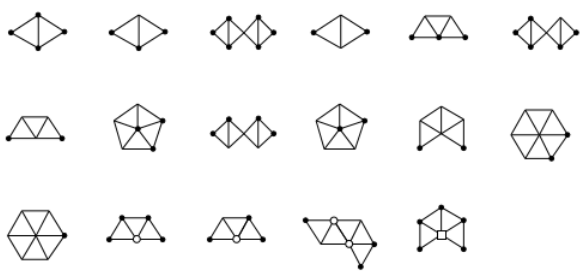


Figure 31: *17 out of 633 configurations [17]*

5 Application within cartography

Each map in the plane can be coloured with four colours such that neighbour areas are in different colours, as it is shown in previous chapters. Application of four colour theorem for colouring political maps is tested in this chapter. We use a world political map as an example. The software used is QGIS [25] with its unofficial plugin TopoColour [26], which implements algorithms for graph colouring.

Geographical maps of administrative units, i.e. political maps, have some specialities that should be considered prior to the application of graph colouring algorithms.

As it is defined in 3.1.1, neighbouring countries (administrative units) are those which shares common boundary line. Existence of common point does not imply neighbours. Geographical maps are abstract and generalised models of reality, and it is possible that some short boundary line in reality is represented as point, due to model or cartographic generalisation. Application is therefore possible only to model of geographical reality, users should be aware of these constraints, and in case of unexpected results, know how to deal with it.

Maritime boundaries are often not shown on political maps, and almost never are administrative areas on the sea coloured with different colours. This is certainly true for world political maps, where colouring is usually applied to land parts of countries. This means that countries that share only maritime boundaries will not be considered as neighbours and could be coloured automatically with the same colour.

Further, countries are often consisted of more or less distant land parts, e.g. islands or exclaves. For example, some countries at certain administrative level contains overseas territories. This could potentially lead to non-planar graphs representing neighbours.

5.1 Methodology and programs used

The data used for the world countries were taken from GADM database of Global Administrative Areas [27].

There is no unique solution for model of countries and its boundaries, and it is often result of the point of view of certain diplomacy. For the purpose of this paper, we will not change the original data, because we will use it for the purpose of demonstration of application of graph colouring, and not for making special world political map. The dataset consists of 256 top-level administrative units, i.e. countries.

Data was first loaded in QGIS as separate vector layer and transformed to Eckert VI map projection, one of map projections suitable for world maps.

TopoColour plugin implements algorithms from graph colouring theory with purpose of colouring polygons in vector layer. It also allows creating graph representing detected neighbours in dataset.

Typical procedure for colouring areas in vector layer is as follows:

– Start the TopoColour plugin and select polygon vector layer and one column in attribute table that has unique value for each administrative unit. Finding of neighbours starts. It can take a while for complex geometries (e.g. up to one hour or more).
– When neighbours are found, user selects "greedy" or "random" algorithm and starts the computing of colours. Number of colours needed is given as a result. For "random" algorithms, successive computations can give different number of colours.
– Save the colour numbers to one column in attribute table and style the layer.

Greedy algorithm gives five colours for political map of the world. It does not guarantee optimum number of colours. Brute force approach for four colours and 256 countries would yield 4^{256} different colour assignments, and it is not feasible even with modern computers. Random algorithm usually gives six or seven colours for this political map.

After computation of colours is done, layer can be styled in order to get nice coloured map (Fig. 32).

It is known that world political map can be coloured with four colours. In order to achieve this we start with automatically defined colours, eliminate less used colour by replacing it with one of the four colours, and rearranging the colours of neighbours. It is not too hard to accomplish that, and result can be seen on Fig. 33. This also means that graph representing neighbouring countries is planar (Fig. 34) and algoritms used are not giving optimal solution.

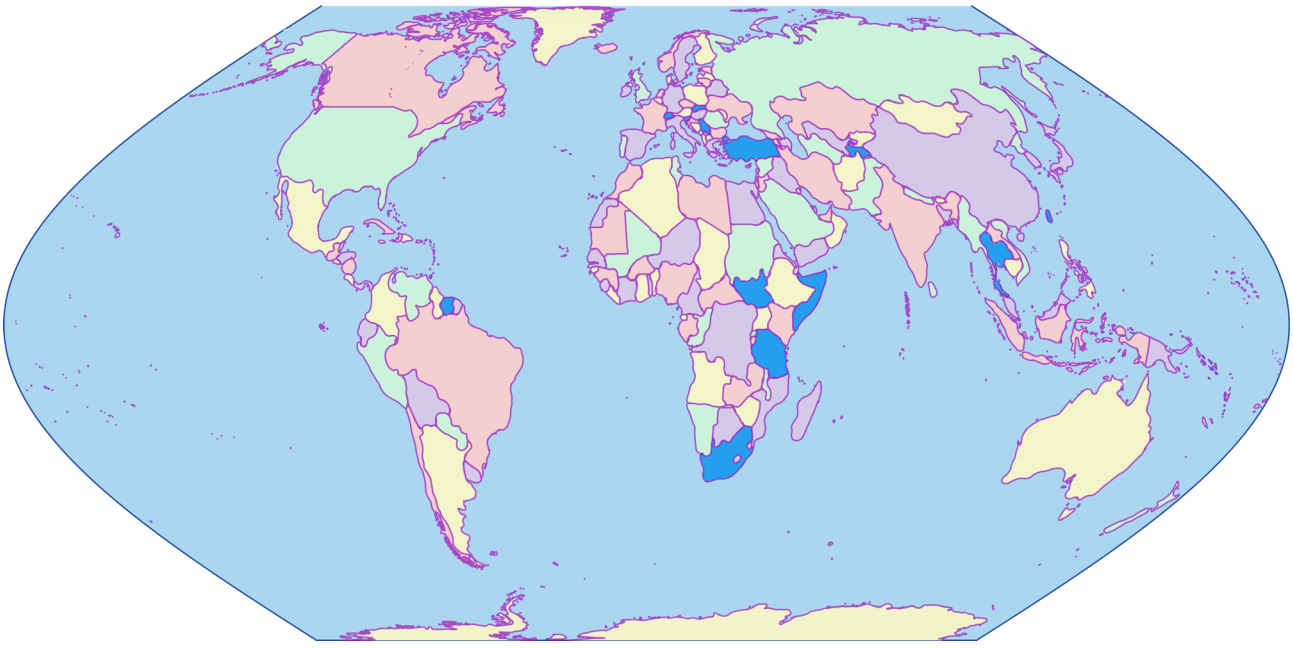


Figure 32: *World political map coloured with five colours obtained by greedy algorithm implemented in QGIS plugin TopoColour. Less used colour is blue, and is a good candidate for manual elimination.*

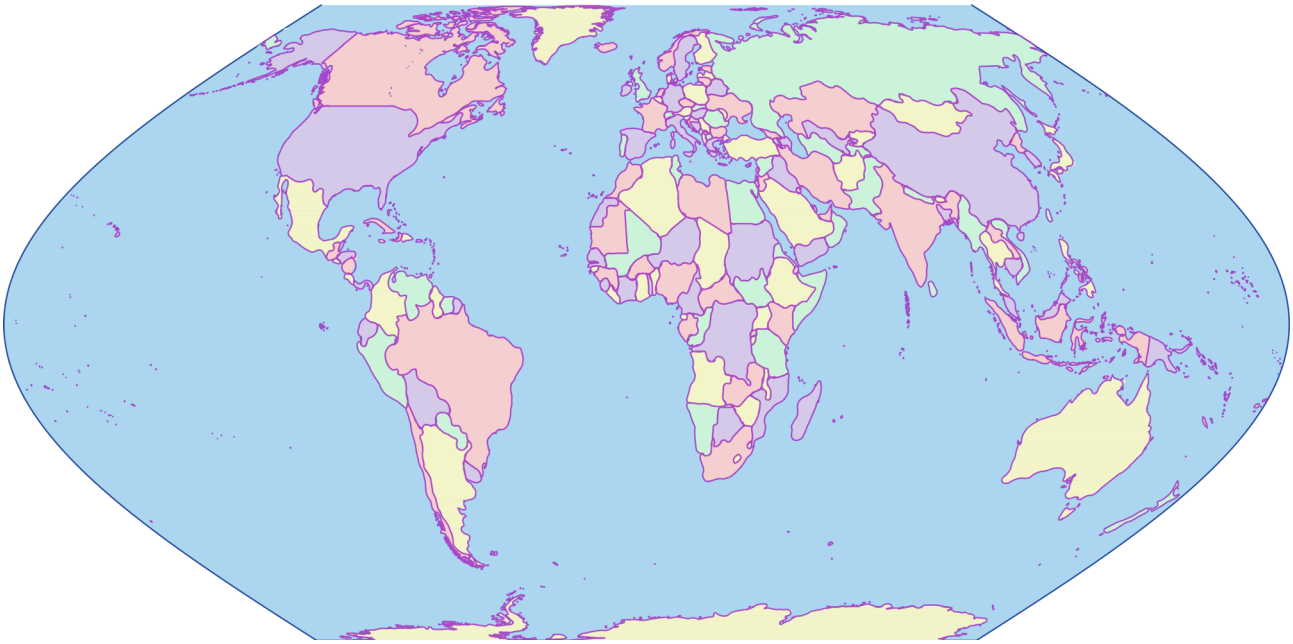


Figure 33: *Manual elimination of the fifth colour from automatically coloured world political map gives a map with four colours.*

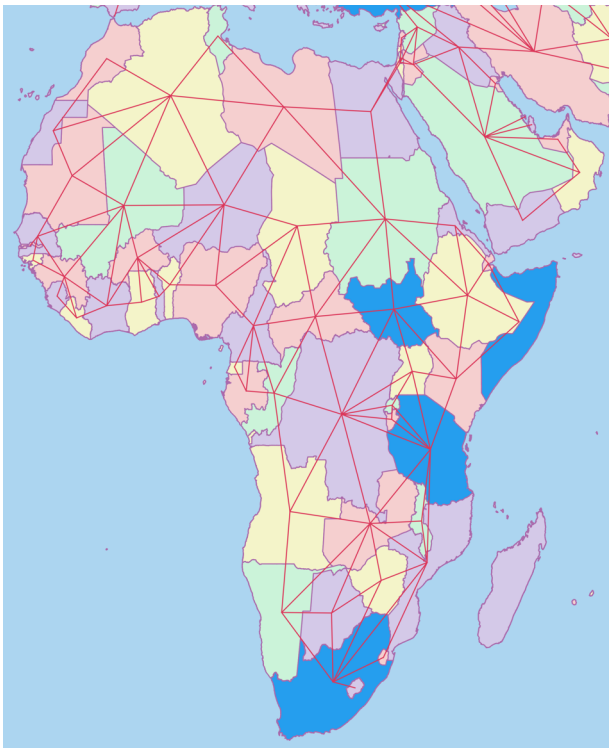


Figure 34: *Graph representing neighbouring countries (clipped to Africa region)*

6 Conclusion

Graph colouring is widely applied in many scientific fields. In this paper, the focus is on the application in cartography. Since the map colouring with four colours is rarely mentioned in cartographic books, we were motivated to research this connection in this paper.

The four colours conjecture has proved to be one of the greatest and long lasting problems in mathematics. The problem itself has attracted the attention both of mathematicians and laypersons. It took more than one century to prove this conjecture, which was achieved only with the development of information science and with computer assistance. It is also the first more significant theorem that has been proved in such a way. The theorem faced a lot of negative comments because of that and was not well accepted by the mathematical public of that time.

Algorithms implementing graph colouring and four colour theorem, which are still not so widely available in cartographic software, provide analysis and processing of map data with aim of colouring administrative units or creating political maps. In this process one should take care of geometry of boundaries because even very small differences in coordinates can give unexpected results. Special care has to be given to model of geographical reality, e.g. definition of administrative entities, maritime boundaries between countries, overseas territories etc. We can conclude that automatization of colouring of administrative units can

greatly help mapmakers, but for the final map, manual interventions are still required.

References

[1] J. BEBAN BRKIĆ, *Diskretna matematika*, internal script, Faculty of Geodesy University of Zagreb, Zagreb, 2012.

[2] N. FRANČULA, *Digitalna kartografija*, script, Faculty of Geodesy University of Zagreb, Zagreb, 2004.

[3] N. FRANČULA, *Kartografske projekcije*, script, Faculty of Geodesy University of Zagreb, Zagreb, 2004.

[4] S. FRANGEŠ, *Kartografija*, lectures from the Faculty of Civil Engineering, Architecture, Geodesy and Geoinformatics University of Split, Split, 2012.

[5] R. FRITSCH, G. FRITSCH, *The Four-Color Theorem*, Springer, Germany, 1998.

[6] A. GOLEMAC, A. MIMICA, T. VUČIČIĆ, Od koenigsberških mostova do kineskog poštara, *Hrvatski matematički elektronički časopis*, 2012.

[7] S. GRAČAN, Četiri su dovoljne!, *Miš* **243** (2007), 26–35.

[8] G. GRDIĆ, *Teorem o četiri boje*, bachelor's thesis, Department of Mathematics University of Rijeka, Rijeka, 2015.

[9] G. HAKE, D. GRUNREICH, *Kartographie*, Walter de Gruyter, Berlin, New York, 1994.

[10] T. JOGUN, *Izrada političke karte svijeta iz podataka Openstreetmapa*, master's thesis, Faculty of Geodesy University of Zagreb, Zagreb, 2016.

[11] S. LIPSCHUTZ, M. LIPSON, *Discrete Mathematics (3rd Edition)* , Schaum's Outline Series, McGraw-Hill, New York, 1997.

[12] M. MONMONIER, *How to Lie with Maps (2nd Edition)*, University of Chicago Press, Chicago, London, 1996.

[13] A. NAKIĆ, M.-O. PAVČEVIĆ, *Bojanja grafova*, script, Faculty of Electrical Engineering and Computing University of Zagreb, Zagreb, 2014.

[14] A. NAKIĆ, M.-O. PAVČEVIĆ, *Planarnost*, script, Faculty of Electrical Engineering and Computing University of Zagreb, Zagreb, 2014.

[15] J. NEUMANN, *Enzyklopädisches Wörterbuch Kartographie in 25 Sprachen*, 2. erweiterte Ausgabe, Saur, München, 462, 1997.

[16] M.-O. PAVČEVIĆ, *Uvod u teoriju grafova*, Element, Zagreb, 2007.

[17] N. ROBERTSON, D. SANDERS, R. THOMAS, P. SEYMOUR, A New Proof of the Four-colour Theorem, *Electronic Research Announcements of the American Mathematical Society*, Volume2, No 1, (1996), 17–25.

[18] K. A. SALIŠČEV, *Kartografija: ubenik dlja universitetov*, M.: "Vicaja kola", Moskva, 1982.

[19] H. SREBRO, Political Considerations in Cartographic Maps, *Proceedings of the 26th International Cartographic Conference*, August 25-30, Dresden, Germany, 2013.

[20] R. J. WILSON, *Introduction to Graph Theory (4th Edition)*, Prentice Hill, Addison Wesley Longman Limited, Edinburgh Gate, Harlow, Essex CM20 2JE, England, 1998.

[21] `http://www-history.mcs.st-andrews.ac.uk/HistTopics/The_four_colour_theorem.htm` (15.10.2016.)

[22] `http://www.ics.uci.edu/~eppstein/junkyard/euler/` (29.10.2016.)

[23] `http://www.math.ualberta.ca/~isaac/math222/f05/euler.pdf` (23.9.2016.)

[24] `http://www.halapa.com/pravipdf/boje4.pdf` (15.05.2016.)

[25] `http://www.qgis.org/` (20.05.2016.)

[26] `http://github.com/nyalldawson/topocolour` (20.05.2016.)

[27] `http://www.gadm.org` (20.05.2016.)

Mate Glaurdić
e-mail: mate.glaurdic@gmail.com

Jelena Beban-Brkić
email: jbeban@geof.hr

Dražen Tutić
email: dtutic@geof.hr

Faculty of Geodesy, University of Zagreb,
Kačićeva 26, 10 000 Zagreb, Croatia

How to get KoG?

The easiest way to get your copy of KoG is by contacting the editor's office:

Marija Šimić Horvath
msimic@arhitekt.hr
Faculty of Architecture
Kačićeva 26, 10 000 Zagreb, Croatia
Tel: (+385 1) 4639 176
Fax: (+385 1) 4639 465

The price of the issue is €15 + mailing expenses €5 for European countries and €10 for other parts of the world.

The amount is payable to:

ACCOUNT NAME:	Hrvatsko društvo za geometriju i grafiku Kačićeva 26, 10000 Zagreb, Croatia
IBAN:	HR8623600001101517436

Kako nabaviti KoG?

KoG je najbolje nabaviti u uredništvu časopisa:

Marija Šimić Horvath
msimic@arhitekt.hr
Arhitektonski fakultet
Kačićeva 26, 10 000 Zagreb
Tel: (01) 4639 176
Fax: (01) 4639 465

Za Hrvatsku je cijena primjerka 100 KN + 10 KN za poštarinu.

Nakon uplate za:

HDGG (za KoG), Kačićeva 26, 10000 Zagreb
IBAN: HR8623600001101517436

poslat ćemo časopis na Vašu adresu.

Ako Vas zanima tematika časopisa i rad našeg društva, preporučamo Vam da postanete članom HDGG-a (godišnja članarina iznosi 150 KN). Za članove društva časopis je besplatan.

INSTRUCTIONS FOR AUTHORS

SCOPE. "KoG" publishes scientific and professional papers from the fields of geometry, applied geometry and computer graphics.

SUBMISSION. Scientific papers submitted to this journal should be written in English, professional papers should be written in Croatian or English. The papers have not been published or submitted for publication elsewhere.
The manuscript should be sent in PDF format via e-mail to one of the editors:

Sonja Gorjanc
sgorjanc@grad.hr

Ema Jurkin
ema.jurkin@rgn.hr

The first page should contain the article title, author and coauthor names, affiliation, a short abstract in English, a list of keywords and the Mathematical subject classification.

UPON ACCEPTANCE. After the manuscript has been accepted for publication authors are requested to send its LaTex file via e-mail to one of the addresses:

sgorjanc@grad.hr, ema.jurkin@rgn.hr

Figures should be titled by the figure number that match to the figure number in the text of the paper.

The corresponding author and coauthors will receive hard copies of the issue free of charge.

UPUTE ZA AUTORE

PODRUČJE. "KoG" objavljuje znanstvene i stručne radove iz područja geometrije, primijenjene geometrije i računalne grafike.

UPUTSTVA ZA PREDAJU RADA. Znanstveni radovi trebaju biti napisani na engleskom jeziku, a stručni na hrvatskom ili engleskom. Rad ne smije biti objavljen niti predan na recenziju u drugim časopisima.
Rukopis se šalje u PDF formatu elektronskom poštom na adresu jedne od urednica:

Sonja Gorjanc
sgorjanc@grad.hr

Ema Jurkin
ema.jurkin@rgn.hr

Prva stranica treba sadržavati naslov rada, imena autora i koautora, podatke o autoru i koautorima, sažetak na hrvatskom i engleskom, ključne riječi i MSC broj.

PO PRIHVAĆANJU RADA. Tekst prihvaćenog rada autor dostavlja elektronskom poštom kao LaTeX datoteku. Slike trebaju imati nazive koji odgovaraju rednom broju slike u tekstu članka. Adresa:

sgorjanc@grad.hr, ema.jurkin@rgn.hr

Svaki autor i koautor dobiva po jedan primjerak časopisa.