Irish Mathematical Society
Cumann Matamaitice na hÉireann

Bulletin

Number 91 **Summer 2023**

ISSN 0791-5578

Irish Mathematical Society Bulletin

The aim of the *Bulletin* is to inform Society members, and the mathematical community at large, about the activities of the Society and about items of general mathematical interest. It appears twice each year. The *Bulletin* is published online free of charge.

The *Bulletin* seeks articles written in an expository style and likely to be of interest to the members of the Society and the wider mathematical community. We encourage informative surveys, biographical and historical articles, short research articles, classroom notes, book reviews and letters. All areas of mathematics will be considered, pure and applied, old and new.

Correspondence concerning books for review in the *Bulletin* should be directed to

`mailto://reviews.ims@gmail.com`

All other correspondence concerning the *Bulletin* should be sent, in the first instance, by e-mail to the Editor at

`mailto://ims.bulletin@gmail.com`

and only if not possible in electronic form to the address

The Editor
Irish Mathematical Society Bulletin
Department of Mathematics and Statistics
Maynooth University
Co. Kildare W23 HW31

Submission instructions for authors, back issues of the *Bulletin,* and further information about the Irish Mathematical Society are available on the IMS website

`http://www.irishmathsoc.org/`

The Irish Mathematical Society is a registered charity
(RCN: 20020279).

Irish Mathematical Society Bulletin 91, Summer 2023

CONTENTS

EDITORIAL

In this issue we print a letter from Tom Carroll, the previous President of the IMS, which informs members about the recent action of the Society, taken to ensure this country's continued involvement with the International Mathematical Union.

All who knew him will miss Rex Dark, a true stalwart, whose obituary appears in this issue. A short article by Des MacHale which poses some questions about the possible orders of a group with trivial centre, incorporates the fruit of a conversation with Rex.

The review of Des MacHale's *Comic Sections Plus* marks the first instance in the Bulletin of a mother-daughter team of authors. One can only hope that this marks the dawn of a new age of work-life balance among practioners of our ancient art.

The Annual Scientific Meeting will be held on Thursday 31st August and Friday 1st September 2023 at the University of Limerick. The meeting is organised jointly by members of the Department of Mathematics and Statistics at University of Limerick and the Department of Mathematics and Computer Studies at Mary Immaculate College. As we go to press, confirmed speakers include Norma Bargary (UL) Patrick Browne (TUS) John Butler (TUD) Julie Crowley (MTU) James Cruickshank (Galway) Patrick Farrell (Oxford) Emma Greenbank (UL) Thomas Heuttemann (QUB) Cónall Kelly (UCC) Bernd Kreussler (MIC) Myrto Manolaki (UCD) and Katrin Wendland (TCD).

Organisers of all meetings supported by the Society should submit their reports to the editor by the 15th of December, for publication in the Winter issue. The LaTeX template for reports (template-report.tex) may be downloaded from the IMS website `https://irishmathsoc.org/bulletin/index.php?file=instruct`. We do not dictate any rigid format to be used by reporters on our scientific meetings or sponsored meetings. Reports have sometimes included all abstracts, but it has become more usual just to give a link to the website where these can be found. This is probably more appropriate in this networked environment, although one might worry about the permanence of some of these conference websites. It might be worth the effort of reporters giving a self-contained account, for the record, since the IMS website has a good chance of remaining up indefinitely. Our template for meeting reports uses the `bimsplain` style, which is much less restrictive than the `bims` style used for articles, classroom notes and the like, and it should allow reasonably easy transfer of content from other formats.

Members are encouraged to submit (and promote the submission by others of) papers for the Bulletin that are of general interest to readers. Apart from short research articles and classroom notes, the Bulletin has carried articles on history of mathematics and mathematical education, and well-written and interesting material of this kind remains welcome. Expert surveys of active research areas are particularly appreciated, and I am pleased to have two in the present issue. I would prefer it if the authors of short proofs of FLT, RH, and the like would spare me their attentions.

The Bulletin is entirely the work of volunteers. The work of the Editor is made much lighter by the generous assistance of the Editorial Board members, by the Book Reviews Editor, Eleanor Lingham, the Problem Page Editor, Ian Short, the Website Manager, Michael Mackey, and the DOI manager and guru of last resort, David Malone. I am pleased to report that Colm Mulcahy has agreed to take on the editing of obituaries for the future.

For a limited time, beginning as soon as possible after the online publication of this Bulletin, a printed and bound copy may be ordered online on a print-on-demand basis at a minimal price[1].

Editor, Bulletin IMS, Department of Mathematics and Statistics, Maynooth University, Co. Kildare W23 HW31, Ireland.
E-mail address: ims.bulletin@gmail.com

[1]Go to www.lulu.com and search for *Irish Mathematical Society Bulletin*.

LINKS FOR POSTGRADUATE STUDY

The following are the links provided by Irish Schools for prospective research students in Mathematics:

DCU: mailto://maths@dcu.ie
TUD: mailto://chris.hills@tudublin.ie
ATU: mailto://creedon.leo@atu.ie
MTU:
http://www.ittralee.ie/en/CareersOffice/StudentsandGraduates/PostgraduateStudy/
UG: mailto://james.cruickshank@nuigalway.ie
MU: mailto://mathsstatspg@mu.ie
QUB:
http://web.am.qub.ac.uk/wp/msrc/msrc-home-page/postgrad_opportunities/
TCD: http://www.maths.tcd.ie/postgraduate/
UCC: http://www.ucc.ie/en/matsci/postgraduate/
UCD: mailto://nuria.garcia@ucd.ie
UL: mailto://sarah.mitchell@ul.ie

The remaining schools with Ph.D. programmes in Mathematics are invited to send their preferred link to the editor.

E-mail address: ims.bulletin@gmail.com

Letters to the Editor

The IMS and the IMU
From Tom Carroll

At last December's committee meeting, it was suggested that I write a letter to the Bulletin on the matter of Ireland's membership of the International Mathematical Union (the IMU).

The International Mathematical Union (`https://www.mathunion.org`) encourages and supports international mathematical activities globally across all areas of pure mathematics, applied mathematics, and mathematics education. The International Congress of Mathematicians (the ICM) is organised under its auspices. Every four years, prestigious prizes including the Fields Medals are awarded at the Opening Ceremony of the ICM. The Secretariat of the IMU is currently based in Berlin and runs day-to-day business. The Executive Committee, which includes the President Hiraku Nakajima (Japan) and Secretary General Christoph Sorger (France), is drawn from the highest international echelons of mathematics.

The members of the IMU are countries, currently numbering 85 in total. The level of membership ranges from one to five and reflects the mathematical stature of the member country. Ireland has been a member, a Group II member, since the 1950s. The Group V members are the mathematical (and political) powerhouses internationally: Brazil, Canada, China, France, Germany, Israel, Italy, Japan, Republic of Korea, Russia, the United Kingdom, and the United States of America. Group n countries are entitled to n votes at the General Assembly (GA) held every four years around the time of the ICM. Derek Kitson (Secretary of the IMS) and I represented Ireland at the GA in Helsinki last June. The membership dues increase nonlinearly as a function of the group number: for Group II member countries, including Ireland, these are currently €2,920 p.a. while Group V members contribute €17,520.

Each member country is represented via an Adhering Organization (AO), which may be its principal academy, a mathematical society, its research council or some other institution or association of institutions, or an appropriate agency of its government. Up to late 2019/early 2020, Ireland's AO was the Royal Irish Academy (RIA). As I understand it, the main reason the RIA decided to no longer act as Ireland's AO was its inability, or unwillingness, to continue to pay Ireland's annual membership dues. The IMS took over this role in 2020 - see Pauline Mellon's President's Report in *IMS Bulletin 86* for the background to this decision. The feeling then, and now, is that Ireland's membership is essential if we are to consider ourselves as seriously engaged in the world of Mathematics. Letting our membership lapse, as would otherwise have been the case, was unthinkable. Here is a link to Ireland's page on the IMU website (`https://www.mathunion.org/imu-members/ireland`) where the IMS is listed as the AO.

Though a minuscule sum of money at a national scale, the annual IMU dues constitute a significant commitment for our Society. Ireland's 2020 membership was largely funded through the generosity of various research centres and departments (see the President's Report mentioned above). Membership dues since then have been paid directly from IMS funds, in part offset by the Society's reduced outlay during the covid period, in supporting conferences for example. Repeated efforts to elicit contributions from various arms of government and the backing of VPs for Research in our various institutions have been unsuccessful. The executive and the committee continue to explore new avenues towards meeting the Society's commitment to paying the annual IMU membership dues so that this encumbrance does not limit its other activities and initiatives.

It's fair to say that taking on the role of AO to the IMU was not on the Society's radar prior to December 2019. Nevertheless, 'We are where we are, however we got here. What matters is where we go next.' I personally believe that our Society can play a positive and proactive role as AO, as does the LMS for the UK and the DMV for Germany, and raise the profile of Ireland within the IMU. The IMS and its members are, after all, those with the greatest stake in Ireland's international mathematical reputation.

Tom Carroll, IMS President 2020-2022, Committee Member 2023-2024
UCC
Received 31-5-2023
t.carroll@ucc.ie

NOTICES FROM THE SOCIETY

Officers and Committee Members 2023

President	Dr Leo Creedon	ATU
Vice-President	Dr Rachel Quinlan	UG
Secretary	Dr Derek Kitson	MIC
Treasurer	Dr Cónall Kelly	UCC

Dr T. Carroll, Dr R. Flatley, Dr R. Gaburro, Prof. M. Mathieu, Prof. A. O'Shea, Dr R. Ryan, Assoc. Prof. H. Šmigoc, Dr N. Snigireva.

Local Representatives

Belfast	QUB	Prof M. Mathieu
Carlow	SETU	Dr D. Ó Sé
Cork	MTU	Dr J. P. McCarthy
	UCC	Dr S. Wills
Dublin	DIAS	Prof T. Dorlas
	TUD, City	Dr D. Mackey
	TUD, Tallaght	Dr C. Stack
	DCU	Prof B. Nolan
	TCD	Prof K. Soodhalter
	UCD	Dr R. Levene
Dundalk	DKIT	Mr Seamus Bellew
Galway	UG	Dr J. Cruickshank
Limerick	MIC	Dr B. Kreussler
	UL	Mr G. Lessells
Maynooth	MU	Prof S. Buckley
Sligo	ATU	Dr L. Creedon
Tralee	MTU	Prof B. Guilfoyle
Waterford	SETU	Dr P. Kirwan

Applying for I.M.S. Membership

(1) The Irish Mathematical Society has reciprocity agreements with the American Mathematical Society, the Deutsche Mathematiker Vereinigung, the Irish Mathematics Teachers Association, the London Mathematical Society, the Moscow Mathematical Society, the New Zealand Mathematical Society and the Real Sociedad Matemática Española.

(2) The current subscription fees are given below:

Institutional member	€200
Ordinary member	€30
Student member	€15
DMV, IMTA, NZMS or RSME reciprocity member	€15
AMS reciprocity member	$20
LMS reciprocity member (paying in Euro)	€15
LMS reciprocity member (paying in Sterling)	£15

The subscription fees listed above should be paid in euro by means of electronic transfer, a cheque drawn on a bank in the Irish Republic, or an international money-order.

(3) The subscription fee for ordinary membership can also be paid in a currency other than euro using a cheque drawn on a foreign bank according to the following schedule:

If paid in United States currency then the subscription fee is US$ 40.
If paid in sterling then the subscription is £30.
If paid in any other currency then the subscription fee is the amount in that currency equivalent to US$ 40.00.

The amounts given in the table above have been set for the current year to allow for bank charges and possible changes in exchange rates.

(4) Any member with a bank account in the Irish Republic may pay his or her subscription by a bank standing order using the form supplied by the Society.

(5) Any ordinary member who has reached the age of 65 years and has been a fully paid up member for the previous five years may pay at the student membership rate of subscription.

(6) Subscriptions normally fall due on 1 February each year.

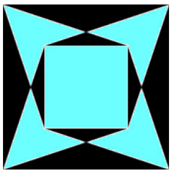
(7) Cheques should be made payable to the Irish Mathematical Society.

(8) Any application for membership must be presented to the Committee of the I.M.S. before it can be accepted. This Committee meets twice each year.

(9) Please send the completed application form, available at
`http://www.irishmathsoc.org/links/apply.pdf`
with one year's subscription to:

Dr Cónall Kelly
School of Mathematical Sciences
Western Gateway Building, Western Road
University College Cork
Cork, T12 XF62
Ireland

E-mail address: `subscriptions.ims@gmail.com`



Irish Math. Soc. Bulletin
Number 91, Summer 2023, 5–8
ISSN 0791-5578

Rex Dark, 1942-2022

ARNOLD FELDMAN, JOHN MCDERMOTT AND MARTIN NEWELL

Rex Dark in 2005 (Photograph: Tracy Feldman)

Rex Dark was born in Huddersfield and raised in Leatherhead. An excellent student from the start, he won several scholarships: first to his preparatory school, then to the

Key words and phrases. Rex Dark, Obituary.
Received on 15-12-2022.
DOI:10.33232/BIMS.0091.5.8.

prestigious Charterhouse School in Godalming, Surrey and from there to Magdalene College, Cambridge. Having achieved a double first degree he undertook research in Group Theory under the supervision of J. E. Roseblade. This led to the award of a Ph.D in 1968, for a thesis entitled "Nilpotent products of groups of prime order". After a few years teaching in Cambridge, Rex came to University College, Galway (later NUIG, now University of Galway) in 1973 on what was originally a one-year appointment. Fortunately for the UCG Mathematics department - and indeed for the college and local community - Rex liked what he found and decided to stay. He remained in Galway for the rest of his working life, taking early retirement in 2003 having reached the age of 60. After he retired from teaching, Rex continued to work on a variety of problems in soluble and insoluble groups. His latest works, with collaborators from Spain and the US, were published in 2022, and he left nearly completed manuscripts that will be submitted for publication by his collaborators.

Rex was a wonderful colleague, and his contributions to the UCG Mathematics department - and to the wider mathematical community - were substantial and very much appreciated. As a lecturer Rex was known for the clarity of his presentations and his willingness to help students with their work in and out of the classroom. His research talks were equally well-organized, clear, and insightful. In addition to the normal teaching and research duties, he undertook his full share of the administrative chores. He was one of the founders of the Irish Mathematics Society and of the series of annual Groups in Galway meetings. He remained a staunch supporter of both through the years, including in his retirement.

Rex was a master of the example and counterexample. His examples were beautifully constructed using his vast knowledge of techniques, including those of his own invention. Underlying them was his broad and deep understanding of groups coupled with extraordinary attention to detail, often involving intricate computations, generally done by hand. He had an extraordinary mental capacity, able to develop intricate computations before writing them down, but also created computer programs as needed. He was endlessly patient, working to perfect any project, producing draft after draft of increasingly comprehensive, impeccably expressed, carefully referenced results until he was satisfied that he had accounted for everything possible.

The impact of Rex's work was substantial. He made significant contributions, either alone or in different collaborations, particularly answering open questions and conjectures by various authors. Early in his career, he produced the first complete group of odd order (further tricky constructions provided an example of a complete group of the smallest possible odd order). His contribution to the study of Camina groups was also remarkable.

The impact of his work in the theory of Fitting classes and injectors deserves special mention; the name of Rex Dark appears among the most relevant figures in the development of the theory. His 1972 paper "Some examples in the theory of injectors of finite soluble groups" so greatly changed the way group theorists looked at Fitting classes that Doerk and Hawkes, in their comprehensive 1992 volume Finite Soluble Groups, devoted two complete sections, more than 45 pages, to what they termed "Dark's construction – the theme & variations". This paper contains the first publication of a Fitting class whose Fischer $\mathfrak{F}$-subgroups were not $\mathfrak{F}$-injectors. (Later, he would discover also the first Fitting set with non-pronormal Fischer $\mathfrak{F}$-subgroups.) B. Hartley, in the review about this paper published in the Mathematical Reviews, wrote referring to this example: "Fitting classes of any intricacy seem rather difficult to handle, and the problem is to select one that is sufficiently complicated but still tractable. This the author succeeds in doing with considerable ingenuity."The main example presented in that 1972 paper was actually less complicated than the first one Rex discovered. This more complicated, unpublished example was modified slightly in Lockett's 1973 paper presenting the first group with non-system-permutable $\mathfrak{F}$-injectors. More than 45 years later, Rex modernized the construction of that unpublished example for use in a paper showing that said example was in a technical sense the smallest possible.

The number of basic questions remaining unanswered in the Theory of Fitting classes indicates the difficulty of the theory, especially in constructing suitable examples, and shows the value of Rex's work. That work also contributed answers to other related fundamental questions, including different constructions of the associated injectors, and permutability, for the dominant Fitting class of finite soluble groups with central π-socle, as well as characterizations of injectors without recourse to the concept of a Fitting set.

His most recent papers involved generalizations of Carter subgroups and injectors in π-separable groups, groups with automorphism groups of smaller order than the groups themselves, and π-special modules. Referring to the first result just mentioned, in the corresponding review published in the Mathematical Reviews, Luis M. Ezquerro writes: "In my opinion, this is the best tribute to Carter on the occasion of his recent death on February 21, 2022."

Rex was also very committed to many non-mathematical interests. In college, for example, he was an active union member and was shop steward for a while. He was also a long-time member of the UCG Mountaineering Club. In addition he was deeply involved with the Galway branch of Alliance Francaise and with the Galway Mountain Rescue Team. Outside UCG, his primary loyalty was perhaps to Saint Nicholas' Collegiate Church. He was a chorister there for a while, and took on the position of treasurer for six years. Moreover, he took to the stage with the Lamplighters, a drama group associated with the church. He continued his service to the Church of Ireland in Westport, where he spent the last of his retirement years.

Rex was a familiar figure making his regular cycling commute from Moycullen to UCG. He also had a small folding bike — one of the first around Galway — which he kept in college to use for quick trips into town. If he wasn't cycling, he was probably walking, thinking nothing of the six mile journey from Valencia to the maths department of the university in an outlying suburb. He would take the train with colleagues to be sociable, but walk when on his own.

He was also an inveterate traveller. In addition to regular journeys to his home in France he attended conferences in many places. He spent sabbaticals and/or working visits with collaborators in Germany, Italy, Spain and the United States. Moreover, he made many trips to see his brother Michael (his only sibling) when the latter worked abroad. Although he eventually acquired a debit card, he traveled without ever having a credit card, a feat that became more extraordinary with every year that passed.

Many warm tributes have been paid to him by former colleagues who remarked in particular on his courtesy, helpfulness, quiet but wry sense of humour and of course his brilliant mind. Former students also remember him with affection. They recall his patient mentoring and his generosity with his time, as well as his enthusiastic and inspiring lectures. Collaborators reported that it was an honor and a privilege to work with him, since Rex contributed his ideas, effort, and knowledge unreservedly, generously and with incredible humility.

Throughout his life Rex was very close to Michael and his family. In the beautiful eulogy given at his funeral, his niece caught the essence of the man when she concluded that he remained the same always — kind, modest, brilliant, and completely original.

(A version of this article has appeared in the London Mathematical Society Newsletter.)

References

[1] Arroyo-Jordá, M.; Arroyo-Jordá, P.; Dark, R.; Feldman, A. D.; Pérez-Ramos, M. D. Injectors in π-separable groups. Mediterr. J. Math. 19 (2022), no. 4, Paper No. 174, 15 pp.
[2] Dark, Rex; Feldman, Arnold D.; Pérez-Ramos, María Dolores Nilpotent length and system permutability. J. Algebra 589 (2022), 287–322.
[3] Dark, Rex; Feldman, Arnold D.; Pérez-Ramos, María Dolores Permutability of injectors with a central socle in a finite solvable group. J. Algebra 476 (2017), 48–84.
[4] Dark, R.; Feldman, A. D.; Pérez-Ramos, M. D. Extraspecially irreducible groups. Adv. Group Theory Appl. 2 (2016), 31–65.
[5] Curran, M. John; Dark, Rex S. Complete groups of order $3p^6$. Adv. Group Theory Appl. 2 (2016), 1–12.
[6] Dark, Rex; Feldman, Arnold D.; Pérez-Ramos, María Dolores Injectors with a central socle in a finite solvable group. J. Algebra 381 (2013), 209–232.
[7] Dark, Rex; Feldman, Arnold D.; Pérez-Ramos, María Dolores Injectors with a normal complement in a finite solvable group. J. Algebra 333 (2011), 139–160.
[8] Dark, Rex; Feldman, Arnold D.; Pérez-Ramos, María Dolores Persistent characterizations of injectors in finite solvable groups. J. Group Theory 12 (2009), no. 4, 511–538.
[9] Dark, Rex; Feldman, Arnold D. Characterization of injectors in finite soluble groups. J. Group Theory 9 (2006), no. 6, 775–785.
[10] Dark, Rex A characterisation of injectors of finite soluble groups. Irish Math. Soc. Bull. No. 56 (2005), 29–36.
[11] Li, Shirong; Dark, Rex S. Finite groups with a system of nilpotent subgroups. Algebra Colloq. 12 (2005), no. 2, 199–204.
[12] Dark, Rex; Feldman, Arnold Fischer subgroups, Fitting height, and pronormality. J. Group Theory 5 (2002), no. 2, 145–162.
[13] Li, Shirong; Dark, Rex S. Finite groups with a system of nilpotent subgroups containing the Sylow subgroup. Math. Proc. R. Ir. Acad. 98A (1998), no. 1, 81–86.
[14] Dark, Rex; Scoppola, Carlo M. On Camina groups of prime power order. J. Algebra 181 (1996), no. 3, 787–802.
[15] Newell, Martin L.; Dark, Rex S. On metabelian groups of exponent eight. Publ. Math. Debrecen 35 (1988), no. 3-4, 295–300 (1989).
[16] Dark, Rex; Newell, Martin Isotropic tensors and symmetric groups. Irish Math. Soc. Newslett. No. 10 (1984), 34–45.
[17] Dark, Rex S.; Newell, Martin L. On 2-generator metabelian groups of prime-power exponent. Arch. Math. (Basel) 37 (1981), no. 5, 385–400.
[18] Newell, M. L.; Dark, R. S. On certain groups with a fourth power endomorphism. Proc. Roy. Irish Acad. Sect. A 80 (1980), no. 2, 167–172.
[19] Dark, R. S.; Newell, M. L. On conditions for commutators to form a subgroup. J. London Math. Soc. (2) 17 (1978), no. 2, 251–262.
[20] Dark, R. S. On hypercentral groups and Artinian modules. J. Algebra 42 (1976), no. 2, 597–599.
[21] Dark, R. S. A complete group of odd order. Math. Proc. Cambridge Philos. Soc. 77 (1975), 21–28.
[22] Dark, Rex S. Some examples in the theory of injectors of finite soluble groups. Math. Z. 127 (1972), 145–156.
[23] Dark, Rex; Rhemtulla, Akbar H. On R_0-closed classes, and finitely generated groups. Canadian J. Math. 22 (1970), 176–184.
[24] Dark, R. S. A prime Baer group. Math. Z. 105 (1968), 294–298.
[25] Dark, R. S. On subnormal embedding theorems for groups. J. London Math. Soc. 43 (1968), 387–390.

(Feldman) 25 River Bend Park, Lancaster, Pennsylvania 17602, USA

(McDermott and Newell) School of Mathematical Sciences, Galway University
E-mail address, Feldman: afeldman@fandm.edu
E-mail address, McDermott: john.p.mcdermott@nuigalway.ie
E-mail address, Newell: martin.newell@nuigalway.ie

Irish Math. Soc. Bulletin
Number 91, Summer 2023, 9–32
ISSN 0791-5578

From CT scans to 4-manifold topology via neutral geometry

BRENDAN GUILFOYLE

Abstract. In this survey paper the ultrahyperbolic equation in dimension four is discussed from a geometric, analytic and topological point of view. The geometry centres on the canonical neutral metric on the space of oriented geodesics of 3-dimensional space-forms, the analysis discusses a mean value theorem for solutions of the equation and presents a new solution of the Cauchy problem over a certain family of null hypersurfaces, while the topology relates to generalizations of codimension two foliations of 4-manifolds.

The air is full of an infinity of straight lines and rays
which cut across each other without displacing each other and
which reproduce on whatever they encounter
the true form of their cause.

Leonardo da Vinci
MS. A. 2v, 1490

1. Introduction

Our starting point is, as the title suggests, the acquisition of density profiles of biological systems using the loss of intensity experienced by a ray traversing the system. Basic mathematical physics arguments imply that this loss is modelled by the integral of the density function along the ray. One goal of Computerized Tomography is to invert the X-ray transform: reconstruct a real-valued function on $\mathbb{R}^3$ from its integrals over families of lines.

The reconstruction of a function on the plane from its value on all lines, or more generally, a function on Euclidean space from its value on all hyperplanes, dates back at least to Johann Radon [62]. One could argue that Allan MacLeod Cormack's 1979 Nobel prize for the theoretical results behind CAT scans [11] is the closest that mathematics has come to winning a Nobel prize, albeit in Medicine. The choice of axial rays reduces the inversion of the X-ray transform to that of the Radon transform over planes in $\mathbb{R}^3$ [43].

The basic problems of tomography - acquisition and reconstruction - arise far more widely than just medical diagnostics, finding application in industry [74], geology [70], archaeology [58] and transport security [56]. Indeed, advances in CT technology, trialed in Shannon Airport recently, could warrant the removal of the 100ml liquid rule for airplane travellers globally [63].

Rather surprisingly, sitting behind the X-ray transform and its many applications is a largely unstudied second order differential equation: the ultrahyperbolic equation.

2020 *Mathematics Subject Classification.* 53A25,35Q99.
Key words and phrases. Ultrahyperbolic equation, neutral geometry, X-ray transform, 4-manifold topology.
Received on 22-12-2022; revised 1-6-2023.
DOI:10.33232/BIMS.0091.9.32.

For a function u of four variables (X_1, X_2, X_3, X_4) the equation is

$$\frac{\partial^2 u}{\partial X_1^2} + \frac{\partial^2 u}{\partial X_2^2} - \frac{\partial^2 u}{\partial X_3^2} - \frac{\partial^2 u}{\partial X_4^2} = 0. \tag{1}$$

The reasons for the relative paucity of mathematical research on the equation despite the link to tomography will be discussed below.

The purpose of this mainly expository paper is to describe recent research on the ultrahyperbolic equation, its geometric context and its applications. It turns out that the ultrahyperbolic equation is best viewed in terms of a conformal class of neutral metrics and that in this context it advances new paradigms that can contribute to the understanding of four dimensional topology. We now discuss the mathematical background of this undertaking before giving a more detailed summary of the paper.

1.1. **Background.** The *X-ray transform* of a real valued function on $\mathbb{R}^3$ is defined by taking its integral over (affine) lines of $\mathbb{R}^3$. That is, given a real function $f : \mathbb{R}^3 \to \mathbb{R}$ and a line γ in $\mathbb{R}^3$, let

$$u_f(\gamma) = \int_\gamma f dr,$$

where dr is the unit line element induced on γ by the Euclidean metric on $\mathbb{R}^3$.

Thus we can view the X-ray transform of a function f (with appropriate behaviour at infinity) as a map $u_f : \mathbb{L}(\mathbb{R}^3) \to \mathbb{R} : \gamma \mapsto u_f(\gamma)$, where $\mathbb{L}(\mathbb{R}^3)$, or $\mathbb{L}$ for short, is the space of oriented lines in $\mathbb{R}^3$. Here we pick an orientation on the line to simplify later local constructions, much as Leonardo does when invoking *rays* as distinct from lines, and note that the space $\mathbb{L}$ double covers the space of lines.

In comparison, the *Radon transform* takes a real-valued function on $\mathbb{R}^3$ and integrates it over *planes* in $\mathbb{R}^3$. By elementary considerations, the space of affine planes in $\mathbb{R}^3$ is three dimensional, equal to the dimension of the underlying space, while the space of oriented lines is four dimensional.

Thus, by dimension count, if we consider the problem of inverting the two transforms, given a function on planes one can reconstruct the original function on $\mathbb{R}^3$, while the problem is over-determined for functions on lines. The consistency condition for a function on line space to come from an integral of a function on $\mathbb{R}^3$ is exactly the ultrahyperbolic equation [46].

Viewed simply as a partial differential equation, equation (1) is neither elliptic nor hyperbolic, and so many standard techniques of partial differential equation do not apply. Indeed, in early editions of their influential classic *Methods of Mathematical Physics*, Richard Courant and David Hilbert showed that the ultrahyperbolic equation in $\mathbb{R}^{2,2}$ has an ill-posed Cauchy boundary value problem when the boundary has Lorentz signature, thus relegating the equation as unphysical in a mechanical sense.

It was Fritz John who in 1937 proved that, to the contrary, the ultrahyperbolic equation can have a well-posed characteristic boundary value problem if the boundary 3-manifold is assumed to be *null*, rather than Lorentz [46]. Later editions of Courant and Hilbert's book acknowledge John's contribution and his discovery of the link to line space, but study of the ultrahyperbolic equation never took off in the way that it did for elliptic and hyperbolic equations.

On the other hand, by reducing the X-ray transform to the Radon transform for certain null configurations of lines, Cormack side-stepped the ultrahyperbolic equation altogether. Moreover, for applied mathematicians, the equation, or its associated John's equations, arises mainly as a compatibility condition if more than a 3-manifold's worth of data is acquired. Its possible utility from that perspective therefore is to check such excess data, rather than to help reconstruct the function.

Our first goal, contained in Section 2 is the geometrization of the ultrahyperbolic equation. In particular, we view it as the Laplace equation of the canonical metric $\mathbb{G}$ of signature $(++--)$ on the space $\mathbb{L}$ of oriented lines in $\mathbb{R}^3$ [36]. The fact that $\mathbb{G}$ is conformally flat and has zero scalar curvature means that a conformal multiple of a harmonic function satisfies the flat ultrahyperbolic equation (1). Fritz John did not explicitly use the neutral metric, but at the cost of the introduction of unmotivated multiplicative factors in calculations, factors that can now be related with the conformal factor of the metric.

The introduction of the neutral metric not only clarifies the ultrahyperbolic equation, but it highlights the role of the conformal group in tomography. Properties such as conformal flatness of a metric, zero distance between points or nullity of a hypersurface are properties of the conformal class of a metric. Moreover, mathematical results can be extended by applying conformal maps [9].

Section 2 describes how these neutral conformal structures arise in the space of oriented geodesics of any 3-dimensional space-form, namely $\mathbb{R}^3$, $\mathbb{S}^3$ and $\mathbb{H}^3$. The commonality between these three spaces allows one to apply many of the results (mean value theorem, doubly ruled surfaces, null boundary problems) to non-flat spaces. Surprisingly, electrical impedance tomography calls for negative curvature and so tomography in hyperbolic 3-space is not quite as fanciful as it may at first seem - see [4]. The link between the ultrahyperbolic equation and the neutral metric on the space of oriented geodesics in $\mathbb{H}^3$ as given in Theorem 8 is new and so the full proof is given below.

In Section 3 conformal methods are used to extend both a classical mean value theorem and its interpretation in terms of doubly ruled surfaces in $\mathbb{R}^3$. Aside from the discussion of the conformal extension of the mean value theorem, the section contains a new geometric formula for a solution of the ultrahyperbolic equation given only values on the null hypersurface formed by lines parallel to a fixed plane. In fact, this example was considered by John, but the geometric version we present using the null cone of the neutral metric has not appeared elsewhere.

The final Section turns to global aspects of complex points on Lagrangian surfaces in $\mathbb{L}$ and an associated boundary value problem for the Cauchy-Riemann operator. This proof of the Carathéodory Conjecture using the canonical neutral metric on the space of oriented lines [35] is under review, but significant parts of the arguments have now appeared in print. In particular, the essence as to *why* the Conjecture is true - namely the size of the Euclidean group - has been established [30] and shown to be sharp [26].

The efficacy of second order methods of parabolic partial differentiation in higher codimension has also been proven in this context for both interior [32] and boundary problems [28]. The final argument hinges on the technical point as to whether a hyperbolic angle condition in codimension two in dimension four can be made *sticky* enough to confine the boundary of a line congruence evolving under mean curvature flow. This is the sole remaining part of the proof under review.

Having established the *why*, this approach to the Carathédory Conjecture also lends itself to other independent methods of completion - one needs only to establish the existence of enough holomorphic discs attached to a given Lagrangian surface and the Conjecture follows. Indeed, a local index bound [34] and a conjecture of Toponogov [31] would also follow from existence of such families. This could be proven, for example, by the use of the method of continuity and pseudo-holomorphic curves [25], which would be a first order rather than second order proof. In any event, the acceptance that this infamous Conjecture has been finally put to rest will probably only come about when it has been proven at least twice.

A positive outcome of these developments has been the first application of differential geometry in the theory of complex polynomial: the index bound for an isolated umbilic point on a real analytic surface has been shown to restrict the number of zeros inside the unit circle for a polynomial with self-inversive second derivative [29]. This and related issues are discussed in more detail in Section 4.

The reason codimension two has a special significance in four dimensional topology is briefly discussed and the final section considers topological obstructions to neutral metrics as applied to closed 4-manifolds. In the case where the 4-manifold is compact with boundary, many open questions remain about what geometric information from a neutral metric can be seen at the boundary. Whether for a neutral 4-manifold with null boundary, coming full circle, it is possible to X-ray the inside and explore its topology.

2. The Geometry of Neutral Metrics

This section discusses the geometry of metrics of indefinite signature $(++--)$. While the study of positive definite metrics and Lorentz metrics are very well-developed, the neutral signature case is less well understood, even in dimension four. Rather than the general theory, of which [13] is a good survey, the section will focus on spaces of geodesics and the invariant neutral structures associated with them.

2.1. The Space of Oriented Lines.

The space $\mathbb{L}$ of oriented lines (or *rays*) of Euclidean $\mathbb{R}^3$ can be identified with the set of tangent vectors of $\mathbb{S}^2$ by noting that

$$\mathbb{L} = \{\vec{U}, \vec{V} \in \mathbb{R}^3 \mid \ |\vec{U}| = 1 \ \text{ and } \ \vec{U} \cdot \vec{V} = 0 \ \} = T\mathbb{S}^2, \tag{2}$$

where $\vec{U}$ is the direction vector of the line and $\vec{V}$ the perpendicular distance vector to the origin.

Topologically, $\mathbb{L}$ is a non-compact simply connected 4-manifold which can be viewed as the two dimensional vector bundle over $\mathbb{S}^2$ with Euler number two. One can see the Euler number by taking the zero section, which is the 2-sphere of oriented lines through the origin and perturbing it to another sphere of oriented lines (the oriented lines through a nearby point, for example). The two spheres are easily seen to intersect in two oriented lines, hence the Euler number of the bundle is two.

This space comes with a natural projection map $\pi : \mathbb{L} \to \mathbb{S}^2$ which takes an oriented line to its unit direction vector $\vec{U}$. In fact, there is a wealth of canonical geometric structures on $\mathbb{L}$, where canonical means invariant under the Euclidean group. These include a neutral Kähler structure, a fibre metric and an almost paracomplex structure. All three have a role to play in what follows and so we take some time to describe them in detail.

To start with the Kähler metric on $\mathbb{L}$, one has

Theorem 1. [36] *The space $\mathbb{L}$ of oriented lines of $\mathbb{R}^3$ admits a metric $\mathbb{G}$ that is invariant under the Euclidean group acting on lines. The metric is of neutral signature $(++--)$, is conformally flat and scalar flat, but not Einstein.*

It can be supplemented by a complex structure J_0 and symplectic structure ω, so that $(\mathbb{L}, \mathbb{G}, J_0, \omega)$ is a neutral Kähler 4-manifold.

Here the complex structure J_0 is defined at a point $\gamma \in \mathbb{L}$ by rotation through 90^o about the oriented line γ. This structure was considered in a modern context first by Nigel Hitchin [42], who dated it back at least to Karl Weierstrass in 1866 [73].

The symplectic structure ω is by definition a non-degenerate closed 2-form on $\mathbb{L} = T\mathbb{S}^2$, and it can be obtained by pulling back the canonical symplectic structure on the cotangent bundle $T^*\mathbb{S}^2$ by the round metric on $\mathbb{S}^2$.

These two structures are invariant under Euclidean motions acting on line space and fit nicely together in the sense that $\omega(J\cdot, J\cdot) = \omega(\cdot, \cdot)$. The metric obtained by their composition $\mathbb{G}(\cdot,\cdot) = \omega(J\cdot,\cdot)$, is of neutral signature $(+ + --)$, however. The existence of a Euclidean invariant metric of this signature on line space was first noted by Eduard Study in 1891 [68], but its neutral Kähler nature wasn't discovered until 2005 [36]. Interestingly, the space of oriented lines in Euclidean $\mathbb{R}^n$ admits an invariant metric iff $n = 3$, and in this dimension it is pretty much unique [64]. This accident of low dimensions offers an alternative geometric framework to investigate the semi-direct nature of the Euclidean group in dimension three, one which expresses three dimensional Euclidean quantities in terms of neutral geometric quantities in four dimensions.

This is but one of the many accidents that arise in the classification of invariant symplectic structures, (para)complex structures, pseudo-Riemannian metrics and (para)Kähler structures on the space of oriented geodesics of a simply connected pseudo-Riemannian space of constant curvature or a rank one Riemannian symmetric space [1].

Returning to oriented line space, the neutral metric $\mathbb{G}$ at a point $\gamma \in \mathbb{L}$ can be interpreted as the angular velocity of any line near γ. If the angular velocity is zero - and hence the oriented lines are null-separated - then the lines either intersect or are parallel. One can adopt the projective view, which arises quite naturally, that parallel lines intersect at infinity, and then nullity of a curve with respect to the neutral metric implies the intersection of the underlying infinitesimal lines in $\mathbb{R}^3$. Nullity for higher dimensional submanifolds will be discussed in the next section.

The invariant neutral metric is not flat, although its scalar curvature is zero and its conformal curvature vanishes. The non-zero Ricci tensor has zero neutral length, but its interpretation in terms of a recognisable energy momentum tensor is lacking. Given the difference of signature to Lorentz spacetime, it is also difficult to see the usual physical connection as in general relativity.

Since the metric is conformally flat, there exist local coordinates (X_1, X_2, X_3, X_4) and a strictly positive function Ω so that it can be written as

$$ds^2 = \Omega^2(dX_1^2 + dX_2^2 - dX_3^2 - dX_4^2). \tag{3}$$

Such a metric has zero scalar curvature iff Ω satisfies the ultrahyperbolic equation, thus characterising a Yamabe-type problem for neutral metrics [50]. Such coordinates (X_1, X_2, X_3, X_4) were first constructed using the Plücker embedding on the space of lines by John [46], who showed that the compatibility condition for a function on line space to be the integral of a function on $\mathbb{R}^3$ is exactly the flat ultrahyperbolic equation in these coordinates.

Write $\mathbb{R}^{2,2}$ for $\mathbb{R}^4$ endowed with the flat neutral metric. In Section 3 the ultrahyperbolic equation will be considered in more detail and an explicit formula presented for data prescribed on a certain null hypersurface.

A peculiarity of neutral signature metrics in dimension four is the existence of 2-planes on which the induced metric is identically zero, so-called *totally null* 2-planes. In $\mathbb{R}^{2,2}$ there are a disjoint union of two S^1's worth of totally null 2-planes, termed $\alpha-$planes and $\beta-$planes.

One way to see these is to consider the null cone $\mathcal{C}_0$ at the origin. This is a cone over the 2-torus $S^1 \times S^1$ given by

$$X_1^2 + X_2^2 - X_3^2 - X_4^2 = 0.$$

An $\alpha-$plane is a cone over a diagonal in the torus $t \mapsto (X_1 + iX_2, X_3 + iX_4) = (e^{it}, e^{i(t+t_0)})$, while a $\beta-$plane is a cone over an anti-diagonal in the torus $t \mapsto (X_1 + iX_2, X_3 + iX_4) = (e^{it}, e^{-i(t+t_0)})$.

This null structure exists in the tangent space at a point in any neutral four manifold and if one can piece it together in a geometric way there can be global topological consequences. One natural question is whether the $\alpha-$planes or $\beta-$plane fields are integrable in the sense of Frobenius, thus having surfaces to which the plane fields are tangent. These are guaranteed for the invariant neutral metrics endowed on the space of oriented geodesics of any 3-dimensional space-form, as they are all conformally flat [15].

Roughly speaking, an $\alpha-$surface in a geodesic space is the set of oriented geodesics through a fixed point, while $\beta-$surfaces are the oriented geodesics contained in a fixed totally geodesic surface in the ambient 3-manifold. Thus a neutral metric on a geodesic space allows for the geometrization of both intersection and containment.

Restricting our attention to $\mathbb{R}^3$, the $\alpha-$planes in $\mathbb{L}$ are the oriented lines through a point or the oriented lines with the same fixed direction. The latter are the 2-dimensional fibres of the canonical projection $\pi : \mathbb{L} \to \mathbb{S}^2$ taking an oriented line to its direction.

The distance between parallel lines in $\mathbb{R}^3$ induces a fibre metric on $\pi^{-1}(p)$ for $p \in \mathbb{S}^2$. If ξ is a complex coordinates about the North pole of $\mathbb{S}^2$ given by stereographic projection and η the complex fibre coordinate in the projection $T\mathbb{S}^2 \to \mathbb{S}^2$, then the fibre metric has the form

$$d\tilde{s}^2 = \frac{4d\eta\, d\bar{\eta}}{(1+\xi\bar{\xi})^2}. \tag{4}$$

In Section 3.3 this arises in the X-ray transform from certain null data.

Note that the complex coordinates (ξ, η) on $\mathbb{L}$ are essentially the vectors U and V in definition (2), the direction and perpendicular distance to the origin. They are related to John's conformal flat coordinates (X_1, X_2, X_3, X_4) by

Proposition 2. [8] *For complex coordinates (ξ, η) on $T\mathbb{S}^2$, over the upper hemisphere $|\xi|^2 < 1$ the conformal coordinates (X_1, X_2, X_3, X_4) are*

$$X_1 + iX_2 = \frac{2}{1-\xi^2\bar{\xi}^2}\left(\eta + \xi^2\bar{\eta} - i(1+\xi\bar{\xi})\xi\right)$$

$$X_3 + iX_4 = \frac{2}{1-\xi^2\bar{\xi}^2}\left(\eta + \xi^2\bar{\eta} + i(1+\xi\bar{\xi})\xi\right).$$

We turn now to null 3-manifolds (or hypersurfaces) in a neutral 4-manifold. An example of such is the null cone of a point in $\mathbb{L}$. Fix any oriented line $\gamma_0 \in \mathbb{L}$ and define its *null cone* to be

$$C_0(\gamma_0) = \{\gamma \in \mathbb{L} \mid\ Q(\gamma_0, \gamma) = 0\},$$

where Q is the neutral distance function introduced by John [46]. For convenience introduce the complex conformal coordinates given in terms of the real conformal coordinates of equation (3) by

$$Z_1 = X_1 + iX_2 \qquad\qquad Z_2 = X_3 + iX_4.$$

If two oriented lines $\gamma, \tilde{\gamma}$ have complex conformal coordinates (Z_1, Z_2) and $(\tilde{Z}_1, \tilde{Z}_2)$ then the neutral distance function is

$$Q(\gamma, \tilde{\gamma}) = |Z_1 - \tilde{Z}_1|^2 - |Z_2 - \tilde{Z}_2|^2.$$

Two oriented lines have zero neutral distance iff either they are parallel or they intersect. The null cone arises in the formula for the ultrahyperbolic equation in Theorem 13.

More generally, null hypersurfaces in $\mathbb{L}$ can be understood as 3-parameter families of oriented lines in $\mathbb{R}^3$ as follows. The degenerate hyperbolic metric induced on a null hypersurface $\mathcal{H}$ at a point γ defines a pair of totally null planes intersecting on the null normal of the hypersurface in $T_\gamma\mathcal{H}$, one an $\alpha-$plane, one a $\beta-$plane. These plane fields can be integrable or contact, as explored in [20].

There is a unique $\alpha-$surface in $\mathbb{L}$ containing γ with tangent plane agreeing with the $\alpha-$plane at γ. Such a holomorphic Lagrangian surface is either the oriented lines through a point, or the oriented lines in a fixed direction. This is the neutral metric interpretation of the classical surface statement that a totally umbilic surface is either a sphere or a plane.

Thus, the $\alpha-$plane at $\gamma \in \mathbb{L}$ identifies a point on each $\gamma \subset \mathbb{R}^3$ (albeit at infinity) which is the centre of the associated $\alpha-$surface. The locus of all these centres in $\mathbb{R}^3$ as one varies over $\mathcal{H}$ will be called the *focal set* of the null hypersurface. A null hypersurface is said to be *regular* if the focal set is a submanifold of $\mathbb{R}^3$.

Proposition 3. *A regular null hypersurface $\mathcal{H}_n$ with focal set of dimension n must be one of the following:*

$\mathcal{H}_0$: *The set of oriented lines parallel to a fixed plane,*
$\mathcal{H}_1$: *The set of oriented lines through a fixed curve,*
$\mathcal{H}_2$: *The set of oriented lines tangent to a fixed surface.*

Assuming the fixed curve and fixed surface are convex, we have $\mathcal{H}_0 = \mathcal{H}_2 = S^1 \times \mathbb{R}^2$ and $\mathcal{H}_1 = S^2 \times \mathbb{R}$. The null cone of a point $\gamma \in \mathbb{L}$ is clearly an example of null hypersurface $\mathcal{H}_1$, the fixed curve being the line $\gamma \subset \mathbb{R}^3$.

On the other hand, the formula presented in Section 3.3 assumes data on a null hypersurface $\mathcal{H}_0$. Both the $\alpha-$ and $\beta-$planes in $\mathcal{H}_0$ are integrable, so it can be foliated by $\alpha-$surfaces (all the oriented lines in a fixed direction) and by $\beta-$surfaces (all oriented lines contained in a plane parallel to the fixed plane).

The α-foliation underpins the projection operator in the formula and it is not clear how the formula would look for data on null hypersurfaces of type $\mathcal{H}_1$ or $\mathcal{H}_2$, as the $\alpha-$planes are not in general integrable.

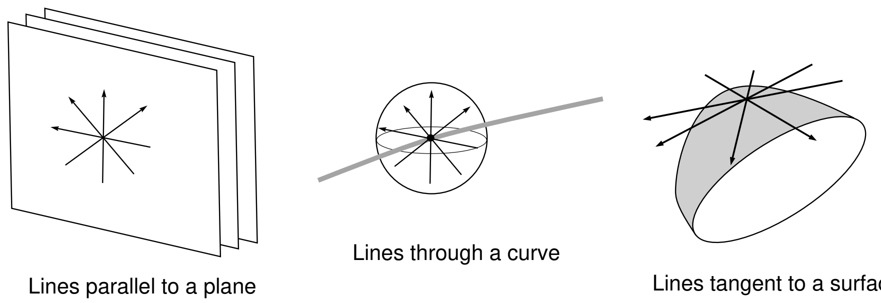


Figure 1. Regular null hypersurfaces in oriented line space

In Figure 1 the three types of regular null hypersurfaces $\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2$ are shown. The left null hypersurface is $\mathcal{H}_0$, the standard configuration for acquiring data in CT scans, and is discussed in Section 3.3.

Reconstruction using either of the other two null hypersurfaces would have advantages if one seeks to reduce the amount of radiation exposure during the scan. In particular, using the oriented lines $\mathcal{H}_1$ through a fixed line would reduce the exposure of each point to a semi-circle of radiation rather than the full circle in the $\mathcal{H}_0$. On the other hand, using the oriented lines $\mathcal{H}_2$ tangent to a convex surface would leave the interior occluded, and hence shielded completely from radiation. Whether either of these two configurations can be practically acquired by a physical scanner is another matter.

2.2. **Paracomplex Structures.** The complex structure J_0 on the space of oriented geodesics of a 3-dimensional space form evaluated at an oriented geodesic is obtained by rotation through 90^o about the geodesic. This almost complex structure is integrable in the sense of Nijinhuis, which for any almost complex structure J says

$$N^k_{ij} = J^m_j \partial_m J^k_i - J^m_i \partial_m J^k_j + J^k_m(\partial_i J^m_j - \partial_j J^m_i) = 0,$$

and thus a complex structure. This is due to the fact that the ambient space has constant curvature [42].

One can also take reflection of an oriented line in a fixed oriented line $\gamma \in \mathbb{L}$ to generate a map $J_1 : T_\gamma\mathbb{L} \to T_\gamma\mathbb{L}$ such that $J_1^2 = 1$ and the $\pm1-$eigenspaces are 2-dimensional. This *almost paracomplex structure* is not integrable in the sense of Nijinhuis and thus not a *paracomplex structure*. It is however anti-isometric with respect to the canonical neutral metric $\mathbb{G}$:

$$\mathbb{G}(J_1\cdot, J_1\cdot) = -\mathbb{G}(\cdot,\cdot).$$

Theorem 4. [19] *The space of oriented lines of Euclidean 3-space admits an invariant commuting triple (J_0, J_1, J_2) of a complex structure, an almost paracomplex structure and an almost complex structure, respectively, satisfying $J_2 = J_0J_1$. The complex structure J_0 is isometric, while J_1 and J_2 are anti-isometric. Only J_0 is parallel w.r.t. $\mathbb{G}$, and only J_0 is integrable.*

Composing the neutral metric $\mathbb{G}$ with the (para)complex structures J_0, J_1, J_2 yields closed 2-forms Ω_0 and Ω_1, and a conformally flat, scalar flat, neutral metric $\tilde{\mathbb{G}}$, respectively. The neutral 4-manifolds $(\mathbb{L}, \mathbb{G})$ and $(\mathbb{L}, \tilde{\mathbb{G}})$ are isometric. Only J_0 is parallel w.r.t. $\tilde{\mathbb{G}}$.

An almost paracomplex structure is an example of an *almost product structure*, in which a splitting of the tangent space at each point of the manifold is given, in this case $4 = 2 + 2$. Such pointwise splittings can only be extended over a manifold subject to certain geometric and topological conditions. For example

Theorem 5. [19] *A conformally flat neutral metric on a 4-manifold that admits a parallel anti-isometric or isometric almost paracomplex structure has zero scalar curvature.*

The parallel condition for an isometric almost paracomplex structure can be expressed in terms of the first order invariants of the eigenplane distributions:

Theorem 6. [19] *Let j be an isometric almost paracomplex structure on a pseudo-Riemannian 4-manifold. Then j is parallel iff the eigenplane distributions are tangent to a pair of mutually orthogonal foliations by totally geodesic surfaces.*

Canonical examples for neutral conformally flat metrics are the indefinite product of two surfaces of equal constant Gauss curvature, which have exactly this double foliation. It is instructive in this case to use the isometric paracomplex structure $j = I \oplus -I$ to flip the sign of the product metric. The result is a Riemannian metric which turns out to be Einstein. This construction holds more generally:

Theorem 7. [19] *Let (M, g) be a Riemannian 4-manifold endowed with a parallel isometric paracomplex structure j, and let the associated neutral metric be $g'(\cdot,\cdot) = g(j\cdot,\cdot)$. Then, g' is locally conformally flat if and only if g is Einstein.*

This transformation will be used in Section 4.3 to find global topological obstructions to parallel isometric paracomplex structures.

2.3. **The Space of Oriented Geodesics of Hyperbolic 3-Space.** In this section we consider the space $\mathbb{L}(\mathbb{H}^3)$ of oriented geodesics in three dimensional hyperbolic space $\mathbb{H}^3$ of constant sectional curvature -1. The canonical neutral metric on this space has been considered in detail [22] [23] [65], but its relation to the ultrahyperbolic equation has not. To illustrate the ideas of this paper, and explore the commonality with the flat case, proofs are provided in this section.

The space $\mathbb{L}(\mathbb{H}^3)$ of oriented geodesics in hyperbolic 3-space is diffeomorphic to that of oriented lines $\mathbb{L}(\mathbb{R}^3)$ in Euclidean 3-space $\mathbb{L}(\mathbb{H}^3) = \mathbb{L}(\mathbb{R}^3) = T\mathbb{S}^2$, but the projection map does not have the same geometric significance. In fact each oriented geodesic has *two* Gauss maps (the beginning and end directions at the boundary of the ball model for $\mathbb{H}^3$) and there is a natural embedding into $S^2 \times S^2$. Thus it is natural to view $\mathbb{L}(\mathbb{H}^3)$ as $S^2 \times S^2$ with the diagonal removed or, more geometrically, the *reflected* diagonal removed [23].

The canonical neutral metric $\tilde{\mathbb{G}}$ on $\mathbb{L}(\mathbb{H}^3)$ is conformally flat and scalar flat, thus relating the solutions of the flat ultrahyperbolic equation with harmonic functions, as in the case of $\mathbb{L}(\mathbb{R}^3)$.

Theorem 8. *For any compactly supported or asymptotically constant function f on hyperbolic 3-space, its X-ray transform is harmonic with respect to the canonical neutral metric:*

$$\triangle_{\tilde{\mathbb{G}}} u_f = 0,$$

where $\triangle_{\tilde{\mathbb{G}}}$ is the Laplacian of $\tilde{\mathbb{G}}$.

Proof. Consider the upper half-space model of hyperbolic 3-space $\mathbb{H}^3$, that is $(x_1, x_2, x_3) \in \mathbb{R}^3, x_3 \in \mathbb{R}_{>0}$ with metric

$$ds^2 = \frac{dx_1^2 + dx_2^2 + dx_3^2}{x_3^2}.$$

We can locally model the space of oriented geodesics in this model by $(\xi, \eta) \in \mathbb{C}^2$ where the unit parameterised geodesic is [23]

$$z = x_1 + ix_2 = \eta + \frac{\tanh r}{\bar{\xi}} \qquad\qquad x_3 = \frac{1}{|\xi| \cosh r}. \tag{5}$$

With respect to these coordinates the neutral metric is

$$ds^2 = -\frac{i}{4}\left(\frac{1}{\xi^2}d\xi^2 - \frac{1}{\bar{\xi}^2}d\bar{\xi}^2 + \bar{\xi}^2 d\eta^2 - \xi^2 d\bar{\eta}^2\right),$$

and the Laplacian is

$$\triangle_{\tilde{\mathbb{G}}} u = 8\mathrm{Im}\left(\frac{1}{\bar{\xi}^2}\partial_\eta^2 u + \partial_\xi(\xi^2 \partial_\xi u)\right).$$

Note that

$$\frac{\partial}{\partial r} = \frac{1}{\cosh^2 r}\left(\frac{1}{\bar{\xi}}\frac{\partial}{\partial z} + \frac{1}{\xi}\frac{\partial}{\partial \bar{z}} - \frac{\sinh r}{|\xi|}\frac{\partial}{\partial t}\right).$$

Now a straight-forward calculation establishes the following identity

$$\triangle_{\tilde{\mathbb{G}}} u_f = 4i \int_{-\infty}^{\infty} \frac{\partial}{\partial r}\left(\frac{1}{\bar{\xi}}\partial_z f - \frac{1}{\xi}\partial_{\bar{z}} f\right) dr = 4i \left[\frac{1}{\bar{\xi}}\partial_z f - \frac{1}{\xi}\partial_{\bar{z}} f\right]_{-\infty}^{\infty}.$$

Thus, by integration by parts, as long as the transverse gradient of f falls off at the boundary faster than $|\xi|$, the boundary terms vanish and we get

$$\triangle_{\tilde{\mathbb{G}}} u_f = 0.$$

□

In Section 3.1 unit (pseudo-)circles in flat planes are proven to be the domains of integration of a mean value theorem for solutions of the ultrahyperbolic equation and to generate doubly ruled surfaces in the underlying $\mathbb{R}^3$. We now present a local conformally flat coordinate system for $\mathbb{L}(\mathbb{H}^3)$ using the hyperboloid model of hyperbolic 3-space $\mathbb{H}^3$, which lets one explicitly construct such doubly ruled surfaces in $\mathbb{H}^3$.

In the hyperboloid model in Minkowski space $\mathbb{R}^{3+1}$, $\mathbb{H}^3$ is the hyperboloid $x_0^2 - x_1^2 - x_2^2 - x_3^2 = 1$ and the oriented geodesics are the intersections with oriented planes of Lorentz signature through the origin in $\mathbb{R}^{3+1}$.

An oriented geodesic in $\mathbb{H}^3$ in the ball model can be uniquely determined by the directions at the boundary $(\mu_1, \mu_2) \in \mathbb{S}^2 \times \mathbb{S}^2$. These directions (μ_1, μ_2) are exactly the null directions on the Lorentz plane.

The relationships between the complex coordinates $(\mu_1, \mu_2) \in \mathbb{C}^2$ obtained by stereographic projection on each $\mathbb{S}^2$ factor and the complex coordinates (ξ, η) introduced in Theorem 8 is

$$\xi = \tfrac{1}{2}\left(\bar{\mu}_1 + \tfrac{1}{\mu_2}\right)^{-1} \qquad\qquad \eta = \tfrac{1}{2}\left(-\mu_1 + \tfrac{1}{\bar{\mu}_2}\right).$$

Proposition 9. *If (μ_1, μ_2) are the standard holomorphic coordinates on $\mathbb{L}(\mathbb{H}^3)$, consider the complex combination*

$$Z_1 = \frac{(1+\mu_2\bar{\mu}_2)\bar{\mu}_1 + (1+\mu_1\bar{\mu}_1)\bar{\mu}_2 + i[(1-\mu_2\bar{\mu}_2)\bar{\mu}_1 - (1-\mu_1\bar{\mu}_1)\bar{\mu}_2]}{1-\mu_1\bar{\mu}_1\mu_2\bar{\mu}_2}$$

$$Z_2 = \frac{(1+\mu_2\bar{\mu}_2)\bar{\mu}_1 + (1+\mu_1\bar{\mu}_1)\bar{\mu}_2 - i[(1-\mu_2\bar{\mu}_2)\bar{\mu}_1 - (1-\mu_1\bar{\mu}_1)\bar{\mu}_2]}{1-\mu_1\bar{\mu}_1\mu_2\bar{\mu}_2}.$$

The flat neutral metric $ds^2 = dZ_1 d\bar{Z}_1 - dZ_2 d\bar{Z}_2$ pulled back by the above is equal to $\Omega^2\tilde{\mathbb{G}}$ where

$$\Omega = \frac{|1+\mu_1\bar{\mu}_2|^2}{1-|\mu_1|^2|\mu_2|^2}.$$

The inverse mapping from (μ_1, μ_2) to (Z_1, Z_2) is given by

$$\mu_1 = \tfrac{1}{2}(\bar{A}+\bar{B}) - \frac{\bar{A}-\bar{B}}{2|A-B|^2}\left(|A|^2 - |B|^2 + 2 - \sqrt{(|A|^2-|B|^2+2)^2 - |A-B|^2|A+B|^2}\right) \tag{6}$$

$$\mu_2 = \tfrac{1}{2}(\bar{A}-\bar{B}) - \frac{(\bar{A}+\bar{B})}{2|A+B|^2}\left(|A|^2 - |B|^2 + 2 - \sqrt{(|A|^2-|B|^2+2)^2 - |A-B|^2|A+B|^2}\right) \tag{7}$$

where $A = \frac{1}{2}(Z_1 + Z_2)$ and $B = \frac{1}{2i}(Z_1 - Z_2)$.

Proof. A direct calculation. □

In Section 3.2 these transformations will be used to construct surfaces in $\mathbb{H}^3$ that are ruled by geodesics in two distinct ways - doubly ruled surfaces.

3. The Ultrahyperbolic Equation

In this section solutions of the ultrahyperbolic equation (1) are studied. A mean value property for such solutions is presented along with its interpretation in terms of doubly ruled surfaces in $\mathbb{R}^3$. Classically it was known that a non-flat doubly ruled surface in $\mathbb{R}^3$ is either a one-sheeted hyperboloid or a hyperbolic paraboloid [40]. The construction of doubly ruled surfaces is extended to hyperbolic 3-space and the analogue of the 1-sheeted hyperboloid is exhibited. An explicit geometric formula is then given for the ultrahyperbolic equation with data given on a certain null hypersurface.

3.1. **Mean Value Theorem.** The X-ray transform takes a function $f : \mathbb{R}^3 \to \mathbb{R}$ to $u_f : \mathbb{L} \to \mathbb{R}$ by integrating over lines. In 1937 Fritz John showed that if a function f satisfies certain fall-off conditions at infinity (which hold for compactly supported functions), then u_f satisfies the ultrahyperbolic equation (1), [46].

The link between the ultrahyperbolic equation (1) and the neutral metric is

Theorem 10. [8] *Let $u : \mathbb{R}^{2,2} \to \mathbb{R}$ and $v : \mathbb{L} \to \mathbb{R}$ be related by $v = \Omega^{-1}u$, where Ω is the conformal factor.*

Then u is a solution of the ultrahyperbolic equation (1) iff v is in the kernel of the Laplacian of the neutral metric: $\Delta_{\mathbb{G}} v = 0$.

Leifur Asgeirsson [2] had earlier shown that solutions of the ultrahyperbolic equation satisfy a mean value property. In particular, for $u : \mathbb{R}^{2,2} \to \mathbb{R}$ a solution of equation (1) satisfies

$$\int_0^{2\pi} u(a + r\cos\theta, b + r\sin\theta, c, d)\ d\theta = \int_0^{2\pi} u(a, b, c + r\cos\theta, d + r\sin\theta)\ d\theta, \qquad (8)$$

for all $a, b, c, d \in \mathbb{R}$ and $r > 0$. The two domains of integration are circles of equal radius lying in a pair of orthogonal planes $\pi, \pi^\perp$ in $\mathbb{R}^{2,2}$ with definite induced metrics on them.

It can be shown that the mean value theorem holds over a much larger class of curves, namely the image of these circles under any conformal map of $\mathbb{R}^{2,2}$. We refer to such curves as *conjugate conics* and these turn out to be pairs of circles, hyperbolae and parabolae lying in orthogonal planes of various signatures:

Theorem 11. [8] [9] *Let S and $S^\perp$ be curves contained in orthogonal affine planes π and $\pi^\perp$ in $\mathbb{R}^{2,2}$, respectively, which are one of the following pairs:*

(1) Circles *with equal and opposite radii $\pm r_0$ when the two planes are definite,*
(2) Hyperbolae *with equal and opposite radii $\pm r_0$ when the two planes are indefinite,*
(3) Parabolae *in non-intersecting degenerate affine planes determined by the property that every point on $S \subset \pi$ is null separated from every point on $S^\perp \subset \pi^\perp$.*

Then the following mean value property holds for any solution u of the ultrahyperbolic equation:

$$\int_S u\ dl = \int_{S^\perp} u\ dl,$$

where dl is the line element induced on the curves by the flat metric g.

One can view this as a conformal extension of the original mean value theorem, one that intertwines the classical conic sections, the ultrahyperbolic equation and neutral geometry.

3.2. **Doubly Ruled Surfaces.** John also pointed out the relationship between the two circles in Asgeirsson's theorem and the double ruling of the hyperboloid of 1 sheet [46]. In fact, conjugate conics have been shown to correspond to the pairs of families of lines of all non-planar doubly ruled surfaces in $\mathbb{R}^3$.

Theorem 12. [9] *Let $S, S^\perp$ be two curves in $\mathbb{R}^{2,2}$ representing the two one-parameter families of lines $L, L^\perp$ in $\mathbb{R}^3$. Then $S, S^\perp$ are a pair of conjugate conics in $\mathbb{R}^{2,2}$ if and only if L and $L^\perp$ are the two families of generating lines of a non-planar doubly ruled surface in 3-space.*

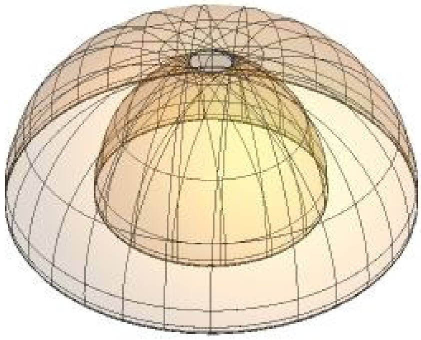


Figure 2. Halfspace model

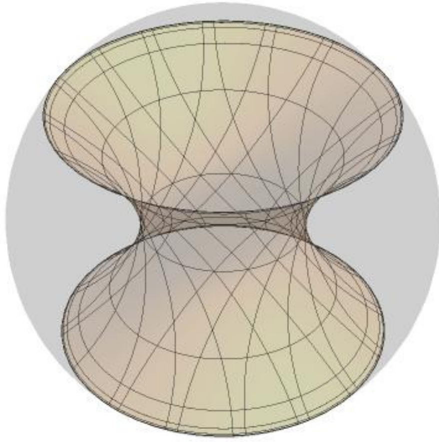


Figure 3. Ball model

The geometric reason these curves yield a doubly ruled surface is that every point on one curve is zero distance from every point on the other curve - this follows from the neutral Pythagoras Theorem! But, as mentioned earlier, zero distance between oriented lines implies intersection, we see that every line of one ruled surface intersects every line of the other ruling, hence the double ruling.

While this result was originally proven in $\mathbb{R}^3$, it holds in any 3-dimensional space of constant curvature, where the canonical neutral Kähler metric plays the same role. To demonstrate this, let us construct doubly ruled surfaces in 3-dimensional hyperbolic space $\mathbb{H}^3$.

Recall the conformal coordinates for $\mathbb{L}(\mathbb{H}^3)$ given in equations (6) and (7). To generate the hyperbolic equivalent of the 1-sheeted hyperboloid, the two curves (parameterized by u) are circles of radii $\pm r_0$ in two definite planes:

$$Z_1 = r_0 e^{iu} \qquad\qquad Z_2 = 0,$$

and

$$Z_1 = 0 \qquad\qquad Z_2 = r_0 e^{iu}.$$

For the curves we can view the doubly ruled surfaces in either the upper half-space model or the ball model of $\mathbb{H}^3$. For the former, one uses the equations (5), while for the latter one can use

$$x_1 + ix_2 = \frac{\mu_2(1+\mu_1\bar{\mu}_1)e^v - \mu_1(1+\mu_2\bar{\mu}_2)e^{-v}}{(1+\mu_1\bar{\mu}_1)(1+\mu_2\bar{\mu}_2)\cosh v + [(1+\mu_1\bar{\mu}_2)(1+\mu_2\bar{\mu}_1)(1+\mu_1\bar{\mu}_1)(1+\mu_2\bar{\mu}_2)]^{\frac{1}{2}}}$$

$$x_3 = \frac{(1+\mu_1\bar{\mu}_1)(1-\mu_2\bar{\mu}_2)e^v - (1+\mu_2\bar{\mu}_2)(1-\mu_1\bar{\mu}_1)e^{-v}}{2\left((1+\mu_1\bar{\mu}_1)(1+\mu_2\bar{\mu}_2)\cosh v + [(1+\mu_1\bar{\mu}_2)(1+\mu_2\bar{\mu}_1)(1+\mu_1\bar{\mu}_1)(1+\mu_2\bar{\mu}_2)]^{\frac{1}{2}}\right)}.$$

Figure 1 is a plot of a doubly ruled surface in the upper half-space model while Figure 2 is in the ball model of hyperbolic 3-space. These are the hyperbolic equivalent of the 1-sheeted hyperboloid, although they satisfy a fourth order (rather than second order) polynomial equation.

3.3. **Cauchy Problem for the Ultrahyperbolic Equation.** One way to reconcile the difference between the dimension of $\mathbb{L}(\mathbb{R}^3)$ and that of $\mathbb{R}^3$ is to consider the problem of determining the value of a solution $v : \mathbb{L} \to \mathbb{R}$ of the Laplace equation

$$\triangle_{\mathbb{G}} v = 0,$$

on all of oriented line space $\mathbb{L}$, given only the values of the function on a null hypersurface $\mathcal{H} \subset \mathbb{L}$.

Consider the case where the data is known on the hypersurface generated by all oriented lines parallel to a fixed plane in $P_0 \subset \mathbb{R}^3$ - the case of regular dimension zero focal set $\mathcal{H}_0$ in Proposition 3.

This null hypersurface is suitable as a boundary for the Cauchy problem, as proven by John [46]. In fact, it can be foliated both by $\alpha-$planes and $\beta-$planes - the former being the oriented lines parallel to P_0 in a fixed direction, while the latter are all oriented lines parallel to P_0 at a fixed height.

Denote

$$\mathcal{H} = \{\gamma \in \mathbb{L} \mid \ \gamma \parallel P_0 \ \}.$$

Clearly $\mathcal{H} = \mathbb{S}^1 \times \mathbb{C}$ and for convenience, suppose that P_0 is horizontal in standard coordinates, so that in complex coordinates the hypersurface is $\xi = e^{i\theta}$, since the only restriction on the oriented line is that its direction lies along the equator.

The distance between parallel lines in $\mathbb{R}^3$ induces the metric (4) and associated distance function $\|.\|$. In fact, there is an invariant metric on $\mathcal{H}$ with volume form $d^3Vol = d\eta \, d\bar{\eta} \, d\theta$.

Suppose that $\gamma_0 \notin \mathcal{H}$ and consider the intersection of this null hypersurface with the null cone $C_0(\gamma_0) \cap \mathcal{H} = \mathbb{S}^1 \times \mathbb{R}$. This surface intersects each fibre in an affine line. Let $Pr_0(\gamma)$ be the projection of γ onto this affine line with respect to the fibre metric: $Pr_0 : \mathbb{S}^1 \times \mathbb{R}^2 \to \mathbb{S}^1 \times \mathbb{R}$.

We now prove the following explicit geometric formula that determines the value of a solution of the ultrahyperbolic equation from its value on the null hypersurface of type $\mathcal{H}_0$ in $\mathbb{L}$:

Theorem 13. *If $v : \mathbb{L} \to \mathbb{R}$ is a function satisfying the ultrahyperbolic equation, then at an oriented line γ_0*

$$v(\gamma_0) = -\tfrac{1}{2\pi^2} \iiint_{\gamma \in \mathcal{H}} \frac{v(\gamma) - v(Pr_0(\gamma))}{\|\gamma - Pr_0(\gamma)\|^2} \, d^3Vol,$$

where $Pr_0(\gamma)$ is projection onto the intersection of the null cone of γ_0 with the α-plane through γ that lies in the null hypersurface $\mathcal{H}$.

Proof. Our starting point is Fritz John's formula (equation (13) of [46]) which gives the solution of the ultrahyperbolic equation at an oriented line γ_0 by the cylindrical average over all planes parallel to γ_0:

$$v(\gamma_0) = -\tfrac{1}{\pi} \int_0^\infty \frac{F(R) - F(0)}{R^2} dR, \tag{9}$$

where

$$F(R) = \tfrac{1}{2\pi} \int_0^{2\pi} \iint_{P_{(R,\alpha)}} \rho(r, s) dr ds \, d\alpha,$$

$P_{(R,\alpha)}$ is the plane parallel to γ_0 at a distance R and angle α, and (r, s) are flat coordinates on that plane.

Consider the map

$$z = \frac{1}{1 + \nu\bar{\nu}} \left(2\nu R + (e^{iA} - \nu^2 e^{-iA})r + i(e^{iA} + \nu^2 e^{-iA})s\right) \tag{10}$$

$$x_3 = \frac{1}{1 + \nu\bar{\nu}} \left((1 - \nu\bar{\nu})R - (\bar{\nu}e^{iA} + \nu e^{-iA})r - i(\bar{\nu}e^{iA} - \nu e^{-iA})s\right). \tag{11}$$

For fixed $R \in \mathbb{R}$, $\nu \in \mathbb{C}$ and $A \in [0, 2\pi)$, the map $(r, s) \mapsto (z(r, s), x_3(r, s)) \in \mathbb{R}^3$ paramaterizes the plane a distance R from the origin with normal direction ν. Changing A rotates the r- and s-axes in the plane.

By a translation we can assume γ_0 contains the origin and so has complex coordinates $(\xi = \xi_0, \eta = 0)$. Let us restrict attention to planes that are parallel γ_0. Thus the normal direction of $P_{(R,\nu)}$ is perpendicular to the direction of γ_0, we have

$$\nu = \frac{\xi_0 + e^{i\alpha}}{1 - \bar{\xi}_0 e^{i\alpha}},$$

where $\alpha \in [0, 2\pi)$.

The quantity R is then just the distance from the plane to the line γ_0. Finally we want to rotate the ruling by s on the plane so that it is horizontal and thus a curve in $\mathcal{H}$. Clearly this is achieved by

$$\nu = r_0 e^{iA},$$

or more explicitly

$$A = \tfrac{1}{2i} \ln \left[\frac{(\xi_0 + e^{i\alpha})(1 - \xi_0 e^{-i\alpha})}{(\bar{\xi}_0 + e^{-i\alpha})(1 - \bar{\xi}_0 e^{i\alpha})} \right] \qquad r_0 = \left[\frac{(\xi_0 + e^{i\alpha})(\bar{\xi}_0 + e^{-i\alpha})}{(1 - \xi_0 e^{-i\alpha})(1 - \bar{\xi}_0 e^{i\alpha})} \right]^{\frac{1}{2}}.$$

The first of these is invertible for fixed ξ_0, $A \leftrightarrow \alpha$.

The horizontal ruling for $P_{(A,\alpha)}$ is

$$z = \frac{2\nu}{1 + \nu\bar{\nu}} R + \frac{1 - \nu\bar{\nu}}{1 + \nu\bar{\nu}} r e^{iA} + i s e^{iA}$$

$$x_3 = \frac{1 - \nu\bar{\nu}}{1 + \nu\bar{\nu}} R - \frac{2|\nu|}{1 + \nu\bar{\nu}} r.$$

The direction of the ruling is

$$\frac{\partial}{\partial s} = i e^{iA} \frac{\partial}{\partial z} - i e^{-iA} \frac{\partial}{\partial \bar{z}}$$

so that the complex coordinates are $\xi = i e^{iA}$ and

$$\eta = \tfrac{1}{2}(z - 2x_3\xi - \bar{z}\xi^2) = -(r - iR) \left(\frac{r_0 - i}{r_0 + i} \right) e^{iA}.$$

Thus we have parameterized $\mathcal{H}$ by coordinates (R, α, r) and a straightforward calculation shows that the fibre metric is simply

$$d\eta d\bar{\eta} = dR^2 + dr^2 \qquad \text{and} \qquad d^3Vol = dr dR d\alpha.$$

The null cone of γ_0 consists of all lines that either intersect or are parallel to it. For non-horizontal γ_0 the null cone intersects the null hypersurface $\mathcal{H}$ at the lines that intersect γ_0, namely those with coordinates $(R = 0, \alpha, r)$ which is a line through the origin in each fibre. We have chosen γ_0 to contain the origin in $\mathbb{R}^3$, which is why the line in the fibre is through the origin. More generally the intersection of the null cone with a fibre is an affine line (not necessarily through the origin), as claimed.

Thus the fibre projection is simply $Pr_0(R, \alpha, r) = (0, \alpha, r)$ and

$$R = \|\gamma - Pr_0(\gamma)\|.$$

Now putting this together with the integral formula

$$\begin{aligned} v(\gamma_0) &= -\tfrac{1}{2\pi^2}\int_0^\infty\int_0^{2\pi}\frac{1}{R^2}\left[\iint_{P_{(R,\alpha)}}\rho(r,s)drds-\iint_{P_{(0,\alpha)}}\rho(r,s)drds\right]dRd\alpha \\ &= -\tfrac{1}{2\pi^2}\int_0^\infty\int_0^{2\pi}\int_{-\infty}^\infty\frac{v(R,\alpha,r)-v(0,\alpha,r)}{R^2}drdRd\alpha \\ &= -\tfrac{1}{2\pi^2}\iiint_{\gamma\in\mathcal{H}}\frac{v(\gamma)-v(Pr_0(\gamma))}{\|\gamma-Pr_0(\gamma)\|^2}\,d^3Vol, \end{aligned}$$

as claimed. □

4. Topological Considerations

In this section global topological aspects of neutral metrics and almost product structures are explored. These include the relationship between umbilic points on surfaces in $\mathbb{R}^3$ and complex points on Lagrangian surfaces in $\mathbb{L}$, and an associated boundary value problem for the Cauchy-Riemann operator. The significance of these constructions for a number of conjectures from classical surface theory is indicated.

Some background on the problems of 4-manifold topology are discussed with particular attention to codimension two. The significance of neutral metrics to these issues is that they are uniquely capable of quantifying codimension two topological phenomena, and thus can be used as geometric tools to resolve certain long-standing questions. For the case of closed 4-manifolds, we end with a discussion of topological obstructions that arise to certain neutral geometric structures.

4.1. **Global Results.** Topological aspects of neutral metrics become evident in the identification of complex points on Lagrangian surfaces in $\mathbb{L}$ with umbilic points on surfaces in $\mathbb{R}^3$ [37].

The Lagrangian surface $\Sigma\subset|mathbbL$ is formed by the oriented normal lines to the surface $S\subset\mathbb{R}^3$ and the index $i(p)\in\mathbb{Z}/2$ of an isolated umbilic point $p\in S$ on a convex surface is exactly one half of the complex index of the corresponding complex point $\gamma\in\Sigma$: $I(\gamma)=2i(p)\in\mathbb{Z}$. Thus problems of classical surface theory can be explored through studying Lagrangian surfaces in the four dimensional space of oriented lines $\mathbb{L}$ with its neutral metric $\mathbb{G}$.

The metric induced on a Lagrangian surface is Lorentz or degenerate - the degenerate points being the umbilic points of S and the null cone at γ being the principal directions of S at p. The indices of isolated umbilic points carry geometric information from the neutral metric and vice versa.

If an isolated umbilic point p has half-integer index then the principal foliation around p is non-orientable - it defines a line field rather than a vector field about the umbilic point. The foliation is orientable if the index is an integer. The following theorem establishes a topological version of a result of Ferdinand Joachimsthal [45] for surfaces intersecting at a constant angle:

Theorem 14. [33] *If S_1 and S_2 are smooth convex surfaces intersecting with constant angle along a curve that is not umbilic in either S_1 or S_2, then the principal foliations of the two surfaces along the curve are either both orientable, or both non-orientable.*

That is, if $i_1\in\mathbb{Z}/2$ is the sum of the umbilic indices inside the curve of intersection on S_1 and $i_2\in\mathbb{Z}/2$ is the sum of the umbilic indices inside the curve of intersection on S_2 then

$$2i_1=2i_2 \mod 2.$$

Pushing deeper, if one considers the problem of finding a holomorphic disc in $\mathbb{L}$ whose boundary lies on a given Lagrangian surface Σ, one encounters a classical problem of Riemann-Hilbert for the Cauchy-Riemann operator. Given a totally real surface Σ in a complex surface $\mathbb{M}$, the Riemann-Hilbert problem seeks a map $f : (D, \partial D) \to (\mathbb{M}, \Sigma)$ which is holomorphic: it lies in the kernel of the Cauchy-Riemann operator $\bar{\partial} f = 0$. For this to be an elliptic boundary value problem it is required that the boundary surface Σ be totally real i.e. has no complex points. In the Riemannian case Lagrangian implies totally real, and so Lagrangian boundary conditions are often used when the ambient metric is Riemannian.

In our case, due to the neutral signature of the metric formed by the composition of the symplectic structure (which defines Lagrangian) and the complex structure (which defines holomorphic), new features arise. In particular, Lagrangian surfaces may not be totally real, and therefore at complex points they are not suitable as a boundary condition for the $\bar{\partial}$-operator. If, however, the boundary surface is assumed to be spacelike with respect to the metric, then by the neutral Wirtinger identity it is also totally real and is suitable.

The deformation from Lagrangian to spacelike by the addition of a holomorphic twist can be achieved over an open hemisphere. This *contactification* of the problem throws away the surface S in $\mathbb{R}^3$, as the perturbed spacelike surface $\tilde{\Sigma}$ in $\mathbb{L}(\mathbb{R}^3)$ forms a 2-parameter family of twisting oriented lines in $\mathbb{R}^3$ that are not orthogonal to any surface. Any holomorphic disc with boundary lying on $\tilde{\Sigma}$ yields a holomorphic disc with boundary lying on Σ by subtracting the holomorphic twist and so the problems are equivalent over a hemisphere.

The Riemann-Hilbert problem then follows the standard case, with the linearisation at a solution defining an elliptic boundary value problem with analytic index $\mathcal{I}$ given by

$$\mathcal{I} = \text{Dim Ker } \bar{\partial} - \text{Dim Coker } \bar{\partial}.$$

The analytic index for the problem is well-known to be related to the Keller-Maslov index $\mu(\partial D, \Sigma)$ along the boundary by

$$\mathcal{I} = \mu + 2.$$

The Keller-Maslov index in the case of a section of $\mathbb{L}$ is given by the sum i of the umbilic indices inside the curve ∂D in the boundary Σ, as viewed in $\mathbb{R}^3$ [37]:

$$\mu = 4i.$$

For the Keller-Maslov class to control the dimension of the space of holomorphic discs, one needs the dimension of the cokernel to be zero. If the problem is Fredholm regular, by a small perturbation the cokernel vanishes and the space of holomorphic discs is indeed determined by the number of enclosed umbilic points.

Remarkably, the Riemann-Hilbert problem associated with a convex sphere containing a single umbilic point *is* Fredholm regular:

Theorem 15. [30] *Let $\Sigma \subset \mathbb{L}$ be a Lagrangian sphere with a single isolated complex point. Then the Riemann-Hilbert problem with boundary Σ is Fredholm regular.*

The reason behind this result is that the Euclidean isometry group acts holomorphically and symplectically on $\mathbb{L}$, thus preserving the problem. The action is also transitive and so fixing the single complex point one considers the equivariant problem, the result being that it is Fredholm regular, as in the totally real case.

The non-existence of a convex sphere containing a single umbilic point is the famous conjecture of Constantin Carathéodory, and Theorem 15 gives the reason the Conjecture is true. Namely, were such a remarkable surface S to exist, the Riemann-Hilbert problem with boundary given by the normal lines Σ would be Fredholm regular and so have the property that the dimension of the space of parameterised holomorphic discs with boundary lying on it would be entirely determined by the number of umbilic points in the interior on S.

$$\mathcal{I} = \text{Dim Ker } \bar{\partial} = 4i + 2 \tag{12}$$

This property would also hold for a dense set of perturbations of S in an appropriate function space. To show that such a surface S cannot exist, one can seek to find violations of equation (12), in particular, a holomorphic disc which encloses a totally real disc on the boundary Σ.

By equation (12), if the boundary encloses a totally real disc, then $\mathcal{I} = 2$. However, since the Möbius group acts on the space of parameterized holomorphic discs, the space of unparameterized holomorphic discs is $2 - 3 = -1$. Thus, over an umbilic-free region of the remarkable surface S it should be impossible to solve the $\bar{\partial}$-problem.

The proof of the Carathéodory Conjecture in [35] follows from the existence of holomorphic discs with boundary enclosing umbilic-free regions, as established by evolving to them using mean curvature flow of a spacelike surface in $\mathbb{L}$, thus disproving equation (12).

At this point in time two thirds of the proof given in [35] has appeared in print, with the final part containing the boundary estimates for mean curvature flow currently under review.

In fact, the interior estimates required to prove long time existence and convergence hold for more general spacelike mean curvature flow with respect to indefinite metrics satisfying certain curvature conditions [32].

The final step of the proof of the Conjecture is the establishment of boundary estimates for mean curvature flow in $\mathbb{L}$ and sufficient control to show that the flow weakly converges in an appropriate function space to a holomorphic disc. The boundary conditions used for mean curvature flow (a second order system) include a constant angle condition and an asymptotic holomorphicity condition.

The constant angle condition is defined between a pair of spacelike planes that intersect along a line and is hyperbolic in nature. The asymptotic holomorphicity condition ensures that the ultimate disc is holomorphic rather than just maximal.

The sizes of the constant hyperbolic angle and the added holomorphic twist are free parameters in the evolution and can be used to control the flowing surface. If one views it as a codimension two capillary problem, the effect of the parameter changes is to increase the friction at the boundary, stopping it from skating off the hemisphere, thus preserving strict parabolicity.

An analogous result in the rotationally symmetric case for mean curvature flow in the space of oriented lines with Dirichlet and Neumann boundary conditions shows that the evolving surface can be made to converge to a holomorphic disc - in this case to a family of holomorphic discs called the Bishop family [6] - or to a maximal surface, depending on the boundary condition imposed [28].

For the full flow one can then show that:

Theorem 16. [35] *Let S be a $C^{3+\alpha}$ smooth oriented convex surface in $\mathbb{R}^3$ without umbilic points and suppose that the Gauss image of S contains a closed hemisphere. Let $\Sigma \subset \mathbb{L}$ be the oriented normal lines of S forming a Lagrangian surface in the space of oriented lines.*

Then $\exists f : D \to \mathbb{L}$ with $f \in C^{1+\alpha}_{loc}(D) \cap C^0(\overline{D})$ satisfying

(i) *f is holomorphic,*
(ii) *$f(\partial D) \subset \Sigma$.*

This would conclude the proof of the Carathéodory Conjecture for $C^{3+\alpha}$ smooth surfaces.

The appearance of Gauss hemispheres here is noteworthy, for this meets with a conjecture of Victor Toponogov that a complete convex plane must have an umbilic point, albeit at infinity [71]. Toponogov showed that such planes have hemispheres as Gauss image and established his conjecture under certain fall-off conditions at infinity.

In fact, the same reasoning as above that pits Fredholm regularity against mean curvature flow proves the Toponogov Conjecture:

Theorem 17. [31] *Every $C^{3+\alpha}$-smooth complete convex embedding of the plane P, satisfies* $\inf_P |\kappa_1 - \kappa_2| = 0$.

The proof follows from applying Theorem 16 in this case, while Fredholm regularity is established easily, as a putative counter-example is by assumption totally real (even at infinity).

Without the high degree of symmetry of the Euclidean group, one would not expect Fredholm regularity to hold and this obstructs the generalisation of the Carathéodory Conjecture to non-Euclidean ambient metrics. This turns out to be the case and the delicate nature of the problem is revealed:

Theorem 18. [26] *For all $\epsilon > 0$, there exists a smooth Riemannian metric g on $\mathbb{R}^3$ and a smooth strictly convex 2-sphere $S \subset \mathbb{R}^3$ such that*

(i) *S has a single umbilic point,*
(ii) $\|g - g_0\|^2 \leq \epsilon$,

where $\|.\|$ is the L_2 norm on $\mathbb{R}^3$ with respect to the flat metric g_0.

The proof here is constructive: the Euclidean metric is deformed while keeping the standard round 2-sphere fixed (although not round in the deformed metric) and one can essentially brush the principal foliation of the surface into any configuration one chooses by changing the ambient geometry.

Finally, establishing the local index bound $i(p) \leq 1$ for any isolated umbilic point p has long been the preferred route to proving the Carathéodory Conjecture in the real analytic case [38] [44]. The above methods can also be used to find a slightly weaker local index bound for isolated umbilics on smooth surfaces:

Theorem 19. [34] *The index of an isolated umbilic p on a $C^{3,\alpha}$ surface in $\mathbb{R}^3$ satisfies $i(p) < 2$.*

The proof follows from the extension of Theorem 15 to surfaces of higher genus by removing hyperbolic umbilic points and adding totally real cross-caps to the Lagrangian section. The existence of holomorphic discs over open hemispheres again contradicts Fredholm regularity and the local index bound follows.

Once again, the role of the Euclidean isometry group is paramount, and even a small perturbation of the ambient metric means that the index bound does not hold.

Theorem 20. [26] *For all $\epsilon > 0$ and $k \in \mathbb{Z}/2$, there exists a smooth Riemannian metric g on $\mathbb{R}^3$ and a smooth embedded surface $S \subset \mathbb{R}^3$ such that*

(i) *S has an isolated umbilic point of index k,*
(ii) $\|g - g_0\|^2 \leq \epsilon$,

where $\|.\|$ is the L_2 norm on $\mathbb{R}^3$ with respect to the flat metric g_0.

Finally, the local umbilic index bound $i(p) \leq 1$ of Hamburger [38] for real analytic surfaces has recently been used to prove results on the zeros of certain holomorphic polynomials. In particular, a polynomial whose zero set is invariant under inversion in the unit circle is called *self-inversive* [7] [47] [57] [66] [72].

Theorem 21. [29] *Let P_N be a polynomial of degree N with self-inversive second derivative and suppose that none of the roots of P_N lies on the unit circle. Then the number of roots (counted with multiplicity) of P_N inside the unit circle is less than or equal to $\lfloor N/2 \rfloor + 1$.*

This result is in the spirit of a converse to the Gauss-Lucas theorem [51] in which the zeros of the first derivative of a polynomial are restricted by the zeros of the polynomial. Here, however, by methods of differential geometry, the locations of the zeros of the second derivative restrict the zeros of the polynomial - the first such application. It is also worth noting that the result is sharp.

The method of proof is to take a polynomial with self-inversive second derivative and to construct a real analytic strictly convex with an isolated umbilic point whose index is determined by the number of zeros inside the unit circle.

4.2. **Four Manifold Topology.** The proof by Grigori Perelman of Thurston's Geometrization Conjecture [59][60][61] naturally raises the question as to whether closed 4-dimensional manifolds can be geometrized in some way. The approach in three dimensions, however, does not apply in higher dimensions and even basic things are harder.

For example, any finitely presented group can be the fundamental group of a smooth closed 4-manifold, while the fundamental group of a prime 3-manifold must be a quotient of the isometry group of one of the eight Thurston homogenous geometries [69], and so it is clear that new geometric paradigms are required.

To make matters worse, while in three dimensions there is no distinction between smooth, piecewise-linear and topological structures on closed manifolds, in higher dimension this may not be true. If one considers open manifolds, these problems are compounded further. In each dimension $n \geq 3$ there are uncountably many *fake* $\mathbb{R}^n$'s - open topological manifolds that are homotopy equivalent to, but not homeomorphic to $\mathbb{R}^n$ [12][24][54]. While many of these involve infinite constructions, an example of Barry Mazur in dimension four requires only the attachment of two thickened cells [53].

Four dimensions also has its share of peculiar problems that do not arise in higher dimensions. In particular, the Whitney trick, in which closed loops are contracted to a point across a given disc, plays a major role in many higher dimensional results, for example Stephen Smale's proof of the h-cobordism theorem [67]. The issue is that, while in dimensions five and greater a generic 2-disc is embedded, in dimension four a generic 2-disc is only immersed and will have self-intersections, making it unsuitable to contract loops across.

Against this array of formidable difficulties, the Disc Theorem of Micheal Freedman [16] utilizes a doubly infinite codimension two construction to claim that there is a topological work-around for the Whitney trick. This result leads to the proof of the topological Poincaré Conjecture in dimension four, as well as the classification of all simply connected closed topological 4-manifolds based almost entirely on their intersection form in the second homology.

Contradictions with Donaldson's ground-breaking work on smooth 4-manifolds [14] lead to extraordinary families of exotic manifolds (homeomorphic but not diffeomorphic) not seen in any other dimension. Since the work of John Milnor [55] it has been known that exotic differentiable structures in dimensions seven and above exist, but only in finitely many families. According to the Disc Theorem exotic differentiable structures in dimension four occur in uncountable families - indeed, no 4-manifold is known to have only countably many distinct differentiable structures.

Both the original Disk Theorem [16] and subsequent attempts to complete it [3] [17] [18] [27] [39] have depended upon the iterative attachment of 1- and 2-handles or there generalizations, as one attempts to push unwanted codimension two intersections to infinity. The ultimate homeomorphism that is sought is shown to exist using Bing shrinking and what is called decomposition space theory [5].

One key aspect of these efforts is that they all involve codimension two constructions - gluing in thickened 2-discs or more general surfaces into 4-manifolds. The work in this survey involves geometric paradigms associated with neutral metrics which can gain more control of these codimension two constructions.

Unlike Riemannian metrics which exist on all smooth manifolds, neutral metrics see the topology of the underlying manifold and can be used to express topological invariants. The next section considers closed 4-manifolds and illustrates the manner in which the existence of certain neutral metrics restricts the topology of the underlying 4-manifold. These are modest steps in the direction of understanding a tiny part of the wild world of 4-manifolds in which there is a splitting $4 = 2 + 2$.

4.3. **Closed Neutral 4-manifolds.** The simplest topological invariant of a closed 4-manifold M is its *Euler number* $\chi(M)$. Let $H_n(M,\mathbb{R})$ be the n^{th} homology group of M with real coefficients and b_n be the associated Betti numbers $n = 0, 1, ..., 4$. For a closed connected 4-manifold we have $b_0 = b_4 = 1$, and $b_3 = b_1$ by Poincaré duality and the Euler number is defined

$$\chi(M) = \sum_{n=0}^{4}(-1)^n \dim H_n(M,\mathbb{R}) = 2 - 2b_1 + b_2,$$

The Chern-Gauss-Bonnet Theorem states that one can express this geometrically as

$$\chi(M) = \frac{\epsilon}{32\pi^2}\int_M |W(g)|^2 - 2|Ric(g)|^2 + \tfrac{2}{3}S^2\ d^4V_g,$$

for any metric g of definite ($\epsilon = 1$) or neutral signature ($\epsilon = -1$) [49].

On a closed 4-manifold there is a natural symmetric bilinear pairing on the integral second homology $H_2(M,\mathbb{Z})$. It is the sum of the number of transverse intersection points between two surfaces representing the homology classes.

The intersection form can be diagonalised over $\mathbb{R}$ and the number of positive and negative eigenvalues is denoted b_+ and b_-, respectively. Thus $b_2 = b_+ + b_-$ and the *signature* $\tau(M) = b_+ - b_-$ is another topological invariant of M.

The existence of a neutral metric on a closed 4-manifold is equivalent to the existence of a field of oriented tangent 2-planes on the manifold [52]. Moreover:

Theorem 22. [41] [48] [52] *Let M be a closed 4-manifold admitting a neutral metric. Then*

$$\chi(M) + \tau(M) = 0 \ mod\ 4 \qquad \text{and} \qquad \chi(M) - \tau(M) = 0 \ mod\ 4. \tag{13}$$

If M is simply connected, these conditions are sufficient for the existence of a neutral metric.

Thus, neither $\mathbb{S}^4$ nor $\mathbb{C}P^2$ admit a neutral metric, while the K3 manifold does.

Given a neutral metric g' on M, the Euler number and signature can be expressed in terms of curvature invariants by

$$\chi(M) = \frac{-1}{32\pi^2}\int_M |W^+|^2 + |W^-|^2 - 2|Ric|^2 + \tfrac{2}{3}S^2 \; d^4V_g.$$

$$\tau(M) = b_+ - b_- = \frac{1}{48\pi^2}\int_M |W^+|^2 - |W^-|^2 \; d^4V_g.$$

where $W^\pm$ is the Weyl curvature tensor split into its self-dual and anti-self-dual parts, Ric is the Ricci tensor and S is the scalar curvature of g'.

From these and Theorem 7, the following can be proven

Theorem 23. [19] *Let (M, g') be a closed, conformally flat, scalar flat, neutral 4-manifold. If g' admits a parallel isometric paracomplex structure, then*

$$\tau(M) = 0 \qquad \text{and} \qquad \chi(M) \geq 0.$$

If, moreover, the Ricci tensor of g' has negative norm $|Ric(g')|^2 \leq 0$, then M admits a flat Riemannian metric.

On the other hand, Theorem 7 can also be used on Riemannian Einstein 4-manifolds to find obstructions to parallel isometric paracomplex structures:

Theorem 24. [19] *Let (M, g) be a closed Riemannian Einstein 4-manifold. If g admits a parallel isometric paracomplex structure, then $\tau(M) = 0$.*

The $K3$ 4-manifold, as well as the 4-manifolds $\mathbb{C}P^2 \# k\overline{\mathbb{C}P}^2$ for $k = 3, 5, 7$, admit Riemannian Einstein metrics and isometric almost paracomplex structures, but, as a consequence of Theorem 24, these almost paracomplex structures cannot be parallel.

Acknowledgements:

Most of the work described in this paper was carried out in collaboration with Guillem Cobos, Nikos Georgiou and Wilhelm Klingenberg, with whom it has been a pleasure to learn. Thanks are due to Morgan Robson for assistance with the Figures. Any opinions expressed are entirely the author's.

References

[1] D.V. Alekseevsky, B. Guilfoyle and W. Klingenberg, *On the geometry of spaces of oriented geodesics*, Ann. Global Anal. Geom. **40.4** (2011) 389–409. DOI: http://dx.doi.org/10.1007/s10455-011-9261-5 Erratum: Ann. Global Anal. Geom. **50.1** (2016) 97–99. DOI: https://doi.org/10.1007/s10455-016-9515-3

[2] L. Asgeirsson, *Über eine Mittelwertseigenschaft von Lösungen homogener linearer partieller Differentialgleichungen 2. Ordnung mit konstanten Koeffizienten*, Math. Ann. **113.1** (1937) 321–346. DOI: https://doi.org/10.1007/BF01571637

[3] S. Behrens, et al., The disc embedding theorem, Oxford University Press, 2021.

[4] C.A. Berenstein and E. Casadio Tarabusi, *Integral geometry in hyperbolic spaces and electrical impedance tomography*, SIAM J. Appl. Math. **56.3** (1996) 755–764. DOI: https://doi.org/10.1137/S0036139994277348

[5] R.H. Bing, *Upper semicontinuous dcompositions of E^3*, Annals of Math. (1957) 363–374. DOI: https://doi.org/10.2307/1969968

[6] E. Bishop, *Differentiable manifolds in complex Euclidean space*, Duke Math. J. **32.1** (1965) 1–21. DOI: https://doi.org/10.1215/S0012-7094-65-03201-1

[7] F.F. Bonsall and M. Marden, *Zeros of self-inversive polynomials*, Proc. Amer. Math. Soc. **3.3** (1952) 471–475. DOI: https://doi.org/10.2307/2031905

[8] G. Cobos and B. Guilfoyle, *An extension of Asgeirsson's mean value theorem for solutions of the ultrahyperbolic equation in dimension four*, Differential Geom. Appl. **79** (2021) 101795. DOI: https://doi.org/10.1016/j.difgeo.2021.101795

[9] G. Cobos and B. Guilfoyle, *A conformal mean value theorem for solutions of the ultrahyperbolic equation*, (2022) ArXiv: https://arxiv.org/abs/2210.08155

[10] R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol II. John Wiley and Sons, 2008.

[11] A.M. Cormack, *Representation of a function by its line integrals, with some radiological applications*, Journal of Applied Physics **34.9** (1963) 2722–2727. DOI: https://doi.org/10.1063/1.1729798

[12] M.L. Curtis and K.W. Kwun, *Infinite sums of manifolds*, Topology **3.1** (1965) 31–42. DOI: https://doi.org/10.1016/0040-9383(65)90068-6

[13] J. Davidov, J. G. Grantcharov and O. Mushkarov, *Geometry of neutral metrics in dimension four*, (2008) ArXiv https://arxiv.org/abs/0804.2132

[14] S.K. Donaldson, *An application of gauge theory to 4-dimensional topology*, J. Differential Geom. **18.2** (1983) 279–315. DOI: https://doi.org/10.1215/S0012-7094-96-08316-7

[15] M. Dunajski and S. West, *Anti-self-dual conformal structures in neutral signature*, in Recent Developments in Pseudo-Riemannian Geometry, Editors D.V. Alekseevsky and H. Baum, ESI Lect. Math. Phys., Eur. Math. Soc., Zürich (2008) 113–148.

[16] M.H. Freedman, it The topology of 4-dimensional manifolds, J. Differential Geom. **17.3** (1982): 357-453. DOI: https://doi.org/10.4310/jdg/1214437136

[17] M. H. Freedman, *The disk theorem for four-dimensional*, Proceedings of the International Congress 1 (1983) 647–663.

[18] M.H. Freedman and F. Quinn, Topology of 4-manifolds, (PMS-39) Vol. 46., Princeton University Press, 2014.

[19] N. Georgiou and B. Guilfoyle, *Almost paracomplex structures on 4-manifolds*, Differential Geom. Appl. **82** (2022) 101890. DOI: https://doi.org/10.1016/j.difgeo.2022.101890

[20] N. Georgiou and B. Guilfoyle, *A new geometric structure on tangent bundles*, J. Geom. Phys. **172** (2022) 104415. DOI: https://doi.org/10.1016/j.geomphys.2021.104415

[21] N. Georgiou and B. Guilfoyle, *The causal topology of neutral 4-manifolds with null boundary*, New York J. Math. **27** (2021) 477–507. http://nyjm.albany.edu/j/2021/27-20.html

[22] N. Georgiou and B. Guilfoyle, *A characterization of Weingarten surfaces in hyperbolic 3-space*, Abh. Math. Sem. Univ. Hambg. **80.2** (2010) 233–253. DOI: https://doi.org/10.1007/s12188-010-0039-7

[23] N. Georgiou and B. Guilfoyle, *On the space of oriented geodesics of hyperbolic 3-space*, Rocky Mountain J. Math. **40.4** (2010) 1183–1219. DOI: https://www.jstor.org/stable/44239772

[24] L.C. Glaser, *Uncountably many contractible open 4-manifolds*, Topology **6.1** (1967) 37–42. DOI: https://doi.org/10.1016/0040-9383(67)90011-0

[25] M. Gromov, *Pseudo holomorphic curves in symplectic manifolds*, Invent. Math. **82.2** (1985) 307–347. DOI: https://doi.org/10.1007/BF01388806

[26] B. Guilfoyle, *On isolated umbilic points*, Comm. Anal. Geom. **28.8** (2020) 2005–2018. DOI: https://doi.org/10.4310/CAG.2020.v28.n8.a8

[27] B. Guilfoyle, *Why is the 4-dimensional smooth Poincaré Conjecture still open?* University College Dublin Distinguished Visitor Series (2019) Lecture 4 https://www.youtube.com/watch?v=VZs1UG2Wtn8

[28] B. Guilfoyle and W. Klingenberg, *Parabolic evolution with boundary to the Bishop family of holomorphic discs*, (2023) ArXiv: https://arxiv.org/abs/2304.05702.

[29] B. Guilfoyle and W. Klingenberg, *Roots of polynomials and umbilics of surfaces*, Results in Math. (to appear) ArXiv: https://arxiv.org/abs/2204.13163

[30] B. Guilfoyle and W. Klingenberg, *Fredholm-regularity of holomorphic discs in plane bundles over compact surfaces*, Ann. Fac. Sci. Toulouse Math. Série 6, **29.3** (2020) 565–576. DOI: https://doi.org/10.5802/afst.1639

[31] B. Guilfoyle and W. Klingenberg, *Proof of the Toponogov Conjecture on complete surfaces*, (2020) ArXiv: https://arxiv.org/abs/2002.12787

[32] B. Guilfoyle and W. Klingenberg, *Higher codimensional mean curvature flow of compact spacelike submanifolds*, Trans. Amer. Math. Soc. **372.9** (2019) 6263–6281. DOI: https://doi.org/10.1090/tran/7766

[33] B. Guilfoyle and W. Klingenberg, *A global version of a classical result of Joachimsthal*, Houston J. Math. **45.2** (2019) 455–467. ArXiv https://arxiv.org/abs/1404.5509

[34] B. Guilfoyle and W. Klingenberg, *From global to local: an index bound for umbilic points on smooth convex surfaces*, (2012) IHES preprint M-12-18 https://arxiv.org/abs/1207.5994

[35] B. Guilfoyle and W. Klingenberg, *Proof of the Carathéodory Conjecture*, (2011) ArXiv https://arxiv.org/abs/0808.0851

[36] B. Guilfoyle and W. Klingenberg, *An indefinite Kähler metric on the space of oriented lines*, J. London Math. Soc. **72.2** (2005) 497–509. DOI:https://doi.org/10.1112/S0024610705006605
[37] B. Guilfoyle and W. Klingenberg, *Generalised surfaces in* $\mathbb{R}^3$, Math. Proc. R. Ir. Acad. **104A(2)** (2004) 199–209. ArXiv: https://arxiv.org/abs/math/0406185
[38] H. Hamburger, *Beweis einer Carathéodoryschen Vermutung I, II and III*, Ann. Math. **41** (1940) 63–86, Acta. Math. **73** (1941) 175–228, Acta. Math. **73** (1941) 229–332. DOI: https://doi.org/10.2307/1968821 DOI: https://doi.org/10.1007/BF02392230
[39] K. Hartnett, *New Math Book Rescues Landmark Topology Proof*, Quanta Magazine, September 9 2021. https://www.quantamagazine.org/new-math-book-rescues-landmark-topology-proof-20210909/
[40] D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, American Mathematical Soc., 1999.
[41] F. Hirzebruch and H. Hopf, *Felder von Flächenelementen in 4-dimensionalen Mannigfaltigkeiten*, Math. Ann. **136** (1958) 156–172. DOI: https://doi.org/10.1007/BF01362296
[42] N.J. Hitchin, *Monopoles and geodesics*, Comm. Math. Physics **83.4** (1982) 579–602. DOI: https://doi.org/10.1007/BF01208717
[43] G.N. Hounsfield,*Computerized transverse axial scanning (tomography): Part 1. Description of system*, Brit. J. Radiol. **46.552** (1973) 1016–1022. DOI: https://doi.org/10.1259/0007-1285-46-552-1016
[44] V. V. Ivanov, *An analytic conjecture of Carathéodory*, Siberian Math. J. **43** (2002) 251–322. DOI: https://doi.org/10.1023/A:1014797105633
[45] F. Joachimsthal, *Demonstrationes theorematum ad superficies curvas spectantium*, J. Reine Angew. Math. **30** (1846) 347–350. DOI: https://doi.org/10.1515/crll.1846.30.347
[46] F. John, *The ultrahyperbolic differential equation with four independent variables*, Duke Math. J. **4.2** (1938) 300–322. DOI: https://doi.org/10.1007/978-1-4612-5406-5_8
[47] D. Joyner and T. Shaska, *Self-inversive polynomials, curves, and codes*, Higher genus curves in mathematical physics and arithmetic geometry **703** (2018) 189–208.
[48] H. Kamada, *Self-dual Kähler metrics of neutral signature on complex surfaces*, Tohoku Math. Publ. **24** (2002) 1–94. DOI: https://doi.org/10.2748/tmpub.24.1
[49] P. Law, *Neutral Einstein metrics in four dimensions*, J. Math. Phys. **32.11** (1991) 3031–3042. DOI: https://doi.org/10.1063/1.529048
[50] J.M. Lee and T.H. Parker, *The Yamabe problem*, Bull. Amer. Math. Soc. (N.S.) **17.1** (1987) 37–91. DOI: https://doi.org/10.1090/S0273-0979-1987-15514-5
[51] M. Marden, Geometry of Polynomials, Mathematical Surveys and Monographs Vol. 3, American Mathematical Society, Providence, RI (1966).
[52] Y. Matsushita, *Fields of 2-planes and two kinds of almost complex structures on compact 4-dimensional manifolds*, Math. Z. **207.1** (1991) 281–291. DOI: https://doi.org/10.1007/BF02571388
[53] B. Mazur, *A note on some contractible 4-manifolds*, Ann. of Math. **73.1** (1961) 221–228. DOI: https://doi.org/10.2307/1970288
[54] D.R. McMillan, *Some contractible open 3-manifolds*, Trans. Amer. Math. Soc. **102.2** (1962) 373–382. DOI: https://doi.org/10.2307/1993684
[55] J. Milnor, *Differentiable structures on spheres*, Amer. J. Math. **81.4** (1959) 962–972. DOI: https://doi.org/10.2307/2372998
[56] A. Mouton, G.T. Flitton, S. Bizot, N. Megherbi, and T.P. Breckon, *An evaluation of image denoising techniques applied to CT baggage screening imagery*, In 2013 IEEE International Conference on Industrial Technology (ICIT) (2013, February) 1063-1068. DOI: https://doi.org/10.1109/ICIT.2013.6505819.
[57] P.J. O'Hara and R.S. Rodriguez, *Some properties of self-inversive polynomials*, Proc. Amer. Math. Soc. **44.2** (1974) 331–335. DOI: https://doi.org/10.1090/S0002-9939-1974-0349967-5
[58] E.M. Payne, *Imaging Techniques in Conservation*, Journal of Conservation and Museum Studies, **10.2** (2012) 17—29. DOI: https://doi.org/10.5334/jcms.1021201
[59] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, (2002). ArXiv Preprint https://arxiv.org/abs/math/0211159
[60] G. Perelman, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, (2003). ArXiv Preprint https://arxiv.org/abs/math/0307245
[61] G. Perelman, *Ricci flow with surgery on three-manifolds*, (2003). ArXiv Preprint https://arxiv.org/abs/math/0303109

[62] J. Radon, *Über die Bestimmung von Funktionen durch ihre Integralwerte langs gewisser Mannigfaltigkeiten, Berichte Verhandlunger Gesellschaft Wissenschaft*, Math.-Phys. Klasse **69** (1917) 262–271.

[63] M. Russell, *How Next Generation Airport Scanners Are Ending The 100ml Liquid Rule*, Simple Flying (30 March 2022). Retrieved 11 April 2022. https://simpleflying.com/airport-scanners-ending-100ml-liquid-rule/

[64] M. Salvai, *On the geometry of the space of oriented lines of Euclidean space*, Manuscripta Math. **118.2** (2005) 181–189. DOI: https://doi.org/10.1007/s00229-005-0576-z

[65] M. Salvai, *On the geometry of the space of oriented lines of the hyperbolic space*, Glasg. Math. J. **49.2** (2007) 357–366. DOI: https://doi.org/10.1017/S0017089507003710

[66] T. Sheil-Small, Complex polynomials, Studies In Advanced Mathematics **73**, Cambridge University Press, Cambridge, 2002.

[67] S. Smale, *On the structure of manifolds*, Amer. J. Math. **84.3** (1962) 387—399. DOI: https://doi.org/10.2307/2372978

[68] E. Study, *Von den Bewegungen und Umlegungen: I. und II. Abhandlung*, Math. Annalen **39.4** (1891) 441–565. DOI: https://doi.org/10.1007/BF01199824

[69] W.P. Thurston, The geometry and topology of 3-manifolds, Princeton Lecture Notes (1979), available at http://www.msri.org/publications/books/gt3m/

[70] S. Tonai, Y. Kubo, M.-Y. Tsang, S. Bowden, K. Ide, T. Hirose, N. Kamiya, Y. Yamamoto, K. Yang, Y. Yamada and Y. Morono, *A New Method for Quality Control of Geological Cores by X-Ray Computed Tomography: Application in IODP Expedition 370*, Frontiers in Earth Science, **7** (2019) 117. DOI: https://doi.org/10.3389/feart.2019.00117

[71] V.A. Toponogov, *On conditions for existence of umbilical points on a convex surface*, Siberian Mathematical Journal, **36** (1995) 780–784. DOI: https://doi.org/10.1007/BF02107335

[72] R.S. Vieira, *On the number of roots of self-inversive polynomials on the complex unit circle*, Ramanujan J. **42.2** (2017) 363–369. DOI: https://doi.org/10.1007/s11139-016-9804-2

[73] K. Weierstrass, *Mathematics of Surfaces III*, Monatsberichte der Berliner Akademie **371** (1866) 612–625.

[74] Y. Zhang, W. Verwaal, M.F.C. Van de Ven, A.A.A. Molenaar, and S.P. Wu, *Using high-resolution industrial CT scan to detect the distribution of rejuvenation products in porous asphalt concrete*, Construction and Building Materials **100** (2015) 1–10. DOI: https://doi.org/10.1016/j.conbuildmat.2015.09.064

Brendan Guilfoyle is Professor of Mathematics in the Munster Technological University. He obtained an MSc. from Trinity College Dublin in general relativity and a doctorate from University of Texas at Austin under the supervision of Prof. Karen Uhlenbeck. His mathematical research interests include geometry, topology and analysis in low dimensions that arise in physical problems.

Brendan Guilfoyle, School of Science Technology Engineering and Mathematics, Munster Technological University, Kerry, Tralee, Co. Kerry, Ireland.

E-mail address: brendan.guilfoyle@mtu.ie

Irish Math. Soc. Bulletin
Number 91, Summer 2023, 33–35
ISSN 0791-5578

Trivial Centre Group Orders

DES MACHALE

In memoriam Rex Dark

Abstract. We discuss the possible orders of finite groups that have trivial centre.

1. Introduction

The centre $Z(G)$ of a group G is defined to be $\{z \in G | zx = xz, \text{ for all } x \in G\}$. $Z(G)$ is a characteristic subgroup of G, which of course contains the identity element 1 of G. A group in which $Z(G) = \{1\}$ is said to have trivial centre. We discuss the question:

> For which natural numbers n does there exist a group G with $|G| = n$ and G has trivial centre?

This is sequence A060702 in Sloane [1] and begins 1, 6, 10, 12, 14, 18, 20, 21, 22, 24, 26, 30, 34, 36, 38, 39, 42, 46, 48, 50, 52, 54, 55, 56, 57, 58, 60, 62, 66, 68, 70, 72, 74, 75, 78, ...

We call these numbers the trivial centre group orders (TZ-numbers). The determination of all TZ-numbers seems to be a difficult problem, possibly out of reach at the moment, but we can list many classes of numbers which belong to this set.

(1) For each n, $4n + 2 = 2(2n + 1)$ is a TZ-number. This is because the dihedral group D_{2n+1} of order $2(2n + 1)$ has trivial centre.
(2) If p is an odd prime and $q > p$ is a prime such that p divides $q - 1$, then pq is a TZ-number. This is because under these conditions, there exists a unique group of order pq with trivial centre. This sequence begins 21, 39, 55, 57, 93, 111, 129, 183, 201, 203, 205, 219, ...
(3) Let p be a prime such that $p \equiv 1 \pmod 4$; then there are five isomorphism classes of groups of order $4p$. Two of those are abelian; there is D_{2p}, dihedral, and Q_p, dicyclic, given by $\langle a, b | a^{2p} = 1; b^2 = a^p, b^{-1}ab = a^{-1}\rangle$. All of these have non-trivial centre. But there is a fifth isomorphism class of groups, the semi-direct product of a cyclic group of order p by its unique cyclic subgroup of order 4 in its automorphism group. This group has trivial centre. Thus if $p \equiv 1 \pmod 4$ then $4p$ is a TZ-number. This sequence begins 20, 52, 68, 116, 148, 164, 212, ... ([1] A350115)
(4) Except for some small values of n, $n!$ and $n!/2$ are TZ-numbers because the symmetric group S_n and the alternating group A_n both have trivial centre.
(5) The simple non-abelian orders are clearly TZ-numbers. This sequence begins 60, 168, 360, 504, 660, 1092, 2448, 2520, 3420, 4080, ... ([1] A001034).
(6) We remark that the product of two TZ-numbers is also a TZ-number. This is because for direct products, $Z(G_1 \times G_2) = Z(G_1) \times Z(G_2)$. But not every

2020 *Mathematics Subject Classification.* 20D99.
Key words and phrases. Finite Group, Trivial Centre.
Received on 8-3-2023.
DOI:10.33232/BIMS.0091.33.35.

multiple of a TZ-number is a TZ-number. For example, 14 is a TZ-number but 28 is not.

(7) A perfect group G is a group which satisfies $G' = G$. Surprisingly, not all perfect groups have trivial centre. An example is $\mathrm{SL}(2, 5)$. However, if G is perfect, it is known that $G/Z(G)$ has trivial centre (Grün's lemma).

(8) A complete group G is a group with trivial centre in which every automorphism is inner.
The sequence of complete orders begins 1, 6, 20, 24, 42, 54, 110, 120, 144, 156, 168, 216, 252, 272, 320, ... ([1] A341298).
In 1975, Rex Dark discovered a non-trivial complete group G which had odd order. It had order $33,209,467,522,096,377 = 3 \cdot 19 \cdot 17^{12}$ [3].
More recently, he showed that the smallest possible non-trivial complete group of odd order has order $352,947 = 3 \cdot 7^6$.
We can also list several classes of numbers which are *not* TZ-numbers.

(9) These include the cyclic orders, the abelian orders and more generally the nilpotent orders, which include the primes and the prime powers. We recall [2] that n is a nilpotent number, i.e. every group of order n is nilpotent, if n is of the form $p_1^{a_1} p_2^{a_2} \ldots p_t^{a_t}$, p_i distinct and $p_i^k \not\equiv 1 \pmod{p_j}$ for all integers i, j and k, with $1 \leq k \leq a_i$. This sequence begins 1, 2, 3, 4, 5, 7, 8, 9, 11, 13, 15, 16, 17, 19, 23, 25, 27, 29, 31, 32, 33, 35, ... ([1] A056867).
Since a finite nilpotent group has non-trivial centre, none of the terms of this sequence, except the first, is a TZ-number.

(10) Let p be a prime such that $p \equiv 3 \pmod 4$, $p > 3$. Then there are precisely four isomorphism classes of groups of order $4p$ – two abelian, D_{2p}, dihedral, and Q_p, dicyclic. All of these groups have non-trivial centre. Thus $4p$ is never a TZ-number when $p \equiv 3 \pmod 4$. This sequence begins 28, 44, 76, 92, 124, 172, 188, ...

Some time ago, the author made the following

Conjecture. *If $n = 6t$, for some natural number t, then n is a TZ-number.*

I made some progress with this conjecture, for which the numerical evidence is overwhelming, but could not finish it. Then I discussed it over coffee with Rex Dark at a conference and this is the proof we came up with:

Theorem. *If n is of the form $6t$, for some natural number t, then n is a TZ-number.*

Proof. Let $n = 2^k \cdot 3 \cdot m$, with $k \geq 1$ and m odd. We show there is a group G with $|G| = n$ and $Z(G) = \{1\}$.

Case 1: Suppose first that k is odd, say $k = 2r + 1$.

Take $H = S_3$, $K = K_1 \times K_2 \times \ldots \times K_r$, with $K_i \simeq C_2 \times C_2$ and $L = C_m$.

Then $\mathrm{Aut}(C_2 \times C_2) \simeq S_3$, so each of the groups K_i can be regarded as a faithful H module. We can also make S_3 act on L by taking A_3 to centralise L and $S_3/A_3 \simeq C_2$ to invert L elementwise. Thus H acts on K_1, K_2, ... , K_r and on L, and we form the corresponding semidirect product $G = H \cdot (K \times L)$. Clearly, $|G| = 6 \cdot 4^r \cdot m = n$ and $Z(G) = \{1\}$. We note that this construction still works when $r = 0$ (so $K = \{1\}$) and/or when $m = 1$ (so $L = \{1\}$).

Case 2: Next suppose that k is even, say $k = 2r$, $r \geq 1$, and $m = 1$.

Take $H = C_3$, $K = K_1 \times K_2 \times \ldots \times K_r$, with $K_i \simeq C_2 \times C_2$.

Then $C_3 \subseteq \mathrm{Aut}(C_2 \times C_2)$, so each of the groups K_i can be regarded as a faithful H module, and we form the corresponding semidirect product $G = HK$. Then $|G| = 3 \cdot 4^r = n$ and $Z(G) = \{1\}$.

Case 3: Finally, suppose that $k = 2r$ (with $r \geq 1$) and $m > 1$.

As in Case 1, we can construct a group G_1 with $|G_1| = 2^{2r-1} \cdot 3$ and $Z(G_1) = \{1\}$. We also take G_2 to be dihedral of order $2m$ and we form $G = G_1 \times G_2$. Then $|G| = 2^{2r-1} \cdot 3 \cdot 2m = n$ and $Z(G) = Z(G_1) \times Z(G_2) = \{1\}$, and we are done. □

2. Questions

Apart from a complete description of TZ-numbers, several other questions remain. Among these are:

Q1: What is the density of TZ-numbers? Our remarks (1) to (10) and our theorem could possibly throw some light on this question. Actual numbers in blocks of 100 less than 2000 appear to indicate that a figure hovering around 49.5% of natural numbers are TZ-numbers.
However, the preponderance of p-groups would seem to indicate that the proportion of *groups* with trivial centre is very small.

Q2: The consecutive numbers 20, 21, 22; 54, 55, 56, 57, 58; and 200, 201, 202, 203, 204, 205 are all TZ-numbers. We ask if there exist arbitrarily long sequences of this type.

Q3: Are there any other positive integers k (not a multiple of 6) for which kn is always a TZ-number? Clearly, k cannot be 10, 14, 20, 21 or 22.

References

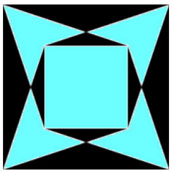
[1] N.A.J. Sloane: *The Online Encyclopaedia of integer sequences.* `https://oeis.org`

[2] Jonathan Pakianathan and Krishnan Shankar: *Nilpotent Numbers*, American Mathematical Monthly, 107, August-September 2000, 631-634.

[3] Rex Dark: *A Complete Group of odd order*, Math. Proc. Cam. Phil. Soc, Vol 77, January 1975, 21-28.

Des MacHale is Emeritus Professor of Mathematics at University College Cork where he taught for forty years. His main interests are in finite groups and rings, but he also dabbles in Number Theory, Euclidean Geometry, Trigonometric Inequalities, Combinatorial Geometry and Problem Posing and Solving. He has written several biographical books on George Boole but some would say his magnum opus is *Comic Sections Plus, the Book of Mathematical Jokes, Humour, Wit and Wisdom*, cf. `https://www.logicpress.ie/authors/machale`.

School of Mathematical Sciences, University College Cork
E-mail address: `d.machale@ucc.ie`



Irish Math. Soc. Bulletin
Number 91, Summer 2023, 37–47
ISSN 0791-5578

Segre's theorem on ovals in Desarguesian projective planes

PATRICK J. BROWNE, STEVEN T. DOUGHERTY AND PADRAIG Ó CATHÁIN

Abstract. Segre's theorem on ovals in projective spaces is an ingenious result from the mid-twentieth century which requires surprisingly little background to prove. This note, suitable for undergraduates with experience of linear and abstract algebra, provides a complete and self-contained proof. All necessary pre-requisites, principally evaluation of homogeneous polynomials at projective points and Desargues' theorem are presented in full. While following the broad outline of Segre's proof, careful parameterisation of certain tangent lines results in shorter and simpler computations than the original.

One of the most significant advances in the history of mathematics was the discovery in the 17th century, principally by Descartes, that geometry could be understood in algebraic terms. For example, a circle is defined geometrically as the set of points equidistant from a given point. Algebraically, this could be understood as the set of x and y satisfying $(x-a)^2 + (y-b)^2 = r^2$. Within this framework lines and conic sections could be described in an algebraic manner. The power of this connection was that one could maintain geometric intuition but have the power of algebra to construct rigorous proofs. It is no exaggeration to say that this discovery led to the development of calculus, differential equations, linear algebra and most of modern mathematics. In this paper, we shall investigate a fascinating connection between a geometric object and an algebraic description, this time in a finite projective space.

The study of projective geometry has its roots in the attempts of renaissance artists to accurately depict three dimensional scenes on a two-dimensional canvas. In the eighteenth century, it was realised that the mathematical study of geometry is rather easier in projective spaces than in Euclidean spaces, and projective geometries remain central objects in modern mathematics. In this note we consider only the lowest dimensional projective spaces, which are projective planes. Our purpose is to prove a theorem of Segre identifying certain combinatorial configurations with the set of points at which a suitable polynomial vanishes, [12]. The most important step in the proof is the celebrated Lemma of the Tangents, for which several 'co-ordinate free' proofs are available, [4, 2]. Our purpose is to present an accessible account, broadly similar to Segre's but with all necessary background material and improving on certain technicalities in the original proof.

Projective planes may be defined axiomatically as follows.

Definition 0.1. A projective plane consists of *points*, *lines* and an *incidence relation* relating points and lines which obey the following axioms:

(1) There is a unique line incident with any two distinct points.
(2) Any two distinct lines are incident with a unique point.
(3) There exist four points, no three incident with a line.

2020 *Mathematics Subject Classification.* 51E21.
Key words and phrases. Segre, Oval.
Received on 12-1-2023; revised 14-4-2023.
DOI:10.33232/BIMS.0091.37.47.

The third axiom is necessary so that certain degenerate geometric objects are not planes (e.g. where all points lie on a single line). A projective plane may be constructed from a two-dimensional vector space by 'completing' the space with a number of *points at infinity*[1]. It is more convenient mathematically to begin with a three dimensional vector space. Define *projective points* to be one dimensional subspaces and *projective lines* to be two dimensional subspaces, with incidence given by containment[2].

Verifying the axioms for a projective plane requires only elementary linear algebra:

(1) Two distinct one-dimensional subspaces span a unique two-dimensional space.
(2) Two distinct two-dimensional subspaces must intersect in a one-dimensional space (because both spaces live in a three dimensional space - this claim would not hold if we began with a vector space of dimension ≥ 4).
(3) There exist four lines, any three of which span the space, consider for example the lines spanned by

$$(1,0,0), \quad (0,1,0), \quad (0,0,1), \quad (1,1,1)\,,$$

with respect to an arbitrary basis.

While a projective plane is constructed from a three dimensional vector space, it is really a two-dimensional object: the points of the form $[1 : y : z]$ clearly form a two dimensional plane, while the remaining points $[0 : 1 : z]$ and $[0 : 0 : 1]$ are often called *points at infinity.* Each parallel class of lines meets at a point at infinity, and all points at infinity are collinear. By having these two distinct descriptions of the same space, we are able to use whichever one is easier to construct a proof of a given result. Vanishing points in perspective drawing may be considered points at infinity, and (at least with one eye closed) we perceive the real projective plane visually, as opposed to three dimensional Euclidean space.

Particular interest pertains to the projective planes constructed over finite fields, in which case the number of points and lines in the plane is finite. There exists a finite field, unique up to isomorphism, of any prime power order q, which we denote $\mathbb{F}_q$, [9]. (For the less experienced reader, the field of prime order p is precisely the integers modulo p. Nothing is lost by considering this case throughout the paper.) It is an easy exercise to see that the number of one- and two-dimensional subspaces of a three-dimensional vector space over $\mathbb{F}_q$ is $q^2 + q + 1$, while the number of one-dimensional subspaces in a two-dimensional space is $q + 1$. We conclude that a finite projective plane constructed from a vector space necessarily has $q^2 + q + 1$ points and an equal number of lines. Additionally, there are $q + 1$ points incident with any line, and $q + 1$ lines incident with any point. Projective planes have been considered by mathematicians of the highest calibre: Hilbert and Artin both wrote undergraduate-accessible accounts of the foundations of geometry with a particular emphasis on projective planes, [7, 1]. Finite projective planes are a more specialised topic, to which monographs have nevertheless been devoted. Hughes and Piper, and one of the authors have written at advanced undergraduate level, while Dembowski's work is more demanding, [8, 6, 5].

We typically denote the line $\{(at, bt, ct) : t \in \mathbb{F}_q\}$ by the *projective (or homogeneous) coordinates* $[a : b : c]$. Sometimes it is convenient to normalise projective coordinates so that the first non-zero entry is 1: provided a is non-zero the projective points $[a : b : c]$ and $[1 : a^{-1}b : a^{-1}c]$ are equal.

[1]There is nothing infinite about these points in a finite plane. In the infinite projective plane formed from a Euclidean plane, such points would seem to be located at an infinite distance from every other point.

[2]The reader should be aware that the projective dimension is typically one less than the standard vector-space dimension. In the remainder of this note, dimensions are projective.

Definition 0.2. A projective plane is *Desarguesian* if it is constructed from a three dimensional vector space over a field (or more generally, a division algebra). Equivalently, if it admits projective co-ordinates in a field (or division ring). The projective plane over the field $\mathbb{F}$ is denoted $\mathrm{PG}_2(\mathbb{F})$. Otherwise, it is *non-Desarguesian.*

In this note we consider only Desarguesian projective planes over finite fields. The Desarguesian plane over a field of order q is said also to be of order q, though note that there are $q+1$ points on each line. There do exist finite projective planes that are non-Desarguesian. The smallest of these has $91 = 9^2+9+1$ points. Some planes of this type may be constructed from so-called *quasi-fields*, but others seem to have no discernible algebraic structure. It is one of the foundational questions of finite projective planes to determine for which orders non-Desarguesian planes exist. It is known that there is no such plane containing $43 = 6^2+6+1$ points or $111 = 10^2+10+1$ points, but existence is open for a plane on $157 = 12^2+12+1$ points, and for infinitely many larger values.

1. Conics and ovals in a projective plane

Because a projective point corresponds to many distinct points in the underlying vector space, it does not make sense to evaluate a polynomial at a projective point, but it **does** make sense to ask whether a homogeneous polynomial is zero or non-zero at a projective point, as

$$F(\lambda x, \lambda y, \lambda z) = \lambda^k F(x, y, z)$$

for a homogeneous polynomial of degree k. The locus of points at which a homogeneous polynomial in three variables evaluates to zero on a Desarguesian projective plane is called the *variety* of the polynomial. We make no attempt to develop the theory of algebraic varieties, the interested reader is referred to Shafarevich, Chapter 1 [13].

(1) A homogeneous polynomial of degree 1 describes a line in a projective plane. For example, the equation $y = 0$ describes the projective points $[1 : 0 : z]$ with $z \in \mathbb{F}_q$ together with the point $[0 : 0 : 1]$. Setting $x+2y-z = 0$ gives the points $[x : y : x+2y]$ where $x, y \in \mathbb{F}_q$. While it may not appear that this set is one dimensional, working projectively and setting $z = \frac{y}{x}$ gives the set $[1 : z : 1+2z]$ where $z \in \mathbb{F}_q$ together with the point $[0 : 1 : 2]$.

(2) Over the real field homogeneous polynomials of degree 2 describe circles, ellipses and hyperbolae. Arguably the greatest achievement of ancient Greek geometry was the unified treatment of these different varieties by considering the intersection of a plane and a cone in three dimensional space, leading to the name *conic sections* for homogeneous polynomials of degree 2.

    In analogy with the real case, we call a homogeneous polynomial of degree two over **any** field a *conic*. We shall see shortly that the theory of conics over a finite field is quite different from the case of the real field. The points of the conic $x^2+y^2+z^2$ correspond to projective points $[x : y : \sqrt{-x^2-y^2}]$ which is an empty set over the real field. On the other hand, it can be verified that $4^2 = -(1^2+2^2) \mod 7$ so this variety is non-empty over $\mathbb{F}_7$. We shall see shortly via purely geometric arguments that a non-degenerate conic section over a finite field of odd order q will always have $q+1$ points.

It will frequently be helpful to think of a variety as the *graph* of a function on a two-dimensional plane. This is achieved in the following way: one breaks the analysis into points in the 2-dimensional plane $[1 : y : z]$, in which the homogeneous equation $F(x, y, z) = 0$ can be rewritten as $y = f(z)$, and then one considers the points at infinity separately. (Of course, normalising as $[x : y : 1]$ would move a different line to infinity and give a different view of the same variety.)

Recall that a tangent to a conic is a line meeting the curve at a single point. The tangent to a curve is routinely found by linearising the curve (which is achieved over $\mathbb{R}$ by computing a normal vector using partial derivatives and taking the orthogonal complement). To a surprising extent, the geometric intuition over $\mathbb{R}$ which is the foundation for calculus carries through to finite fields, though there is no longer any convergence (and so $\epsilon - \delta$ arguments no longer hold). In this note we take formal derivatives over finite fields. The ordinary formulae continue to hold, though they require an entirely different justification, which would lead us too far from the topic of the paper. We justify this by claiming that these methods are provided for illustration, and are not required in the proof of the main theorem. In any case, we define a *tangent* to a variety over a finite field to be a line meeting the variety in a unique point.

Proposition 1.1. *Let $F(x, y, z)$ be a homogeneous function, and $\mathcal{V}$ its variety in projective space. The tangent space to F at $v \in \mathcal{V}$ is the orthogonal complement to the normal vector $\nabla F = [\frac{\partial F}{\partial x} : \frac{\partial F}{\partial y} : \frac{\partial F}{\partial z}]$ evaluated at v (with respect to the standard inner product).*

This is perhaps best illustrated by an example.

Example 1.2. Consider the conic $F(x, y, z) = x^2 - yz$ over a field with at least 3 elements. Normalising the z co-ordinate shows that this is just the standard quadratic function $y = x^2$, completed by the point $[0 : 1 : 0]$ at infinity.

The normal vector to F is given by $\nabla F = [\frac{\partial F}{\partial x} : \frac{\partial F}{\partial y} : \frac{\partial F}{\partial z}] = [2x : -z : -y]$, and the tangent line at a point is the orthogonal complement of this vector.

For example, the point $p = [1 : 1 : 1]$ is in the variety of F by inspection. The normal vector at this point is $[2 : -1 : -1]$ which is orthogonal to, for example, $[0 : 1 : -1]$. Thus a parametrisation of the tangent line at p is given by $T_p = p + t[0 : 1 : -1] = [1 : 1 + t : 1 - t]$, where t is a parameter taking values in $\mathbb{F}_q$ together with the point $[0 : 1 : -1]$. It is easily verified that T_p is tangent to the variety: substituting a generic point on the line into $F(x, y, z)$ gives the equation

$$1 - (1 + t)(1 - t) = t^2 = 0$$

which has a unique solution. Hence, T_p intersects the variety only at p.

The choice of vector orthogonal to $\nabla F(x, y, z)$ is far from unique. Nevertheless, the resulting tangent line is unique (provided the defining equation satisfies technical conditions which will always be met in this paper), though the parameterisation may not be. Considering the tangent line as a two-dimensional subspace in the underlying three-dimensional vector space and reflecting on the non-uniqueness of bases for such spaces may assist the reader.

Definition 1.3. A *conic* in projective space is the locus of points of a homogeneous polynomial of degree 2. A conic is *non-degenerate* if it is non-empty and does not contain an entire projective line.

We note that this is a purely algebraic description of a conic, though motivated by the geometry of the real field.

Proposition 1.4. *A non-degenerate conic in $PG_2(\mathbb{F}_q)$ contains $q + 1$ points. A non-degenerate conic meets a line in at most two points.*

Proof. The generic equation for a conic in $\mathrm{PG}_2(\mathbb{F}_q)$ is $F(x, y, z) = \alpha x^2 + \beta y^2 + \gamma z^2 + \delta xy + \epsilon xz + \zeta yz$. The normal vector is $\nabla F(x, y, z) = [2\alpha + \delta y + \epsilon z : 2\beta + \delta x + \zeta z : 2\gamma + \epsilon x + \zeta y]$, which is linear in x, y, z. A conic is non-degenerate if and only if $\nabla F(x, y, z)$ is non-zero for all $[x : y : z]$ in the corresponding variety. In this case, the normal vector uniquely determines a one-dimensional subspace in the underlying three dimensional vector space. This has a unique two-dimensional orthogonal complement, which is the tangent line to the projective variety.

For the second claim, let p be a point on the conic F. Since there is a unique tangent to the curve at p, and there are precisely $q+1$ lines through p, each of the q remaining lines through p must intersect the conic in additional points. (See Figure 1.) A line through p is described by a linear equation, and the conic by a quadratic equation: substituting one into the other gives a polynomial in one variable of degree at most 2. One root corresponds to p, and so there is precisely one additional point of intersection. $\square$

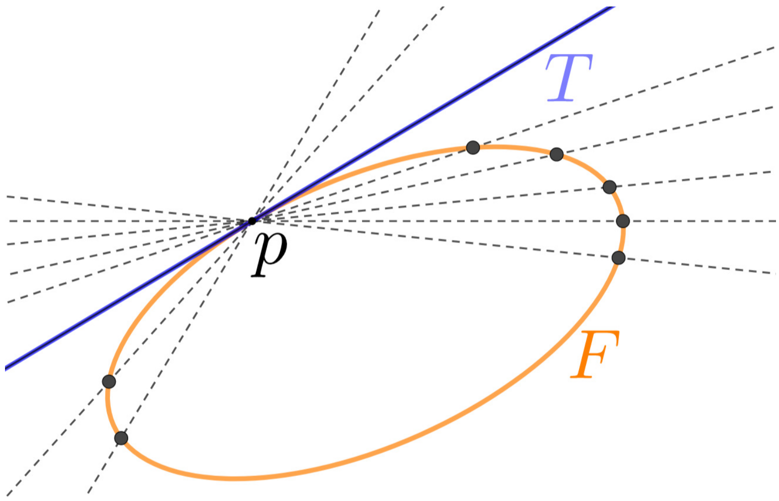


FIGURE 1. A conic F, with tangent T at the point p along with a pencil of lines at p and their intersection points.

Again, Proposition 1.4 is best illustrated with an example. Before this example we remark on a technique that will be used repeatedly.

Remark 1.5. The determinant of a 3×3 matrix having as rows representatives of three projective points vanishes if and only the projective points are collinear. Equivalently, three points in a three dimensional vector space have vanishing determinant if and only if they are coplanar.

Example 1.6. Consider again the conic $F(x,y,z) = x^2 - yz$ over a field with at least 3 elements. Previously, we computed the tangent at $[1:1:1]$. Let us now construct additional points on the curve.

Any line through p can be written parametrically as $[1+\alpha t : 1+\beta t : 1+\gamma t]$. The lines corresponding to (α,β,γ) and (α',β',γ') are distinct if and only if the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ \alpha & \beta & \gamma \\ \alpha' & \beta' & \gamma' \end{pmatrix}$$

is invertible. Let us compute the second point at which the line $[1+t:1+t:1]$ meets the curve:

$$(1+t)^2 - (1+t) = 0 \text{ i.e. } t^2 + t = 0\,.$$

The solution $t=0$ corresponds to p, while the solution $t=-1$ corresponds to the point $[0:0:1]$ on F. In this way, every point on F can be constructed by computing solutions of simple systems of equations.

The reader is encouraged at this point to verify that the tangent to the conic of Example 1.6 at $q = [0 : 0 : 1]$ is the line $\{[1 : 0 : t] : t \in k\} \cup \{[0 : 0 : 1]\}$. Furthermore, the line $x = \alpha y$ contains q for any non-zero α, and the second point of intersection of $[t : \alpha t : 1]$ with the variety is given by

$$t^2 - \alpha t = 0$$

which occurs when $t = \alpha$. Hence, the points on the curve admit the parametrisation $[\alpha : \alpha^2 : 1]$ together with a point 'at infinity' with respect to this parameterisation, which is $[0 : 1 : 0]$.

In contrast to the definition of a conic, the next definition is purely combinatorial.

Definition 1.7. An *oval* in a projective plane is a subset of the points of the plane meeting no line in more than 2 points.

Proposition 1.8. *An oval in a projective plane of order q contains at most $q+2$ points if q is even and $q+1$ points if q is odd.*

Proof. Denote the oval by $\mathcal{O}$, let $p \in \mathcal{O}$ and $r \notin \mathcal{O}$. Each line through p can intersect the oval in at most one additional point, hence there at most $q+2$ points on the oval. If there are $q+2$ points in $\mathcal{O}$ then every line intersecting the oval must do so in two points, there can be no tangents to $\mathcal{O}$.

Suppose now that $\mathcal{O}$ contains $q+2$ points, and consider the lines through r which intersect $\mathcal{O}$. Since each contains precisely two points, the quantity $q+2$, and hence q, must be even. Consequently, when q is odd, an oval contains at most $q+1$ points. □

Following immediately from Propositions 1.4 and 1.8, we have many examples of maximal ovals in projective planes of odd order.

Corollary 1.9. *A conic in a projective plane of odd order is a maximal oval.*

Our goal in this paper is to prove Segre's theorem, which is the converse of this corollary: every maximal oval in a finite projective plane of odd order is actually a conic. The reader may be tempted to think that this converse would be natural to believe, however many prominent researchers in the area did not think that it was true. It was first conjectured by Järnefelt and Kustaanheimo but Marshall Hall said in his review that he found it implausible, [10]. Later Hall reviewed Segre's paper saying that the method of proof was ingenious.

Segre's result is the best possible in the sense that there exist maximal ovals with $q+2$ points in finite planes of even order, not all of which can be constructed from conics. The study of such maximal ovals in planes of even order was a key step in the proof of the non-existence of the projective plane of order 10, [11]. The problem over infinite fields does not appear amenable to classification.

2. The Desargues configuration

Historically, Desargues constructed and worked with projective planes constructed from three dimensional vector spaces. Desargues' theorem is a statement about the collinearity of points lying in a particular configuration. Much later, it was realised that combinatorial objects satisfying the axioms of a projective plane exist. It turns out that the conclusion of Desargues' theorem typically does **not** hold in these exotic planes, and in fact, the validity of Desargues' theorem is a necessary and sufficient condition for co-ordinatisation of a plane by a field. Thus different authors can refer to quite different statements when they write *Desargues' theorem.* We refer the reader to an elementary and historically motivated account of this topic by Blumenthal [3]. In this section, we present the result as it would have been understood by Desargues: that is, in a plane co-ordinatised by a field. Our proof is via linear algebra, and we spend some time illustrating its application, since it will be required in Segre's theorem.

Definition 2.1. A *triangle* is a collection of three non-collinear points in a projective plane. Let $P = \{p_1, p_2, p_3\}$ and $S = \{s_1, s_2, s_3\}$ be two triangles in a projective plane. If the lines $|p_1 s_1|$ and $|p_2 s_2|$ and $|p_3 s_3|$ intersect in a point, then P and S are *in perspective from a point.* This point is called the *center of perspectivity.*

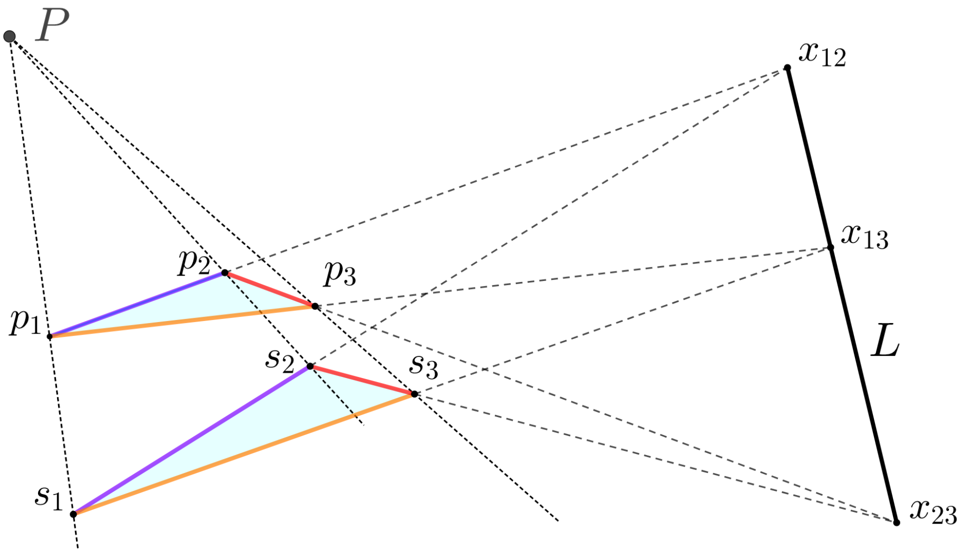


FIGURE 2. The triangles $p_1p_2p_3$ and $s_1s_2s_3$ are in perspective. The point P is the centre of perspectivity and L the line of perspectivity. The points x_{ij} show the construction of the line L. One interpretation of Desargues' theorem is that x_{23} is necessarily on the line spanned by x_{12} and x_{13}.

Desargues' theorem states, that in a Desarguesian projective plane, two triangles in perspective from a point **must** satisfy an additional condition. Geometrically, the intersection points of congruent sides of the triangle must be collinear. Algebraically, this is expressed as the vanishing of a certain determinant. The following proof is the ninth provided by Tan in an article surveying proof techniques for this result, [14].

Theorem 2.2 (Desargues' theorem)**.** *Let $P = \{p_1, p_2, p_3\}$ and $S = \{s_1, s_2, s_3\}$ be triangles in perspective in a projective plane. Denote by x_{ij} the intersection of the congruent sides $|p_i p_j|$ and $|s_i s_j|$ of the two triangles.*

Then the points x_{12}, x_{13} and x_{23} are collinear.

Proof. By hypothesis, the triangles P and S are in perspective from a point, which we may choose without loss of generality[3] as $c = [1 : 1 : 1]$. Again without loss of generality, we may label the points of one triangle as $p_1 = [1 : 0 : 0]$, and $p_2 = [0 : 1 : 0]$ and $p_3 = [0 : 0 : 1]$.

By hypothesis, the point s_i is on the line through p_i and c, so $s_1 = [1 + t_1 : 1 : 1]$ and $s_2 = [1 : 1 + t_2 : 1]$, and $s_3 = [1 : 1 : 1 + t_3]$. The intersection of two lines is most conveniently computed as the simultaneous solution of their linear equations. To find

[3]We will not justify this claim fully, but remark that this is the projective version of the claim that since all bases of a vector space are equivalent, we lose no generality by working with the standard normal basis.

the line through s_1 and s_2, it suffices to compute conditions under which the unknowns x_1, x_2, x_3 render the following matrix rank deficient:

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ 1+t_1 & 1 & 1 \\ 1 & 1+t_2 & 1 \end{pmatrix}.$$

The necessary and sufficient condition, given by the vanishing of the determinant, is $t_2x_1 + t_1x_2 - (t_1t_2 + t_1 + t_2)x_3 = 0$. The line through p_1 and p_2 is given by the equation $x_3 = 0$, and the intersection of these lines is the projective point $x_{12} = [t_1 : -t_2 : 0]$. Similar computations yield $x_{13} = [-t_1 : 0 : t_3]$ and $x_{23} = [0 : t_2 : -t_3]$. These points are collinear if and only if the corresponding matrix, in which points are written as columns,

$$\begin{pmatrix} t_1 & 0 & -t_1 \\ -t_2 & t_2 & 0 \\ 0 & -t_3 & t_3 \end{pmatrix}$$

is rank deficient. But it is easily seen that the column-vector $(1, 1, 1)$ is in the nullspace, which completes the proof. □

We illustrate this theorem with an example, which will be required in the proof Segre's theorem.

Example 2.3. Suppose that $p_1 = [1 : 0 : 0]$ and $p_2 = [0 : 1 : 0]$ and $p_3 = [0 : 0 : 1]$. The lines of the triangle are then $\ell_{12} = [1 : t : 0]$ with equation $x_3 = 0$ as well as ℓ_{13} and ℓ_{23} with equations $x_2 = 0$ and $x_1 = 0$ respectively.

Take $P_1 = [-1 : 1 : 1]$ and $P_2 = [1 : -1 : 1]$ and $P_3 = [1 : 1 : -1]$, which is in perspective with the first triangle through the center of perspectivity $[1 : 1 : 1]$. Its lines are $L_{12} = [-1 + t : 1 - t : 1 + t] = [1 : -1 : t]$ with equation $x_1 + x_2 = 0$. Similarly L_{13} has equation $x_1 + x_3 = 0$ and L_{23} has equation $x_2 + x_3 = 0$.

The intersection of ℓ_1 with L_1 is the point $[1 : -1 : 0]$, the intersection of ℓ_2 with L_2 is $[0 : 1 : -1]$ and the intersection of ℓ_3 with L_3 is $[-1 : 0 : 1]$. Writing these vectors as columns, we perceive that the matrix

$$\begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

is rank-deficient, which means that all three points are contained on a projective line, with equation $x + y + z = 0$.

3. The main theorem

With the necessary background in hand, we proceed to the proof of Segre's theorem. The key step is the famous *Lemma of the Tangents*, which proves that a particular pair of triangles constructed from a conic is in perspective. The main result then applies Desargues' theorem to deduce an algebraic relation between the points on an oval from this configuration. Our proof of Lemma 3.1 is essentially Segre's, while our proof of Theorem 3.2 departs from the original in some details while preserving the essential argument. (Segre achieved his proof without mention of Desargues, though he used the result heavily. He also required a combinatorial result of Qvist on intersections of tangents of an oval which we avoid.)

Lemma 3.1. *Let p_1, p_2, p_3 be three distinct points on an oval in $PG_2(\mathbb{F}_q)$ where q is an odd prime power. Define s_i to be the intersection point of the tangents to the oval at p_{i+1} and p_{i+2}, with subscripts interpreted modulo 3. The triangles $P = \{p_1, p_2, p_3\}$ and $S = \{s_1, s_2, s_3\}$ are in perspective.*

Proof. Without loss of generality, we may choose a co-ordinate system for the projective plane so that

$$p_1 = [1:0:0],\ \ p_2 = [0:1:0],\ \ p_3 = [0:0:1]\,.$$

Observe that the $q+1$ lines through p_1 consist of the q lines $L_1(\alpha)$ described by equations of the form $x_2 = \alpha x_3$ where $\alpha \in \mathbb{F}_q$, and the line $L_1(\infty)$ with equation $x_3 = 0$. The line $L_1(\infty)$ passes through p_2 and the line $L_1(0)$ passes through p_3. Of the remaining $q-1$, precisely one is tangent to the oval, which we denote $L_1(k_1)$.

Define the lines $L_2(\alpha)$ and $L_3(\alpha)$ analogously as the lines passing through the points p_2 and p_3 satisfying equations of the form $x_3 = \alpha x_1$ and $x_1 = \alpha x_2$ respectively. Similarly, write $L_2(k_2)$ and $L_3(k_3)$ for the tangents at p_2 and p_3. The tangents $L_1(k_1)$ and $L_2(k_2)$ are given by the equations $x_2 = k_1 x_3$ and $x_3 = k_2 x_1$, and so intersect at the point $s_3 = [1 : k_1k_2 : k_2]$. Similarly, $s_1 = [k_3 : 1 : k_3k_2]$ and $s_2 = [k_1k_3 : k_1 : 1]$.

To show the triangles $P = \{p_1, p_2, p_3\}$ and $S = \{s_1, s_2, s_3\}$ are in perspective, we must show that the three lines $|p_1s_1| = L_1(k_2k_3)$, and $|p_2s_2| = L_2(k_1k_3)$ and $|p_3s_3| = L_3(k_1k_2)$ meet in a single point. The equations of these lines are respectively

$$x_2 = k_2k_3x_3,\ \ x_3 = k_1k_3x_1,\ \ x_1 = k_1k_2x_2\,. \tag{1}$$

We require a relation between the k_i to show that these equations have a common solution. Let $c = [c_1 : c_2 : c_3]$ be a point on the oval distinct from p_1, p_2, p_3. The entry c_i must be non-zero, otherwise a line $L_j(0)$ with $j \neq i$ would intersect the oval in three points. Denote by $L_i(\lambda_i)$ the line passing through c for $i = 1, 2, 3$. Since $c_2 = \lambda_1 c_3$, it follows that $\lambda_1 = c_2c_3^{-1}$. Similarly, $\lambda_2 = c_3c_1^{-1}$ and $\lambda_3 = c_1c_2^{-1}$. We conclude that

$$\lambda_1\lambda_2\lambda_3 = c_2c_3^{-1}c_3c_1^{-1}c_1c_2^{-1} = 1\,.$$

Denote the remaining $q-2$ points on the oval distinct from p_1, p_2, p_3 by $c_1, \ldots, c_{q-2}$. Each line meets the oval in at most two points, so the line through p_i and c_j is distinct from the line through p_i and c_k for $j \neq k$. Denote by $\lambda_{i,k}$ the unique $\alpha \in \mathbb{F}_q$ such that $L_i(\alpha)$ meets q_k. By the above argument, the identity $\lambda_{1,i}\lambda_{2,i}\lambda_{3,i} = 1$ holds for each $i \in \{1, \ldots, q-2\}$.

The product of the non-zero elements in the field is -1 because the multiplicative group is cyclic and so contains a unique element of order 2 which does not cancel with its inverse in the product. Combining these observations, and using commutativity of multiplication,

$$\prod_{i=1}^{q-2} \lambda_{1,i}\lambda_{2,i}\lambda_{3,i} = \left(\prod_{x\neq k_1} x\right)\left(\prod_{x\neq k_2} x\right)\left(\prod_{x\neq k_3} x\right) = \left(\prod_{x\in\mathbb{F}_q^*} x\right)^3 (k_1k_2k_3)^{-1} = 1\,.$$

Since $\left(\prod_{x\in\mathbb{F}_q^*} x\right)^3 = -1$ we conclude that a non-trivial relationship holds between the three tangents:

$$k_1k_2k_3 = -1\,. \tag{2}$$

Returning at last to the claim: the point $[1 : -k_3 : k_1k_3]$ satisfies the conditions of Equation (1) due to Segre's identity, Equation (2). □

Finally, we show the result which is the aim of this paper, namely Segre's theorem which states that a maximal oval in a projective plane of odd order is in fact a conic.

Theorem 3.2. *The points of a maximal oval in a finite projective plane of odd characteristic satisfy a polynomial equation of degree* 2.

Proof. As in Lemma 3.1, we choose a triangle $P = \{p_1, p_2, p_3\}$ on the oval, and up to projective equivalence we may choose $k_1 = k_2 = k_3 = -1$. With reference to the notation in Lemma 3.1, we have have the points, p_i and the tangent lines $L_i(k_i)$:

$$\begin{array}{lll} p_1 = [1:0:0], & p_2 = [0:1:0], & p_3 = [0:0:1] \\ x_2 = -x_3, & x_3 = -x_1, & x_1 = -x_2, \end{array}$$

Let $c = [c_1 : c_2 : c_3]$ be a point on the oval distinct from the p_i. Let $b_1x_1+b_2x_2+b_3x_3 = 0$ be the unique tangent to the oval at c. As in Lemma 3.1 the co-ordinates c_i are all non-zero, and if b_i were zero then p_i would satisfy the equation of the tangent, a contradiction.

Now, consider the triangle $\{c, p_2, p_3\}$, which by Lemma 3.1 is in perspective to the triangle given by the three tangents

$$b_1x_1 + b_2x_2 + b_3x_3 = 0,\ \ x_3 = -x_1,\ \ x_1 = -x_2\,.$$

By Desargues' Theorem, these triangles are in perspective from a line: we will compute the edges of $\{c, p_2, p_3\}$ and intersect them with the appropriate tangents to derive a relation between the b_i and c_i. First, we intersect the line $|cp_2|$ with the tangent to the oval at p_3. The points of the line $|cp_2|$ are of the form $c + tp_2 = [c_1 : c_2 + t : c_3]$, while the equation of the tangent at p_3 is given by $x_1 = -x_2$. Thus the unique solution is $[c_1 : -c_1 : c_3]$. Similarly, $|cp_3|$ intersects the tangent at p_2 in the point $[c_1 : c_2 : -c_1]$ and the tangent through c intersects $|p_2p_3|$ at the point $[0 : b_3 : -b_2]$.

These three points are collinear, so the determinant of the matrix

$$\begin{pmatrix} 0 & b_3 & -b_2 \\ c_1 & -c_1 & c_3 \\ c_1 & c_2 & -c_1 \end{pmatrix}$$

must vanish. Using that c_1 is non-zero, this is equivalent to the identity

$$b_3(c_1 + c_3) = b_2(c_1 + c_2)\,.$$

An analogous computation with the triangles $\{c, p_1, p_2\}$ and $\{c, p_1, p_3\}$ and the triangles formed from their tangents gives two further identities:

$$b_3(c_2 + c_3) = b_1(c_1 + c_2),\ \ b_1(c_1 + c_3) = b_2(c_2 + c_3)\,.$$

Since $[c_1 : c_2 : c_3]$ lies on the line $b_1x_1 + b_2x_2 + b_3x_3$, the following identity holds:

$$(c_1 + c_2)\,(b_1c_1 + b_2c_2 + b_3c_3) = 0\,.$$

Multiplying out and substituting the identities obtained from Desargues:

$$\begin{aligned} & b_1(c_1 + c_2)c_1 + b_2(c_1 + c_2)c_2 + b_3(c_1 + c_2)c_3 \\ = \ & b_3(c_2 + c_3)c_1 + b_3(c_1 + c_3)c_2 + b_3(c_1 + c_2)c_3 \\ = \ & b_3\left((c_2 + c_3)c_1 + (c_1 + c_3)c_2 + (c_1 + c_2)c_3\right) \\ = \ & 2b_3\left(c_1c_2 + c_2c_3 + c_3c_1\right)\,. \end{aligned}$$

Since $b_3 \neq 0$ and the characteristic is odd, $c_1c_2 + c_2c_3 + c_3c_1 = 0$ holds for the point $[c_1 : c_2 : c_3]$ of the oval. But the point c was an arbitrary point of the oval distinct from the p_i and the equation holds for the points p_i. We have shown that the $q + 1$ points of the oval lie on the conic $x_1x_2 + x_2x_3 + x_3x_1$. This completes the proof. □

Remark 3.3. The reader may be perturbed by the explicit conic constructed in Theorem 3.2. In fact, all conics in $PG_2(\mathbb{F}_q)$ are projectively equivalent (in essentially the same way that all bases of a vector space are equivalent up to choice of basis), and this conic was forced by our choice of basis at the start of the proof.

References

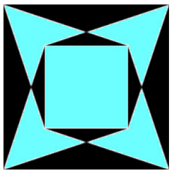
[1] E. Artin. *Geometric algebra*. Interscience Publishers, Inc., New York-London, 1957.
[2] S. Ball. On sets of vectors of a finite vector space in which every subset of basis size is a basis. *J. Eur. Math. Soc. (JEMS)*, 14(3):733–748, 2012.
[3] L. M. Blumenthal. *A modern view of geometry*. Dover Publications, Inc., New York, 1980. Corrected reprint of the 1961 original.
[4] K. Coolsaet. The lemma of tangents reformulated. *Discrete Math.*, 312(3):705–714, 2012.
[5] P. Dembowski. *Finite geometries*. Classics in Mathematics. Springer-Verlag, Berlin, 1997. Reprint of the 1968 original.
[6] S. T. Dougherty. *Combinatorics and Finite Geometry*. Springer, first edition, 2020.
[7] D. Hilbert. *Foundations of geometry*. Open Court, La Salle, Ill., second edition, 1971. Translated from the tenth German edition by Leo Unger.
[8] D. R. Hughes and F. C. Piper. *Projective planes*. Graduate Texts in Mathematics, Vol. 6. Springer-Verlag, New York-Berlin, 1973.
[9] I. M. Isaacs. *Algebra: a graduate course*, volume 100 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2009. Reprint of the 1994 original.
[10] G. Järnefelt and P. Kustaanheimo. An observation on finite geometries. In *Den 11te Skandinaviske Matematikerkongress, Trondheim, 1949*, pages 166–182. Johan Grundt Tanums Forlag, Oslo, 1952.
[11] C. W. H. Lam, L. Thiel, and S. Swiercz. The nonexistence of finite projective planes of order 10. *Canad. J. Math.*, 41(6):1117–1123, 1989.
[12] B. Segre. Ovals in a finite projective plane. *Canadian Journal of Mathematics*, 7:414–416, 1955.
[13] I. R. Shafarevich. *Basic algebraic geometry. 1*. Springer, Heidelberg, third edition, 2013. Varieties in projective space.
[14] K. Tan. Different proofs of Desargues' theorem. *Mathematics Magazine*, 40(1):14–25, 1967.

Patrick Browne received a B.Sc and Ph.D both from the National University of Ireland, Galway, with advisor Dr J. Burns. He has worked in a number of third level institutions around Ireland and currently is a lecturer at TUS in Limerick. His research interests are in the areas of Lie algebras and Riemannian geometry.
Steven T. Dougherty received his Ph.D. from Lehigh University in 1992 and is a professor of mathematics at the University of Scranton. He has been published over 120 times in combinatorics, geometry, coding theory, and number theory and is the author of two books. He has lectured in 12 countries and is the recipient of the 2005 Hasse Prize.
Padraig Ó Catháin received a BA in Mathematics and History from the University of Galway in 2007. Under the direction of Dane Flannery he was awarded the degrees of MLitt in 2008 and PhD in 2012, by the same institution. After a decade at universities in Australia, Finland and the United States, he took a position as Ollamh Cúnta in Fiontar agus Scoil na Gaeilge at Dublin City University in 2022. His research is predominantly in combinatorics, for which he has been awarded the Kirkman medal of the Institute of Combinatorics and its Applications.

(Browne) Technological University of the Shannon: Midlands Midwest.

(Dougherty) Dublin City University.

(Ó Catháin) University of Scranton.
E-mail address, Browne: `patrick.browne@tus.ie`
E-mail address, Dougherty: `Prof.Steven.Dougherty@gmail.com`
E-mail address, Ó Catháin: `padraig.ocathain@dcu.ie`



Irish Math. Soc. Bulletin
Number 91, Summer 2023, 49–72
ISSN 0791-5578

Pairs of Quadratic Forms over the Real Numbers

DAVID B. LEEP AND NANDITA SAHAJPAL

Abstract. This survey paper examines several topics concerning pairs of quadratic forms with real coefficients. We state a theorem that characterizes pairs of real quadratic forms having a nontrivial common zero and give a proof using a method based on point-set topology. This proof relies on determining when various subsets associated with one quadratic form are path-connected. Additionally, we describe how the signature and rank of a quadratic form change over a 2-dimensional family of quadratic forms. Finally, we delve into nonsingular pairs of quadratic forms, simultaneous diagonalization, and provide a proof of the spectral theorem. This paper presents a self-contained exposition of these results.

1. Introduction

Determining whether a quadratic form f with real coefficients has a nontrivial real zero is straightforward. One can start by diagonalizing f using linear algebra techniques or repeatedly applying the completing the square method to express f in the form $d_1X_1^2 + \cdots + d_nX_n^2$. Then, f has a nontrivial real zero if and only if the d_i's are not all positive and not all negative.

Suppose that $f, g \in \mathbb{R}[X_1, \ldots, X_n]$ are quadratic forms in n variables. How does one determine if f, g have a nontrivial common real zero? The answer to this problem is much harder, but it is well known to experts and has been discussed in many places. See [8] for a large bibliography on this subject. One of the main goals of this paper is to answer this question with an exposition that is as self-contained as possible.

In Proposition 4.4, we determine when a pair of quadratic forms with real coefficients has a nontrivial common zero. We follow a method based on point-set topology that Swinnerton-Dyer used in [7, Lemma 1 (i)]. The ideas even go back at least as far as [2]. However, there seems to be a gap in Swinnerton-Dyer's proof. Our exposition will fill in the details of this gap. See Remark 4.7 for specifics.

In Section 2, we present essential material on quadratic forms that we need for this paper. Since it is no extra trouble, we give definitions, statements, and proofs of results in this section that are valid over any field K of characteristic different from 2. We introduce the objects using a basis-free approach and then routinely use convenient bases for efficient calculations.

In Section 3, we investigate some topological properties of the zero sets of one quadratic form with real coefficients. Some of these properties are used in Section 4 to help us solve our problem for pairs of quadratic forms with real coefficients.

In Sections 5 through 8, we study several other topics that are relevant to pairs of quadratic forms over the real numbers. In Section 5, we study how the signature of a quadratic form changes over a 2-dimensional family of quadratic forms. In Section 6,

2020 *Mathematics Subject Classification.* 12D10.
Key words and phrases. Pair of quadratic forms, real quadratic forms, zeros of quadratic forms, rank and signature of a quadratic form.
Received on 24-3-2023; revised 11-5-2023, 21-5-2023.
DOI:10.33232/BIMS.0091.49.72.

we study the ranks of quadratic forms in a 2-dimensional family of quadratic forms, and we relate the rank to the multiplicity of a zero in a naturally occurring polynomial. For quadratic forms $f, g \in \mathbb{R}[X_1, \dots, X_n]$, we also study in Proposition 6.4 the problem of finding a real linear combination $\lambda f + \mu g$ that splits off as many hyperbolic planes as possible. In Section 7, we study nonsingular pairs of quadratic forms and conditions when two quadratic forms can be simultaneously diagonalized. In Section 8, we apply results from Section 7 to pairs of quadratic forms over $\mathbb{R}$. In particular, we use Proposition 8.1 to strengthen a result of Heath-Brown in [3, Lemma 12.1]. See Remark 8.2 for specifics. We end by using our results to give a proof of the Spectral Theorem.

Here are some of the notations and basic notions used throughout this paper. Let K be a field and let $K^\times = K \backslash \{0\}$. We let $\operatorname{char} K$ denote the characteristic of K. Let K^{alg} denote an algebraic closure of K. Unless otherwise noted, we work only with fields K with $\operatorname{char} K \neq 2$. We let $K[X_1, \dots, X_n]$ denote the polynomial ring in n variables.

For $f, g \in K[X_1, \dots, X_n]$, we write $f \mid g$ if f divides g in $K[X_1, \dots, X_n]$. Recall that $f \mid g$ in $K[X_1, \dots, X_n]$ if and only if $f \mid g$ in $K^{alg}[X_1, \dots, X_n]$. A polynomial $f \in K[X_1, \dots, X_n]$ is a homogeneous form of degree m, $m \geq 0$, if each monomial in f has degree m. If f is a homogeneous form, we say that $(a_1, \dots, a_n)$ is a nontrivial zero of f if $f(a_1, \dots, a_n) = 0$ and some $a_i \neq 0$. A quadratic form is a homogeneous form of degree 2. We let $e_1, \dots, e_n$ denote the standard basis of K^n.

We let $\mathbb{R}$ denote the field of real numbers and $\mathbb{C}$ the field of complex numbers. Recall that $\mathbb{C}$ is an algebraically closed field and that $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$, and so $\mathbb{C} = \mathbb{R}^{alg}$.

2. Basic results about quadratic forms

Definition 2.1 (Quadratic Map)**.** Let V be a finite-dimensional vector space over a field K. A quadratic map $f : V \to K$ is a function satisfying the following two conditions.

(1) $f(av) = a^2 f(v)$ for all $v \in V$ and $a \in K$.
(2) The function $B_f : V \times V \to K$ defined by $B_f(v, w) = f(v + w) - f(v) - f(w)$ is a symmetric bilinear form.

We recover the usual notion of a quadratic form by introducing a basis $\{v_1, \dots, v_n\}$ of V. The definition of a quadratic map implies that $f(X_1 v_1 + X_2 v_2) = f(v_1)X_1^2 + B_f(v_1, v_2)X_1X_2 + f(v_2)X_2^2$. A straightforward induction implies that for variables $X_1, \dots, X_n$, we have

$$f(X_1 v_1 + \dots + X_n v_n) = \sum_{i=1}^n f(v_i)X_i^2 + \sum_{1 \leq i < j \leq n} B_f(v_i, v_j)X_iX_j.$$

Let

$$f = \sum_{i=1}^n a_{ii}X_i^2 + \sum_{1 \leq i < j \leq n} a_{ij}X_iX_j \in K[X_1, \dots, X_n],$$

where $a_{ii} = f(v_i)$, $1 \leq i \leq n$, and $a_{ij} = B_f(v_i, v_j)$ for $1 \leq i < j \leq n$. We call f the quadratic form associated to the quadratic map $f : V \to K$ and the basis $\{v_1, \dots, v_n\}$. We see that f is a homogeneous form having degree 2, as expected.

Let $f = \sum_{i=1}^n a_{ii}X_i^2 + \sum_{1 \leq i < j \leq n} a_{ij}X_iX_j \in K[X_1, \dots, X_n]$ be a quadratic form. Associated to f is an $n \times n$ symmetric matrix $M = (m_{ij})$ where

$$m_{ij} = \begin{cases} a_{ii} & \text{if } i = j \\ \frac{1}{2}a_{ij} & \text{if } i < j \\ \frac{1}{2}a_{ji} & \text{if } i > j. \end{cases}$$

We have $f(X_1, \ldots, X_n) = X^t M X$, where $X = (X_1, \ldots, X_n)^t$. It is convenient to regard f as a function $f : K^n \to K$. For a subspace $W \subseteq K^n$, we let $f|_W$ denote the restriction of f to W.

The associated symmetric bilinear form to f is given by $B_f : K^n \times K^n \to K$ defined $B_f(v, w) = v^t M w$ where $v, w \in K^n$ are column vectors. Thus $f(X) = B_f(X, X)$. For a subspace $W \subseteq K^n$, we define the orthogonal complement

$$W^\perp = \{v \in K^n \mid B_f(v, w) = 0 \text{ for all } w \in W\}.$$

It is easily verified that $W^\perp$ is a subspace of K^n. We write $W^{\perp_f}$ if we need to specify the orthogonal complement of W for a particular quadratic form f.

Suppose that $f : K^n \to K$ is a quadratic map and that W is a subspace of K^n such that $K^n = W \oplus W^\perp$. Let $\{v_1, \ldots, v_j\}$ and $\{v_{j+1}, \ldots, v_n\}$ be bases for W and $W^\perp$, respectively. Then the quadratic form associated to f and the basis $\{v_1, \ldots, v_n\}$ is $f_1(X_1, \ldots, X_j) + f_2(X_{j+1}, \ldots, X_n)$, where f_1 and f_2 are the quadratic forms associated to $f|_W$ and $f|_{W^\perp}$, respectively, and the bases $\{v_1, \ldots, v_j\}$ and $\{v_{j+1}, \ldots, v_n\}$.

Suppose $V = W \oplus Y$ and let $f : V \to K$ be a quadratic map. Let $g = f|_W$ and $h = f|_Y$. If $B_f(w, y) = 0$ for all $w \in W$, $y \in Y$, then we write $f = g \perp h$.

For quadratic forms $f, g \in K[X_1, \ldots, X_n]$, the K-pencil of f, g, denoted by $\mathcal{P}_K(f, g)$, consists of all linear combinations $af + bg$ where $a, b \in K$, not both zero.

Note that f, g have a nontrivial common zero over K if and only if $rf + sg$ and $tf + ug$ have a nontrivial common zero over K where $r, s, t, u \in K$ and the matrix $\begin{pmatrix} r & s \\ t & u \end{pmatrix}$ is invertible. Because of this, it is often useful to replace f, g with two other convenient quadratic forms in $\mathcal{P}_K(f, g)$.

Two quadratic forms $f, g \in K[X_1, \ldots, X_n]$ are equivalent, written $f \cong g$, if there exists an invertible $n \times n$ matrix A with entries in K such that $f(X) = g(AX)$, where $X = (X_1, \ldots, X_n)$. In this situation, we say that f is obtained from g by an invertible linear change of variables. If $g(X_1, \ldots, X_n) = X^t N X$, then this is equivalent to the condition $M = A^t N A$. A quadratic form f is equivalent to a diagonal form $d_1 X_1^2 + \cdots + d_n X_n^2$ under an invertible linear change of variables over K because for any symmetric matrix M there is an invertible matrix A such that $A^t M A$ is a diagonal matrix. Such a diagonal form is denoted by $\langle d_1, \ldots, d_n \rangle$.

Definition 2.2 (Rank of a Quadratic Form over K)**.** The rank of a quadratic form f, denoted by $\operatorname{rank}(f)$, is the rank of the matrix M.

If f is equivalent to $\langle d_1, \ldots, d_n \rangle$, then $\operatorname{rank}(f)$ is the number of nonzero d_i.

Definition 2.3 (Radical of a Quadratic Form)**.** Let $f \in K[X_1, \ldots, X_n]$ be a quadratic form with associated symmetric bilinear form B_f. The radical of f over K is the subspace

$$\operatorname{rad}(f) = \{v \in K^n : B_f(v, K^n) = 0\}.$$

We say that a quadratic form f is nonsingular if $\operatorname{rad}(f) = 0$, and is singular if $\operatorname{rad}(f) \neq 0$.

We can write $K^n = V \oplus \operatorname{rad}(f)$ for some subspace $V \subseteq K^n$, and it is straightforward to check that $f|_V$ is nonsingular. We let $\operatorname{Null}(M)$ denote the null space of a matrix M.

Lemma 2.4. *Let $f \in K[X_1, \ldots, X_n]$ be a quadratic form with associated $n \times n$ symmetric matrix M.*

(1) $\operatorname{rad}(f) = \operatorname{Null}(M)$.
(2) $\operatorname{rank}(f) + \dim(\operatorname{rad}(f)) = n$.
(3) *The following statements are equivalent.*
   (a) *f is singular.*

(b) $\mathrm{rad}(f) \neq 0$.
(c) $\mathrm{rank}(f) < n$.
(d) $\det(M) = 0$.

Proof. (1) Let $v \in \mathrm{Null}(M)$. Then $B_f(w, v) = w^t M v = 0$ for all $w \in K^n$, and thus $v \in \mathrm{rad}(f)$. Let $v \in \mathrm{rad}(f)$. Then $B_f(w, v) = 0$ for all $w \in K^n$. Thus $w^t M v = 0$ for all $w \in K^n$, which implies that $Mv = 0$. Therefore, $v \in \mathrm{Null}(M)$.

(2) We have $n = \mathrm{rank}(M) + \dim(\mathrm{Null}(M) = \mathrm{rank}(f) + \dim(\mathrm{rad}(f))$.

(3) The equivalence of the statements follows from the definitions, (1) and (2), and the observation that $\det(M) = 0$ if and only if $\mathrm{Null}(M) \neq 0$. ☐

Definition 2.5 (Hyperbolic Plane)**.** A quadratic form $f \in K[X_1, X_2]$ is called a hyperbolic plane if f is equivalent to the quadratic form X_1X_2.

The following result is useful to identify a hyperbolic plane.

Lemma 2.6. *Let $f \in K[X_1, X_2]$ be a nonsingular quadratic form. The following statements are equivalent.*

(1) *f has a nontrivial zero over K.*
(2) *$f \cong X_1X_2$.*
(3) *$f \cong \langle 1, -1\rangle$.*

Proof. (1) $\Rightarrow$ (2): Let $f = aX_1^2 + 2bX_1X_2 + cX_2^2$ with the associated symmetric matrix $M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$. Assume that f has a nontrivial zero over K. Applying an invertible linear change of variables over K allows us to assume that $f(1, 0) = 0$. Then $a = 0$. Since f is nonsingular, we have $b \neq 0$. Then $f = X_2(2bX_1 + cX_2) \cong X_1X_2$ because X_2 and $2bX_1 + cX_2$ are linearly independent.

(2) $\Rightarrow$ (3): We have $X_1X_2 \cong (X_1 + X_2)(X_1 - X_2) = X_1^2 - X_2^2$.

(3) $\Rightarrow$ (1): $X_1^2 - X_2^2$ has a nontrivial zero over K, namely, $X_1 = 1$, $X_2 = 1$. ☐

For a quadratic form $f \in K[X_1 \ldots, X_n]$, we say that f splits off a hyperbolic plane if f is equivalent to $X_1X_2 + h(X_3, \ldots, X_n)$ for some quadratic form $h \in K[X_3, \ldots, X_n]$. Similarly, we say that f splits off j hyperbolic planes, if

$$f \cong X_1X_2 + X_3X_4 + \cdots + X_{2j-1}X_{2j} + h(X_{2j+1}, \ldots, X_n).$$

Lemma 2.7. *Suppose that $f \in K[X_1, \ldots, X_n]$ is a nonsingular quadratic form that has a nontrivial zero over K. Then $f \cong X_1X_2 + h(X_3, \ldots, X_n)$ for some quadratic form $h \in K[X_3, \ldots, X_n]$.*

Proof. An invertible linear change of variables allows us to assume that $f(1, 0, \ldots, 0) = 0$. Then $f \cong X_1L(X_2, \ldots, X_n) + Q_1(X_2, \ldots, X_n)$ where L is a linear form and Q_1 is a quadratic form, both with coefficients in K. We have $L \neq 0$ because $\mathrm{rad}(f) = (0)$. A second invertible linear change of variables allows us to assume that

$$f \cong X_1X_2 + Q_2(X_2, \ldots, X_n) = X_2(X_1 + c_2X_2 + \cdots + c_nX_n) + h(X_3, \ldots, X_n)$$

where Q_2, h are quadratic forms with coefficients in K. A third invertible linear change of variables allows us to assume that $f \cong X_1X_2 + h(X_3, \ldots, X_n)$. ☐

3. One Quadratic Form over the Real Numbers

In this section, we prove some basic properties of quadratic forms over $\mathbb{R}$. We begin by introducing some definitions and terminology that are specific to quadratic forms over $\mathbb{R}$. Let $f \in \mathbb{R}[X_1, \ldots, X_n]$ be a quadratic form with associated symmetric matrix M.

Definition 3.1 (Definite, Semi-definite, Indefinite)**.**

(1) We say that f is a definite quadratic form over $\mathbb{R}$ if $f(v)$ has the same sign for every $v \in \mathbb{R}^n \backslash 0$. According to that sign, the quadratic form f is called positive definite or negative definite.
(2) We say that f is a semi-definite quadratic form over $\mathbb{R}$ if $f(v)$ is either non-negative or non-positive for every $v \in \mathbb{R}^n \backslash 0$. If $f(v)$ is non-negative for every $v \in \mathbb{R}^n \backslash 0$, then f is called positive semi-definite. If $f(v)$ is non-positive for every $v \in \mathbb{R}^n \backslash 0$, then f is called negative semi-definite.
(3) We say that f is an indefinite quadratic form over $\mathbb{R}$ if f takes both positive and negative values when evaluated at vectors in $\mathbb{R}^n \backslash 0$.

We say that an $n \times n$ symmetric matrix M is positive definite (negative definite, semi-definite, indefinite) if the quadratic form $f(X_1, \dots, X_n) = X^t M X$ has that property.

Proposition 3.2. *A quadratic form $f \in \mathbb{R}[X_1, \dots, X_n]$ has a nontrivial zero over $\mathbb{R}$ if and only if f is not definite.*

Proof. Since f is equivalent to a diagonal form $d_1 X_1^2 + \cdots + d_n X_n^2$, it follows that f has a nontrivial zero over $\mathbb{R}$ if and only if $d_1, \dots, d_n$ do not all have the same sign. □

Definition 3.3 (Signature of a Quadratic Form over $\mathbb{R}$)**.** Suppose that f is equivalent to $\langle d_1, \dots, d_n \rangle$. Let r be the number of elements in the set $\{d_i \mid d_i > 0, 1 \le i \le n\}$, and s be the number of elements in the set $\{d_i \mid d_i < 0, 1 \le i \le n\}$. The signature of f, denoted by $\mathrm{sgn}(f)$, is defined by $\mathrm{sgn}(f) = r - s$.

Proposition 3.4. *The signature of f does not depend on the diagonalization of f.*

Proof. We can write $\mathbb{R}^n = V_1 \oplus V_2 \oplus \mathrm{rad}(f)$ where $\dim(V_1) = r$ and f is positive definite on V_1, $\dim(V_2) = s$ and f is negative definite on V_2, and $\dim(\mathrm{rad}(f)) = t$ where t is the number of d_i's that equal zero.

Similarly, suppose that f is also equivalent to $\langle d_1', \dots, d_n' \rangle$ and write $\mathbb{R}^n = V_1' \oplus V_2' \oplus \mathrm{rad}(f)$ where $V_1', V_2', \mathrm{rad}(f)$ have dimensions r', s', t, respectively, as well as the other properties above, and $\mathrm{sgn}(f) = r' - s'$.

Suppose that $r \neq r'$. We can assume that $r < r'$ and then $s > s'$ because $r + s = r' + s' = n - t$. It follows that $(V_2 \oplus \mathrm{rad}(f)) \cap V_1' \neq (0)$ because

$$\dim(V_2 \oplus \mathrm{rad}(f)) + \dim(V_1') = s + t + r' > s + t + r = n.$$

Let $v \in (V_2 \oplus \mathrm{rad}(f)) \cap V_1'$ with $v \neq 0$. Then $f(v) \le 0$ because $v \in V_2 \oplus \mathrm{rad}(f)$, and $f(v) > 0$ because $v \in V_1'$. This contradiction shows that $r = r'$, and thus $s = s'$. Therefore $r - s = r' - s'$. □

Definition 3.5 (Principal Minor)**.** Let M be an $m \times m$ square matrix. A principal sub-matrix of M is a matrix obtained by deleting any k rows and the corresponding k columns. The leading principal sub-matrix of order k of M is obtained by deleting the last $m - k$ rows and columns of M. The determinant of a principal sub-matrix of a matrix M is called a principal minor of M, and the determinant of a leading principal sub-matrix of M is called a leading principal minor of M.

A version of the following result is stated in many textbooks. See for example, [4, p. 328]. We give a particularly nice proof for the convenience of the reader. Note that a principal minor in [4] is what we call a leading principal minor.

Proposition 3.6 (Sylvester's Criterion)**.** *Let A be a real symmetric $n \times n$ matrix. The following statements are equivalent.*

(1) *A is positive definite.*
(2) *Every principal minor of A is positive.*
(3) *Every leading principal minor of A is positive.*

Proof. (1) $\Rightarrow$ (2) Suppose that A is positive definite. Then each principal $i \times i$ submatrix B is positive definite. Since $B = C^tDC$ for some invertible $i \times i$ matrix C and some diagonal $i \times i$ matrix D with positive entries along its main diagonal, it follows that $\det(B) > 0$. Thus (2) holds.

(2) $\Rightarrow$ (3) is obvious.

(3) $\Rightarrow$ (1) Assume that every leading principal minor of A is positive. The proof is by induction on n with the case $n = 1$ being trivial. Assume that $n \geq 2$ and that the result has been proved for real symmetric $(n-1) \times (n-1)$ matrices. Let

$$A = \begin{pmatrix} M & v \\ v^t & c \end{pmatrix},$$

where M is an $(n-1) \times (n-1)$ matrix, v is an $(n-1) \times 1$ matrix, and $c \in \mathbb{R}$. Since all leading principal minors of A are positive, it follows that every leading principal minor of M is positive. Then $\det(M) > 0$ and so M is invertible, and by induction, M is positive definite. Note that M^{-1} is also symmetric. Let

$$L = \begin{pmatrix} I & 0 \\ v^tM^{-1} & 1 \end{pmatrix}, \quad B = \begin{pmatrix} M & 0 \\ 0 & c - v^tM^{-1}v \end{pmatrix}.$$

Since

$$A = \begin{pmatrix} M & v \\ v^t & c \end{pmatrix} = \begin{pmatrix} I & 0 \\ v^tM^{-1} & 1 \end{pmatrix} \begin{pmatrix} M & 0 \\ 0 & c - v^tM^{-1}v \end{pmatrix} \begin{pmatrix} I & M^{-1}v \\ 0 & 1 \end{pmatrix},$$

we have

$$A = LBL^t.$$

Since $\det(L) = 1$ and $\det(M) > 0$, this gives $\det(A) = \det(L)^2 \det(B) = \det(B) = \det(M)\,(c - v^tM^{-1}v)$. Since $\det(A) > 0$, this gives $c - v^tM^{-1}v > 0$. Therefore B is positive definite. Since $A = LBL^t$, it follows that A is positive definite. $\square$

Definition 3.7 (Path-Connected Topological Space)**.** A topological space $\mathbb{X}$ is path-connected if for any $p, q \in \mathbb{X}$, there is a continuous map $\gamma : [0,1] \to \mathbb{X}$ such that $\gamma(0) = p$ and $\gamma(1) = q$. Such a map is called a path from p to q in $\mathbb{X}$.

Definition 3.8 (Unit $m-$sphere, $\mathbb{S}^m$)**.** Let $m \geq 1$ be any natural number. The unit $m-$sphere is defined as

$$\mathbb{S}^m = \Big\{(x_1, \dots, x_{m+1}) \in \mathbb{R}^{m+1} \Big| \sum_{i=1}^{m+1} x_i^2 = 1\Big\}.$$

Lemma 3.9. *For $m \geq 1$, $\mathbb{S}^m$ is a path-connected subset in $\mathbb{R}^{m+1}$.*

Proof. Let $x = (x_1, \dots, x_{m+1})$. The map $\sigma : \mathbb{R}^{m+1} \backslash 0 \to \mathbb{S}^m$ given by $\sigma(x) = \dfrac{x}{\sqrt{\sum_{i=1}^{m+1} x_i^2}}$ is a well-defined, continuous map such that $\sigma(\mathbb{R}^{m+1} \backslash 0) = \mathbb{S}^m$. Since $\mathbb{R}^{m+1} \backslash 0$ is path-connected for $m \geq 1$ and a continuous image of a path-connected set is also path-connected, it follows that $\sigma(\mathbb{R}^{m+1} \backslash 0) = \mathbb{S}^m$ is a path-connected subset in $\mathbb{R}^{m+1}$. $\square$

The next three propositions determine whether certain subsets in $\mathbb{R}^n$ associated with a quadratic form are path-connected. Some of these results are used later in the proofs of Proposition 4.1 and Proposition 4.4. For the sake of completeness, we give a complete treatment here.

Notation. *Let $f \in \mathbb{R}[X_1, \dots, X_n]$ be a quadratic form, and let $f_{=0}$ denote the set $\{x \in \mathbb{R}^n \backslash 0 \mid f(x) = 0\}$. The notation $f_{>0}$, $f_{\geq 0}$, $f_{<0}$, $f_{\leq 0}$ are defined in a similar fashion.*

Let $\mathbb{X} \subset \mathbb{R}^n \backslash 0$ be any subset. Consider the relation $\sim$ on $\mathbb{X}$ defined by $p \sim q$ if there exists a path from p to q which lies entirely in $\mathbb{X}$. Then $\sim$ is an equivalence relation on $\mathbb{X}$ and the equivalence classes are called the path-connected components of $\mathbb{X}$. The set $\mathbb{X}$ can be written as a disjoint union of these path-connected components.

To study the path-connected components of the above sets, we will apply an invertible linear change of variables to put the quadratic form into a convenient shape that is easy to work with. Since an invertible linear map of $\mathbb{R}^n$ is a homeomorphism, path-connected components are mapped to path-connected components.

Since a quadratic form defined over $\mathbb{R}$ can be diagonalized, it is easy to check that if a quadratic form $f \in \mathbb{R}[X_1, \ldots, X_n]$ is indefinite and has rank $r \leq n$, then $\pm f$ is equivalent to either $X_1^2 + \cdots + X_{r-1}^2 - X_r^2$ or $X_1^2 + \cdots + X_k^2 - X_{k+1}^2 - \cdots - X_r^2$ with $k \geq 2$, $r - k \geq 2$. We focus on these two quadratic forms in Proposition 3.11 and Proposition 3.12.

Proposition 3.10. *Let $f \in \mathbb{R}[X_1, \ldots, X_n]$ be a quadratic form in $n \geq 2$ variables. Assume that* $\operatorname{rad}(f) \neq 0$. *Then $f_{=0}$, $f_{\leq 0}$, and $f_{\geq 0}$ are path-connected.*

Proof. Let $\operatorname{rank}(f) = m$. Then we can assume that $f \in \mathbb{R}[X_1, \ldots, X_m]$, and $m < n$ because $\operatorname{rad}(f) \neq 0$. Let $u = (a_1, \ldots, a_m, \ldots, a_n) \in f_{=0}$. Then $u = (a_1, \ldots, a_n)$ is path-connected to $u' = (a_1, \ldots, a_m, 1, a_{m+2}, \ldots, a_n)$ by a line segment that lies in $f_{=0}$, and u' is path-connected to e_{m+1} by a line segment that lies in $f_{=0}$. Therefore, u is path-connected to e_{m+1} in $f_{=0}$. Since any two points in $f_{=0}$ are path-connected to e_{m+1}, it follows that $f_{=0}$ is path-connected.

The proof that $f_{\leq 0}$ and $f_{\geq 0}$ are path-connected is obtained by replacing each $f_{=0}$ with either $f_{\leq 0}$ or $f_{\geq 0}$. □

Proposition 3.11. *Let $f = X_1^2 + \cdots + X_{r-1}^2 - X_r^2 \in \mathbb{R}[X_1, \ldots, X_n]$, $r \leq n$.*

(1) (a) *If $2 = r = n$, then $f_{>0}$, $f_{\geq 0}$, $f_{<0}$, and $f_{\leq 0}$ each have two path-connected components $\mathcal{B}_1$ and $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$.*
   (b) *If $2 = r < n$, then $f_{\geq 0}$ and $f_{\leq 0}$ are path-connected.*
   (c) *If $2 = r < n$, then $f_{>0}$ and $f_{<0}$ each have two path-connected components $\mathcal{B}_1$ and $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$.*

(2) (a) *If $3 \leq r < n$, then $f_{\leq 0}$ is path-connected.*
   (b) *If $3 \leq r$, then $f_{>0}$ and $f_{\geq 0}$ are path-connected.*
   (c) *If $3 \leq r = n$, then $f_{\leq 0}$ is not path-connected.*
   (d) *If $3 \leq r$ then $f_{<0}$ is not path-connected.*

   Let $u = (a_1, \ldots, a_n), v = (b_1, \ldots, b_n) \in f_{\leq 0}\,(f_{<0})$. In (2c) and (2d), u, v are path-connected in $f_{\leq 0}\,(f_{<0})$ if and only if a_r, b_r have the same sign. In particular, $f_{\leq 0}\,(f_{<0})$ has two path-connected components $\mathcal{B}_1$ and $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$.

(3) (a) *If $2 \leq r < n$, then $f_{=0}$ is path-connected.*
   (b) *Let $2 \leq r = n$.*
      (i) *If $r = 2$, then $f_{=0}$ has four path-connected components.*
      (ii) *Assume $3 \leq r = n$. Let $u = (a_1, \ldots, a_r), v = (b_1, \ldots, b_r) \in f_{=0}$. Then u, v are path-connected in $f_{=0}$ if and only if a_r, b_r have the same sign. In particular, $f_{=0}$ has two path-connected components $\mathcal{B}_1$ and $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$.*

Proof. (1) Note that in this case $f = X_1^2 - X_2^2$.

(a) If $2 = r = n$, then $(1, 0)$ and $(-1, 0)$ lie in $f_{\geq 0}$ but are not path-connected in $f_{\geq 0}$ because a path joining $(1, 0)$ and $(-1, 0)$ would contain a point $(0, b)$ for some nonzero $b \in \mathbb{R}$, but such a point would not lie in $f_{\geq 0}$. Similarly,

	$f_{>0}$	$f_{<0}$	$f_{=0}$	$f_{\geq 0}$	$f_{\leq 0}$
$r = 2 = n$	No†	No†	No	No†	No †
$r = 2,\ r < n$	No †	No†	Yes	Yes	Yes
$r \geq 3,\ r = n$	Yes	No†	No†	Yes	No†
$r \geq 3,\ r < n$	Yes	No†	Yes	Yes	Yes

TABLE 1. Summary of the results from Proposition 3.11. The sets marked with "No†" have two path-connected components $\mathcal{B}_1$ and $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$.

$f_{>0}$ is also not path-connected. Since $r = 2$, by replacing f with $-f$, we get that $f_{<0}$ and $f_{\leq 0}$ are each not path-connected.

(b) If $2 = r < n$, then $f_{\geq 0}$ and $f_{\leq 0}$ are path-connected by Proposition 3.10.

(c) The points $(1, 0, \ldots, 0)$ and $(-1, 0, \ldots, 0)$ lie in $f_{>0}$ but are not path-connected in $f_{>0}$ because a path joining $(1, 0, \ldots, 0)$ and $(-1, 0, \ldots, 0)$ would contain a point $(0, b_2, \ldots, b_n)$ where some $b_i \neq 0$, $2 \leq i \leq n$, but such a point lies in $f_{\leq 0}$. Since $r = 2$, by replacing f with $-f$, we get that $f_{<0}$ is also not path-connected.

In (a) and (c), it follows that the indicated set is the union of two path-connected components $\mathcal{B}_1$ and $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$.

(2) (a) If $3 \leq r < n$, then $f_{\leq 0}$ is path-connected by Proposition 3.10.

(b) Let $3 \leq r$ and let $u = (a_1, \ldots, a_n)$, $v = (b_1, \ldots, b_n) \in f_{>0}$. Then $a_i \neq 0$ and $b_j \neq 0$ for some i, j where $1 \leq i,\ j \leq r - 1$. Note that $u = (a_1, \ldots, a_n)$ is path-connected to $u' = (a_1, \ldots, a_{r-1}, 0, \ldots, 0)$ by a line segment that lies in $f_{>0}$, and $v = (b_1, \ldots, b_n)$ is path-connected to $v' = (b_1, \ldots, b_{r-1}, 0, \ldots, 0)$ by a line segment that lies in $f_{>0}$. Since $r \geq 3$, we have that $\mathbb{R}^{r-1} \backslash \{0\}$ is path-connected and thus u', v' are path-connected by a line segment that lies in $f_{>0}$. Therefore, $f_{>0}$ is path-connected.

If $3 \leq r < n$, then $f_{\geq 0}$ is path-connected by Proposition 3.10.

We can now assume that $3 \leq r = n$. Let $u = (a_1, \ldots, a_n)$, $v = (b_1, \ldots, b_n) \in f_{\geq 0}$. Since $r = n$, we have $a_i, b_j \neq 0$ for some i, j where $1 \leq i,\ j \leq r - 1$. Note that $u = (a_1, \ldots, a_n)$ is path-connected to $u' = (a_1, \ldots, a_{r-1}, 0)$ by a line segment that lies in $f_{\geq 0}$, and $v = (b_1, \ldots, b_n)$ is path-connected to $v' = (b_1, \ldots, b_{r-1}, 0)$ by a line segment that lies in $f_{\geq 0}$.

Since $r \geq 3$, we have that $\mathbb{R}^{r-1} \backslash \{0\}$ is path-connected and thus u', v' are path-connected by a line segment that lies in $f_{\geq 0}$. Therefore, $f_{\geq 0}$ is path-connected.

(c) Let $u = (a_1, \ldots, a_n), v = (b_1, \ldots, b_n) \in f_{\leq 0}$. Then $a_r \neq 0$ and $b_r \neq 0$ because $r = n$. First assume that a_r, b_r have opposite signs. A path from u to v must pass through some point $w = (c_1, \ldots, c_{r-1}, 0)$ where $f(w) > 0$. Thus u and v are not path-connected in $f_{\leq 0}$.

Now suppose that a_r, b_r have the same signs. Note that u and $u' = (0, \ldots, 0, a_r)$ are path-connected by a line segment that lies in $f_{\leq 0}$, and v and $v' = (0, \ldots, 0, b_r)$ are path-connected by a line segment that lies in $f_{\leq 0}$. Since a_r, b_r have the same signs, it follows that u', v' are path-connected by a line segment that lies in $f_{\leq 0}$. Therefore, u, v are path-connected in $f_{\leq 0}$. Since $u \in f_{\leq 0}$ if and only if $-u \in f_{\leq 0}$, it follows that $f_{\leq 0}$ is the union of two path-connected components $\mathcal{B}_1$ and $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$.

(d) Let $u = (a_1, \ldots, a_n), v = (b_1, \ldots, b_n) \in f_{<0}$. Then $a_r \neq 0$ and $b_r \neq 0$. First assume that a_r, b_r have opposite signs. A path from u to v must

pass through some point $w = (c_1, \ldots, c_{r-1}, 0, c_{r+1}, \ldots, c_n)$ where $f(w) \geq 0$. Thus u and v are not path-connected in $f_{<0}$.
Now suppose that a_r, b_r have the same signs. Note that u and $u' = (0, \ldots, 0, a_r, 0, \ldots, 0)$ are path-connected by a line segment that lies in $f_{<0}$, and v and $v' = (0, \ldots, 0, b_r, 0, \ldots, 0)$ are path-connected by a line segment that lies in $f_{<0}$. Since u' is path-connected to v' by a line segment that lies in $f_{<0}$, it follows that u, v are path-connected in $f_{<0}$. Since $u \in f_{<0}$ if and only if $-u \in f_{<0}$, it follows that $f_{<0}$ is the union of two path-connected components $\mathcal{B}_1$ and $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$.

(3) (a) This case follows from Proposition 3.10.

(b) First, assume $2 = r = n$. Then $f = X_1^2 - X_2^2$. The four path-connected components are given by $X_2 = \pm X_1,\ X_1 > 0$, and $X_2 = \pm X_1,\ X_1 < 0$.
Now assume that $3 \leq r = n$. Let $u = (a_1, \ldots, a_r),\ v = (b_1, \ldots, b_r) \in f_{=0}$. Then $a_r \neq 0,\ b_r \neq 0$, and $a_i \neq 0,\ b_j \neq 0$ for some $1 \leq i, j \leq r-1$. Suppose that a_r and b_r have opposite signs. Then a path from u to v in $f_{=0}$ would pass through a nonzero vector of the form $w = (c_1, \ldots, c_{r-1}, 0)$ where $f(w) > 0$. Therefore, u and v are not path-connected in $f_{=0}$.
Now suppose that a_r and b_r have the same signs. There exist $c, d \in \mathbb{R}_{>0}$ such that $u' = c(a_1, \ldots, a_{r-1}), v' = d(b_1, \ldots, b_{r-1}) \in \mathbb{S}^{r-2}$. Then $u'' = \left(u', \dfrac{a_r}{|a_r|}\right)$ and $v'' = \left(v', \dfrac{b_r}{|b_r|}\right)$ lie in $f_{=0}$. Note that $\dfrac{a_r}{|a_r|} = \dfrac{b_r}{|b_r|} = \pm 1$.
We have that u is path-connected to u'' by a line segment that lies in $f_{=0}$ and v is path-connected to v'' by a line segment that lies in $f_{=0}$. Next, u'' is path-connected to v'' by a path that lies in $f_{=0}$ by Lemma 3.9 because either $u'', v'' \in \mathbb{S}^{r-2} \times \{1\}$ or $u'', v'' \in \mathbb{S}^{r-2} \times \{-1\}$. Thus u is path-connected to v in $f_{=0}$.
Therefore, u is path-connected to v in $f_{=0}$ if and only if a_r, b_r have the same sign. It follows that $f_{=0}$ is the union of two path-connected components $\mathcal{B}_1$ and $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$.

□

Proposition 3.12. *Let* $f = X_1^2 + \cdots + X_k^2 - X_{k+1}^2 - \cdots - X_r^2 \in \mathbb{R}^n[X_1, \ldots X_n]$ *such that* $\operatorname{rank}(f) = r \leq n$. *Assume that* $k \geq 2,\ r - k \geq 2$. *Then* $f_{>0}$, $f_{\geq 0}$, $f_{<0}$, $f_{\leq 0}$, *and* $f_{=0}$ *are each path-connected.*

Proof. First, we show that $f_{>0}$ is path-connected. Let $u = (a_1, \ldots, a_n),\ v = (b_1, \ldots, b_n)$ be points in $f_{>0}$. This implies that $a_i \neq 0, b_j \neq 0$ for some $1 \leq i, j \leq k$. There exist $c, d \in \mathbb{R}_{>0}$ such that letting

$$u' = (ca_1, \ldots, ca_k, 0, \ldots, 0),\ v' = (db_1, \ldots, db_k, 0, \ldots, 0),$$

we have $u', v' \in \mathbb{S}^{k-1} \times \{0\}^{n-k} \subset f_{>0}$, where $\mathbb{S}^{k-1}$ is the unit sphere in $\mathbb{R}^k$.

Since $\mathbb{S}^{k-1}$ is path-connected when $k \geq 2$, u' and v' are path-connected in $f_{>0}$. Since u is path-connected to u' by a line segment that lies in $f_{>0}$, and v' is path-connected to v by a line segment that lies in $f_{>0}$, it follows that u is path-connected to v in $f_{>0}$ and hence $f_{>0}$ is path-connected. It follows that $f_{<0} = (-f)_{>0}$ is also path-connected.

If $r < n$, then $\operatorname{rad}(f) \neq 0$ and hence $f_{\geq 0}$, $f_{\leq 0}$, and $f_{=0}$ are each path-connected by Proposition 3.10.

Now assume that $r = n$. We next show that $f_{=0}$ is path-connected. Let u, v be points in $f_{=0}$. Write $u = (p, q)$ and $v = (r, s)$ where $p, r \in \mathbb{R}^k \backslash 0$ and $q, s \in \mathbb{R}^{n-k} \backslash 0$. There exist $c, d \in \mathbb{R}_{>0}$ such that

$$(cp, cq), (dr, ds) \in \mathbb{S}^{k-1} \times \mathbb{S}^{n-k-1} \subset f_{=0}.$$

Note that u is path-connected to (cp, cq) by a line segment that lies in $f_{=0}$, and v is path-connected to (dr, ds) by a line segment that lies in $f_{=0}$. Since $k \geq 2$ and $n-k \geq 2$, it follows that $\mathbb{S}^{k-1} \times \mathbb{S}^{n-k-1}$ is path-connected. Thus (cp, cq) is path-connected to (dr, ds) in $\mathbb{S}^{k-1} \times \mathbb{S}^{n-k-1}$. It follows that u is path-connected to v in $f_{=0}$, and thus $f_{=0}$ is path-connected.

We now show that $f_{\geq 0}$ is path-connected. We have $e_1 \in f_{>0}$ and $e_1 + e_{k+1} \in f_{=0}$. The line segment joining e_1 and $e_1 + e_{k+1}$ lies in $f_{\geq 0}$. Since $f_{>0}$ and $f_{=0}$ are each path-connected and there is a path in $f_{\geq 0}$ joining e_1 and $e_1 + e_{k+1}$, it follows that $f_{>0} \cup f_{=0} = f_{\geq 0}$ is path-connected.

It follows that $f_{\leq 0} = (-f)_{\geq 0}$ is also path-connected. $\square$

The parts of Proposition 3.11 that are needed for the proofs of Proposition 4.1 and Proposition 4.4 are parts 1(a), 1(c), 2(b), 2(d), and 3.

4. Two Quadratic Forms over the Real Numbers

Our proof of Proposition 4.4 below is slightly different from the proof given in [7, Lemma 1 (i)]. We first give Swinnerton-Dyer's proof of [7, Lemma 1 (ii)] in Proposition 4.1 and then use this result to give a simpler proof of [7, Lemma 1 (i)] in Proposition 4.4. We have added many details not included in Swinnerton-Dyer's exposition.

Proposition 4.1. *Let $f, g \in \mathbb{R}[X_1, \ldots, X_n]$ be quadratic forms with $n \geq 2$ and assume that f is indefinite. Then there exist real zeros v, w on $f_{=0}$ such that $g(v) > 0$ and $g(w) < 0$ if and only if $\lambda f + g$ is indefinite for all $\lambda \in \mathbb{R}$.*

Proof. If there exist real zeros v, w on $f_{=0}$ such that $g(v) > 0$ and $g(w) < 0$, then $\lambda f + g$ is indefinite for all $\lambda \in \mathbb{R}$ because $(\lambda f + g)(v) > 0$ and $(\lambda f + g)(w) < 0$.

Now assume that $\lambda f + g$ is indefinite for all $\lambda \in \mathbb{R}$. Suppose that $g \geq 0$ on $f_{=0}$. We will obtain a contradiction. The case $g \leq 0$ on $f_{=0}$ is handled by replacing g with $-g$ and noting that $-g \geq 0$ on $f_{=0}$.

The set $(\lambda f + g)_{<0}$ does not meet $f_{=0}$ for any real λ because $f_{=0}$ lies entirely in $g \geq 0$. For any $\lambda \in \mathbb{R}$, $(\lambda f + g)_{<0}$ is a non-empty, open set. Since $\lambda f + g$ is indefinite, its rank is at least 2. Proposition 3.11, Proposition 3.12, and the comments above Proposition 3.10 imply that the set $(\lambda f + g)_{<0}$ is either path-connected or has two path-connected components $\mathcal{B}_1$, $\mathcal{B}_2$ where $\mathcal{B}_2 = (-1)\mathcal{B}_1$. Since $f_{>0}$ and $f_{<0}$ are disjoint open sets, it follows that each path-connected component lies entirely in either $f_{>0}$ or $f_{<0}$. Since f is a quadratic form, $f(v) = f(-v)$ for all $v \in \mathbb{R}^n$, so it follows that if $(\lambda f + g)_{<0}$ has two path-connected components, then both path components lie entirely in either $f_{>0}$ or $f_{<0}$. Thus $(\lambda f + g)_{<0}$ lies entirely in either $f_{>0}$ or $f_{<0}$.

Define

$$\Lambda_1 = \left\{ \lambda \in \mathbb{R} \;\middle|\; \lambda f + g < 0 \text{ lies in } f > 0 \right\},$$

and

$$\Lambda_2 = \left\{ \lambda \in \mathbb{R} \;\middle|\; \lambda f + g < 0 \text{ lies in } f < 0 \right\}.$$

Then Λ_1 and Λ_2 are disjoint and $\Lambda_1 \cup \Lambda_2 = \mathbb{R}$.

Claim. *Λ_1 and Λ_2 are non-empty subsets of $\mathbb{R}$.*

Since f is indefinite, there exist $v, u \in \mathbb{R}^n \backslash 0$ such that $f(v) > 0$ and $f(u) < 0$. Choose $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_1 < \frac{-g(v)}{f(v)}$ and $\lambda_2 > \frac{-g(u)}{f(u)}$. Then $\lambda_1 f(v) + g(v) < 0$ and so $\lambda_1 \in \Lambda_1$. Similarly, $\lambda_2 f(u) + g(u) < 0$, and so $\lambda_2 \in \Lambda_2$.

Claim. *Λ_1 and Λ_2 are open sets in $\mathbb{R}$.*

Let $\lambda \in \Lambda_1$. Then there exists $v \in \mathbb{R}^n$ such that $\lambda f(v) + g(v) < 0$ and $f(v) > 0$. This implies that $\lambda < \frac{-g(v)}{f(v)}$. If $\lambda' < \frac{-g(v)}{f(v)}$, then $\lambda' f(v) + g(v) < 0$. Hence, $\lambda \in \left(-\infty, \dfrac{-g(v)}{f(v)}\right) \subseteq \Lambda_1$.

Similarly, for $\lambda \in \Lambda_2$ there exists $u \in \mathbb{R}^n$ such that $\lambda f(u) + g(u) < 0$ and $f(u) < 0$. This implies that $\lambda > \frac{-g(u)}{f(u)}$. If $\lambda' > \frac{-g(u)}{f(u)}$, then $\lambda' f(u) + g(u) < 0$. Hence, $\lambda \in \left(\dfrac{-g(u)}{f(u)}, \infty\right) \subseteq \Lambda_2$. This proves the claim.

The previous two claims show that $\mathbb{R}$ can be written as a disjoint union of two non-empty open sets, which is a contradiction. Therefore, there exist real zeros on $f_{=0}$ that give either sign to g. $\square$

Remark 4.2. In Proposition 4.1, suppose that f is a positive semi-definite, but not a definite quadratic form. If there exist real zeros on $f_{=0}$ that give either sign to g, then $\lambda f + g$ is indefinite for all $\lambda \in \mathbb{R}$. However, the converse statement is not true, as the next example shows.

Example 4.3. Let $f = X_1^2 + X_2^2$ and $g = X_1X_3 + X_2X_4$ be quadratic forms in $\mathbb{R}[X_1, X_2, X_3, X_4]$. Then f is positive semi-definite and g is indefinite. For any nonzero $\lambda \in \mathbb{R}$,

$$\lambda f + g = \lambda X_1^2 + \lambda X_2^2 + X_1X_3 + X_2X_4$$

is indefinite because

$$(\lambda f + g)(1, 1, 0, 0) = 2\lambda \text{ and } (\lambda f + g)(1, 1, -2\lambda, -2\lambda) = -2\lambda$$

are opposite in signs when $\lambda \neq 0$. Therefore, $\lambda f + g$ is indefinite for all $\lambda \in \mathbb{R}$, but $g|_{f_{=0}} = 0$.

Proposition 4.4. *Let $f, g \in \mathbb{R}[X_1, \dots, X_n]$ be quadratic forms with $n \geq 3$. Then the following statements are equivalent.*

(1) *The set $f = g = 0$ contains a nontrivial real zero.*
(2) *$\lambda f + \mu g$ is not definite for any real λ, μ, not both zero.*
(3) *For every real λ, μ, not both zero, $\lambda f + \mu g$ has a nontrivial real zero.*

Proof. A definite quadratic form has no nontrivial real zero. A quadratic form that is not definite is either semi-definite or indefinite, and in each case, the quadratic form has a nontrivial real zero. Thus (2) and (3) are equivalent.

(1) $\Rightarrow$ (2). If $f = g = 0$ has a nontrivial real zero, then $\lambda f + \mu g$ is not definite for any real λ, μ, not both zero.

(2) $\Rightarrow$ (1). Suppose that $\lambda f + \mu g$ is not definite for any real λ, μ not both zero. We have the following two cases:

Case 1. Suppose there exists a semi-definite form in $\mathcal{P}_{\mathbb{R}}(f, g)$. Without loss of generality, we may assume that f is a positive semi-definite quadratic form. Since f is not definite, after an invertible linear transformation, we may assume that

$$f(X_1, \dots, X_n) = X_1^2 + \cdots + X_r^2,$$

where $r < n$ is the rank of f, and

$$g = \sum_{1 \leq i \leq j \leq n} a_{ij} X_i X_j.$$

Suppose first that $g(0, \dots, 0, X_{r+1}, \dots, X_n)$ has no nontrivial zero over $\mathbb{R}$. Then $g(0, \dots, 0, X_{r+1}, \dots, X_n)$ is definite, and by replacing g with $-g$ if necessary, we can assume that $g(0, \dots, 0, X_{r+1}, \dots, X_n)$ is positive definite.

For $\lambda \in \mathbb{R}$, consider the symmetric matrix corresponding to $\lambda f + g$:

$$\begin{array}{cc} & \begin{array}{cc} \qquad\qquad r \qquad\qquad & n-r \end{array} \\ \begin{array}{c} r \\ \\ n-r \end{array} & \left(\begin{array}{ccc|c} \lambda+\alpha_{11} & & * & \\ & \ddots & & * \\ * & & \lambda+\alpha_{rr} & \\ \hline & * & & * \end{array}\right) \end{array}$$

Since $g(0,\ldots,0,X_{r+1},\ldots,X_n)$ is positive definite, the $n-r$ leading principal minors starting from the lower right corner of the above matrix are positive by Proposition 3.6. Since λ appears only in the diagonal entries, we can choose λ_0 large enough so that all leading principal minors starting from the lower right corner are positive. Hence, Proposition 3.6 implies that $\lambda_0 f + g$ is a positive definite quadratic form, which is a contradiction.

Thus $g(0,\ldots,0,X_{r+1},\ldots,X_n)$ has a nontrivial zero over $\mathbb{R}$. This zero is then a nontrivial common zero over $\mathbb{R}$ of f, g.

Case 2. Assume that every form in $\mathcal{P}_{\mathbb{R}}(f,g)$ is indefinite. Since f is indefinite and $n \geq 3$, Proposition 3.11, Proposition 3.12, and the comments above Proposition 3.10 imply that $f_{=0}$ is either path-connected or is a union of two path-connected components of the form $\mathcal{B}_1 \cup \mathcal{B}_2$ where $\mathcal{B}_2 = -\mathcal{B}_1$.

Proposition 4.1 implies that there exist $u, v \in f_{=0}$ such that $g(u) > 0$ and $g(v) < 0$. We can assume that u, v lie in the same path-connected component $\mathcal{B}$ of $f_{=0}$ because v can be replaced by $-v$ if necessary.

Consider a path $\gamma : [0,1] \to \mathcal{B}$ where γ is a continuous function satisfying $\gamma(0) = u$ and $\gamma(1) = v$. Since $g : \mathcal{B} \to \mathbb{R}$ is continuous, we have $g \circ \gamma : [0,1] \to \mathbb{R}$ is a continuous function. The image of $g \circ \gamma$ is connected because the continuous image of a connected space is connected. Since $g\circ\gamma(0) = g(u) > 0$ and $g\circ\gamma(1) = g(v) < 0$, it follows that there exists $c \in (0,1)$ such that $g \circ \gamma(c) = 0$. Thus $\gamma(c) \in \mathcal{B} \cap g_{=0} \subseteq f_{=0} \cap g_{=0}$.

□

Example 4.5. Let $f = 2X_1X_2$, $g = X_1^2 - X_2^2$. The pair f, g has no nontrivial zeros over $\mathbb{R}$, or even over $\mathbb{C}$. Every form in the $\mathbb{R}$-pencil is indefinite. This shows that the hypothesis $n \geq 3$ in Proposition 4.4 is necessary.

Example 4.6. This example shows that condition (2) in Proposition 4.4 is not the same as the condition that $\lambda f + \mu g$ is indefinite for every real λ, μ, not both zero. In particular, it really is necessary to consider Case 1 in the proof that (2) $\Rightarrow$ (1). Let $1 \leq j \leq n-2$ and let

$$f = X_1^2 + \cdots + X_j^2$$

$$g = h(X_1,\ldots,X_j) + X_{j+1}^2 + \cdots + X_{n-1}^2 - X_n^2,$$

where h is any quadratic form with real coefficients. Then f is positive semi-definite, but not definite, and $\lambda f + \mu g$ is indefinite for every real λ, μ with $\mu \neq 0$. Also, $(0,\ldots,0,1,1)$ is a real nontrivial common zero of f, g.

Remark 4.7. The proof of Proposition 4.4 is motivated by [7, Lemma 1 (i)]. However, the proof given in [7, Lemma 1 (i)] did not consider Case 1 in the proof of (2) $\Rightarrow$ (1). The argument given in the proof of [7, Lemma 1 (i)] fails when f is semi-definite.

5. Signature of a Quadratic Form

Let $f, g \in \mathbb{R}[X_1, \ldots, X_n]$ be two quadratic forms and let M_f, M_g denote the symmetric matrices corresponding to f, g, respectively. Let

$$D(\lambda, \mu) = \det(\lambda f + \mu g) = \det(\lambda M_f + \mu M_g),$$

and call this the determinant polynomial of f, g. Either $D(\lambda, \mu) = 0$ or $D(\lambda, \mu)$ is a nonzero homogeneous form of degree n in the variables λ, μ. See Example 6.6 below for an example where $D(\lambda, \mu) = 0$. Let

$$T = \{(\lambda, \mu) \in \mathbb{S}^1 \subset \mathbb{R}^2 \mid \det(\lambda f + \mu g) = 0\}.$$

Lemma 5.1. *Assume that $D(\lambda, \mu) = \det(\lambda f + \mu g)$ is a nonzero homogeneous form in the variables λ, μ. Then $|T| \leq 2n$.*

Proof. The hypothesis implies that $D(\lambda, \mu) = \det(\lambda f + \mu g)$ is a nonzero homogeneous form of degree n in the variables λ, μ. Since $D(\lambda, \mu)$ has at most n distinct linear factors defined over $\mathbb{R}$, the equation $D(\lambda, \mu) = 0$ has at most $2n$ distinct zeros on $\mathbb{S}^1$ because each linear factor $a\lambda + b\mu$ gives exactly two zeros $\left(\dfrac{-b}{\sqrt{a^2+b^2}}, \dfrac{a}{\sqrt{a^2+b^2}}\right)$ and $\left(\dfrac{b}{\sqrt{a^2+b^2}}, \dfrac{-a}{\sqrt{a^2+b^2}}\right)$ of $D(\lambda, \mu)$ on $\mathbb{S}^1$. Therefore, $|T| \leq 2n$. $\square$

Next, we define the signature map

$$\begin{aligned} \mathrm{Sgn} : \mathbb{S}^1 &\to \mathbb{Z} \\ (\lambda, \mu) &\mapsto \mathrm{sgn}(\lambda f + \mu g) \end{aligned}$$

For any $n \times n$ matrix M and integer k with $1 \leq k \leq n$, let $M^{(k)}$ denote the upper left $k \times k$ sub-matrix of M, and let $d_k = \det(M^{(k)})$.

In the following, we give the discrete topology to $\mathbb{Z}$, which is the same as the subspace topology inherited from the standard topology on $\mathbb{R}$.

Proposition 5.2. *Assume that $D(\lambda, \mu)$ is nonzero. The signature map* Sgn *is constant on each connected component of $\mathbb{S}^1 - T$ and thus* Sgn *is continuous at all points of $\mathbb{S}^1$ except for the finitely many points that lie in $T \subset \mathbb{S}^1$.*

Proof. The set T is finite by Lemma 5.1. Let $(\lambda_0, \mu_0) \in \mathbb{S}^1 - T$. Then $\lambda_0 f + \mu_0 g$ is a nonsingular quadratic form. Since $\mathrm{rad}(\lambda_0 f + \mu_0 g) = 0$, we can write $\mathbb{R}^n = V \oplus W$ where $(\lambda_0 f + \mu_0 g)(v) > 0$ for all nonzero $v \in V$, $(\lambda_0 f + \mu_0 g)(w) < 0$ for all nonzero $w \in W$, $\dim(V) = r$, $\dim(W) = s$, $\mathrm{sgn}(\lambda_0 f + \mu_0 g) = r - s$. The subspaces V, W are not uniquely determined, but $\dim(V), \dim(W)$ are uniquely determined by Proposition 3.4 and its proof. Let $\{v_1, \ldots, v_r\}$ be a basis of V and $\{v_{r+1}, \ldots, v_{r+s}\}$ a basis of W. Let

$$S_V = \{a_1 v_1 + \cdots + a_r v_r \in V \mid a_1^2 + \cdots + a_r^2 = 1\},$$

$$S_W = \{a_{r+1} v_{r+1} + \cdots + a_n v_n \in W \mid a_{r+1}^2 + \cdots + a_n^2 = 1\}.$$

We take S_V to be the empty set if $r = 0$, and similarly for S_W if $s = 0$. The sets S_V, S_W are compact subsets of $\mathbb{R}^r$, $\mathbb{R}^s$, respectively.

The function $\tau_V : S_V \times \mathbb{S}^1 \to \mathbb{R}$ defined by $\tau_V(v, \lambda, \mu) = v^t M_{\lambda f + \mu g} v - v^t M_{\lambda_0 f + \mu_0 g} v$ is a polynomial function of the entries of v and of λ, μ, and thus a continuous function, and similarly for the corresponding function $\tau_W : S_W \times \mathbb{S}^1 \to \mathbb{R}$. Since $S_V \times \mathbb{S}^1$ and $S_W \times \mathbb{S}^1$ are compact subsets of the metric spaces $\mathbb{R}^r \times \mathbb{R}^2$, $\mathbb{R}^s \times \mathbb{R}^2$, respectively, uniform continuity implies that for every $\varepsilon > 0$, there exists $\delta > 0$ such that if (λ, μ) lies in U_δ, the open neighborhood around (λ_0, μ_0) of radius δ, then

$$|v^t M_{\lambda f + \mu g} v - v^t M_{\lambda_0 f + \mu_0 g} v| < \varepsilon,$$

for every $v \in S_V$, with a similar statement holding for every $v \in S_W$.

If S_V is nonempty, let

$$A_V = \min_{v \in S_V} \{|v^t M_{\lambda_0 f + \mu_0 g} v|\},$$

with A_W defined similarly if S_W is nonempty. Note that if S_V is nonempty, then $A_V > 0$ because S_V is compact, and similarly, $A_W > 0$ if S_W is nonempty.

Let $\varepsilon = \min\{A_V, A_W\}$ if S_V and S_W are both nonempty. Otherwise, let $\varepsilon = A_V$ if only S_V is nonempty, and let $\varepsilon = A_W$ if only S_W is nonempty.

Since $|v^t M_{\lambda_0 f + \mu_0 g} v| \geq \varepsilon$ for all $v \in S_V \cup S_W$, and since for all $(\lambda, \mu) \in U_\delta$, we have $|v^t M_{\lambda f + \mu g} v - v^t M_{\lambda_0 f + \mu_0 g} v| < \varepsilon$, it follows that $v^t M_{\lambda_0 f + \mu_0 g} v$ and $v^t M_{\lambda f + \mu g} v$ have the same sign for all $v \in S_V$ and $(\lambda, \mu) \in U_\delta$, with a similar statement for all $v \in S_W$.

Therefore for all $(\lambda, \mu) \in U_\delta$, we have $(\lambda f + \mu g)(v) > 0$ for all $v \in S_V$ and thus also for all $v \in V$, and $(\lambda f + \mu g)(v) < 0$ for all $v \in S_W$ and thus also for all $v \in W$. It follows that the decomposition $\mathbb{R}^n = V \oplus W$ can be used to compute both $\operatorname{sgn}(\lambda_0 f + \mu_0 g)$ and $\operatorname{sgn}(\lambda f + \mu g)$ and this gives

$$\operatorname{Sgn}(\lambda_0, \mu_0) = \dim(V) - \dim(W) = \operatorname{Sgn}(\lambda, \mu),$$

for all $(\lambda, \mu) \in U_\delta$. This shows that $\operatorname{Sgn} : \mathbb{S}^1 - T \to \mathbb{Z}$ is a locally constant function. Since $\mathbb{Z}$ is given the discrete topology it follows that Sgn is continuous on all the points in $\mathbb{S}^1 - T$. If $\mathcal{C}_i$ is a connected component of $\mathbb{S}^1 - T$, then $\operatorname{Sgn}(\mathcal{C}_i)$ is also connected. Since the only connected sets in $\mathbb{Z}$ are singleton sets, we see that $\operatorname{Sgn}(\mathcal{C}_i)$ is a constant. $\square$

Proposition 5.3. *Assume that $D(\lambda, \mu) \neq 0$. For $(\lambda, \mu) \in \mathbb{S}^1$, the signature of the quadratic form $\lambda f + \mu g$ changes only as we pass through a point T on $\mathbb{S}^1$ and it changes by at most twice the dimension of the radical of the form.*

Proof. The proof of the first part of this proposition follows from Proposition 5.2. We now show that as we pass through a point (λ_0, μ_0) in T on $\mathbb{S}^1_{\mu>0}$, the signature changes by at most twice the dimension of the radical of the form $\lambda_0 f + \mu_0 g$. Let $\operatorname{rank}(\lambda_0 f + \mu_0 g) = r < n$. Without loss of generality, we may assume that $\lambda_0 f + \mu_0 g \in \mathbb{R}[X_1, \ldots, X_r]$. Let $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_s$ denote the connected components of $\mathbb{S}^1 - T$. Proposition 5.2 implies that Sgn is constant on each $\mathcal{C}_i$. Let $\mathcal{C}_1, \mathcal{C}_2$ be the two consecutive components such that (λ_0, μ_0) is the point of singularity that disconnects $\mathcal{C}_1$ and $\mathcal{C}_2$ in $\mathbb{S}^1$. The form $\lambda f + \mu g$ is nonsingular for all $(\lambda, \mu) \in \mathcal{C}_1 \cup \mathcal{C}_2$. For all $(\lambda, \mu) \in \{\mathcal{C}_1 \cup \mathcal{C}_2 \cup (\lambda_0, \mu_0)\}$, $\lambda_0 f + \mu_0 g$ and $\left(\lambda f + \mu g\big|_{X_{r+1} = \cdots = X_n = 0}\right)$ are quadratic forms in r variables, and in this case $\lambda_0 f + \mu_0 g$ is nonsingular when considered as a form in r variables. We define the following map which is the restriction of Sgn defined above.

$$\begin{aligned} \operatorname{Sgn}_1 : \mathcal{C}_1 \cup \mathcal{C}_2 \cup \{(\lambda_0, \mu_0)\} &\to \mathbb{Z} \\ (\lambda, \mu) &\mapsto \operatorname{sgn}\left(\lambda f + \mu g\big|_{X_{r+1} = \cdots = X_n = 0}\right) \end{aligned}$$

From Proposition 5.2, we know that Sgn_1 is a locally constant map at points $(\lambda, \mu) \in \mathcal{C}_1 \cup \mathcal{C}_2 \cup \{(\lambda_0, \mu_0)\}$ where $\lambda f + \mu g\big|_{X_{r+1} = \cdots = X_n = 0}$ is a rank r quadratic form in the variables $X_1, \ldots, X_r$. Since $\lambda_0 f + \mu_0 g$ is a nonsingular form in r variables, we can find $\varepsilon > 0$ such that

$$\operatorname{Sgn}_1(\lambda, \mu) = \operatorname{Sgn}_1(\lambda_0, \mu_0) = \operatorname{Sgn}(\lambda_0, \mu_0)$$

for all $(\lambda, \mu) \in \mathcal{B}_\varepsilon(\lambda_0, \mu_0)$ in $\mathbb{S}^1$. Choose $(\lambda, \mu) \in \mathcal{B}_\varepsilon$ different from (λ_0, μ_0). After performing row and column operations on the symmetric matrix $M_{\lambda f + \mu g}$, it can be written in the form

$$\begin{array}{cc} & \begin{array}{cc} r & n-r \end{array} \\ \begin{array}{c} r \\ \\ n-r \end{array} & \left(\begin{array}{c|c} \begin{matrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_r \end{matrix} & 0 \\ \hline 0 & B \end{array}\right) \end{array}$$

As observed from the above matrix,

$$\operatorname{Sgn}(\lambda,\mu) = \operatorname{Sgn}_1(\lambda,\mu) + \operatorname{sgn}(B) = \operatorname{Sgn}(\lambda_0,\mu_0) + \operatorname{sgn}(B).$$

Since $|\operatorname{sgn}(B)| \le n-r$, we obtain

$$|\operatorname{Sgn}(\lambda,\mu) - \operatorname{Sgn}(\lambda_0,\mu_0)| \le n-r.$$

Choose $(\lambda_1,\mu_1) \in \mathcal{C}_1$ and $(\lambda_2,\mu_2) \in \mathcal{C}_2$ such that (λ_1,μ_1) and (λ_2,μ_2) lie in $\mathcal{B}_\varepsilon$. Then,

$$\begin{aligned} & |\operatorname{Sgn}(\lambda_1,\mu_1) - \operatorname{Sgn}(\lambda_2,\mu_2)| \\ &= |\operatorname{Sgn}(\lambda_1,\mu_1) - \operatorname{Sgn}(\lambda_0,\mu_0) + \operatorname{Sgn}(\lambda_0,\mu_0) - \operatorname{Sgn}(\lambda_2,\mu_2)| \\ &\le |\operatorname{Sgn}(\lambda_1,\mu_1) - \operatorname{Sgn}(\lambda_0,\mu_0)| + |\operatorname{Sgn}(\lambda_0,\mu_0) - \operatorname{Sgn}(\lambda_2,\mu_2)| \\ &\le (n-r) + (n-r) = 2(n-r) \end{aligned}$$

This finishes the proof of the Proposition. □

Example 5.4. Let

$$\begin{aligned} f &= X_1^2 + \cdots + X_m^2 + a_{m+1}X_{m+1}^2 + \cdots + a_nX_n^2 \\ g &= b_{m+1}X_{m+1}^2 + \cdots + b_nX_n^2. \end{aligned}$$

Let $\operatorname{sgn}(g) = c$. Let $\varepsilon > 0$ be a real number. Then for sufficiently small ε, we have $\operatorname{sgn}(g+\varepsilon f) = c+m$ and $\operatorname{sgn}(g-\varepsilon f) = c-m$. Then the difference of the two signatures, which is $2m$, equals two times the dimension of the radical of g.

6. Forms in the pencil containing many hyperbolic planes

Recall that K denotes an arbitrary field with characteristic not 2. Let $H(X,Y) \in K[X,Y]$ be a homogeneous form of degree $n \ge 1$. Then H factors in $K^{alg}[X,Y]$ as a product of linear factors and we can write $H(X,Y) = \prod_{i=1}^r L_i(X,Y)^{e_i}$ where each $e_i \ge 1$, $L_i = \alpha_iX + \beta_iY \in K^{alg}[X,Y]$ is a linear form, $1 \le i \le r$, and $L_1,\dots,L_r$ are distinct in the sense that if $i \ne j$, then L_i and L_j are not scalar multiples of each other. This is the same as saying that L_i and L_j are linearly independent if $i \ne j$.

Suppose that $\gamma,\delta \in K^{alg}$, $(\gamma,\delta) \ne (0,0)$, and $H(\gamma,\delta) = 0$. Then $L_i(\gamma,\delta) = 0$ for some unique value of i. In general, we say that (γ,δ) is a zero of $H(X,Y)$ of multiplicity e if $L(\gamma,\delta) = 0$ for some linear form L where L^e is the exact power of L dividing H.

Lemma 6.1. *Let f, $g \in K[X_1,\dots,X_n]$ be quadratic forms. Assume that the homogeneous polynomial $D(\lambda,\mu) = \det(\lambda f + \mu g)$ of degree n does not vanish identically on K^{alg}. If (λ_0,μ_0) is a zero of $D(\lambda,\mu)$ over K^{alg} of multiplicity m and r is the rank of the quadratic form $\lambda_0 f + \mu_0 g$, then $m \ge n-r$.*

Proof. Since the homogeneous polynomial $D(\lambda,\mu)$ does not vanish on K^{alg}, $D(\lambda,\mu)$ has only finitely many linear forms (up to scalar multiplication) that occur as factors of D. Let (λ_0,μ_0) be a nontrivial zero of $D(\lambda,\mu)$. We can assume that $\mu_0 \ne 0$. After an invertible linear change of variables, we can diagonalize and rewrite $\lambda_0 f + \mu_0 g$ as

$$\lambda_0 f + \mu_0 g = b_1X_1^2 + \cdots + b_rX_r^2,$$

where $\operatorname{rank}(\lambda_0 f + \mu_0 g) = r < n$. Then

$$\begin{aligned}\lambda f + \mu g &= \lambda f + \mu \frac{\lambda_0 f + \mu_0 g - \lambda_0 f}{\mu_0} \\ &= \left(\lambda - \frac{\lambda_0}{\mu_0}\mu\right) f + \frac{\mu}{\mu_0}(b_1 X_1^2 + \cdots + b_r X_r^2)\end{aligned}$$

Let M denote the symmetric matrix corresponding to the quadratic form $\lambda f + \mu g$. Then $D(\lambda, \mu) = \det(M)$, where the matrix M is as shown below and $\alpha = \left(\lambda - \dfrac{\lambda_0}{\mu_0}\mu\right)$:

$$\begin{array}{cc} & \begin{array}{cc} \qquad r \qquad & \qquad n-r \qquad \end{array} \\ \begin{array}{c} r \\ \\ n-r \end{array} & \left(\begin{array}{ccc|ccc} \alpha a_{11} + \frac{\mu b_1}{\mu_0} & & * & & & \\ & \ddots & & & * & \\ * & & \alpha a_{rr} + \frac{\mu b_r}{\mu_0} & & & \\ \hline & & & \alpha a_{r+1,r+1} & & * \\ & * & & & \ddots & \\ & & & * & & \alpha a_{nn} \end{array}\right) \end{array}$$

Each term of the last $n-r$ rows of M contains a factor of $\alpha = \left(\lambda - \dfrac{\lambda_0}{\mu_0}\mu\right)$. This implies that $\left(\lambda - \dfrac{\lambda_0}{\mu_0}\mu\right)^{n-r}$ divides $D(\lambda, \mu)$ over K^{alg}. Thus the linear factor $(\mu_0\lambda - \lambda_0\mu)$ appears at least $n-r$ times in the linear factor decomposition of $D(\lambda,\, \mu)$ over K^{alg}. Therefore, $m_{(\lambda_0,\mu_0)} \geq n - r$. $\square$

For $x \in \mathbb{R}$, recall that $\lceil x \rceil$ denotes the least integer greater than or equal to x.

Lemma 6.2. *Let $f,\, g \in K[X_1, \ldots, X_n]$ be quadratic forms such that the determinant polynomial $D(\lambda, \mu) = \det(\lambda f + \mu g)$ over K is not identically zero. Let L be an extension of K with $K \subseteq L \subseteq K^{alg}$, and let $r \leq \left\lceil \frac{n}{2} \right\rceil$ be a positive integer. Then the following statements are equivalent.*

(a) *Every form in $\mathcal{P}_K(f, g)$ has rank at least r.*
(b) *Every form in $\mathcal{P}_{K^{alg}}(f, g)$ has rank at least r.*
(c) *Every form in $\mathcal{P}_L(f, g)$ has rank at least r.*

Proof. Since $K \subseteq L \subseteq K^{alg}$, if every form in $\mathcal{P}_{K^{alg}}(f, g)$ has rank at least r, then every form in $\mathcal{P}_L(f, g)$ has rank at least r, which further implies that every form in $\mathcal{P}_K(f, g)$ has rank at least r. This shows that (b) $\Rightarrow$ (c) $\Rightarrow$ (a). We finish the proof by showing that (a) $\Rightarrow$ (b).

Suppose that every form in $\mathcal{P}_K(f, g)$ has rank at least r, and suppose that there exists a form $\alpha f + \beta g$ in $\mathcal{P}_{K^{alg}}(f, g)$ such that

$$\operatorname{rank}(\alpha f + \beta g) \leq r - 1.$$

We can assume that either $\alpha = 1$ or $\beta = 1$ because $(\alpha, \beta) \neq (0, 0)$ and (α, β) can be multiplied by any nonzero element of K^{alg}. Assume that $\alpha = 1$. (The other case is handled similarly.) Then $(1, \beta)$ is a zero of the determinant polynomial $D(\lambda, \mu)$ and Lemma 6.1 implies that

$$m_{(1,\beta)} \geq n - (r-1) \geq n - \left(\left\lceil \frac{n}{2} \right\rceil - 1\right) = \begin{cases} \dfrac{n}{2} + 1, & \text{if } n \text{ is even} \\ \dfrac{n+1}{2}, & \text{if } n \text{ is odd} \end{cases} > \frac{n}{2}.$$

The same inequality holds for each conjugate of $(1, \beta)$. Since the degree of $D(\lambda, \mu)$ is n, it follows that $(1, \beta)$ has only one conjugate, and thus $\beta \in K$, a contradiction to (a). Hence every form in $\mathcal{P}_{K^{alg}}(f, g)$ has rank at least r. $\square$

Example 6.3. This example shows that Lemma 6.2 fails to hold if $r > \lceil \frac{n}{2} \rceil$. Let $f = X_1^2 - X_2^2$ and $g = 2X_1X_2$. Then each form in $\mathcal{P}_{\mathbb{R}}(f, g)$ has rank $2 > \lceil \frac{n}{2} \rceil$ because $\det(\lambda f + \mu g) = \det \begin{pmatrix} \lambda & \mu \\ \mu & -\lambda \end{pmatrix} = -(\lambda^2 + \mu^2)$. But over $\mathbb{C}$, there are two forms of rank 1. Namely, $f + ig = (X_1 + iX_2)^2$ and $f - ig = (X_1 - iX_2)^2$.

Proposition 6.4. *Let $f, g \in \mathbb{R}[X_1, \ldots, X_n]$ be quadratic forms and assume that at least one of f, g has rank n. Suppose that every form in $\mathcal{P}_{\mathbb{R}}(f, g)$ has rank at least r. Then there exists a rank n form in $\mathcal{P}_{\mathbb{R}}(f, g)$ that splits off at least $\left\lceil \frac{r}{2} \right\rceil$ hyperbolic planes over $\mathbb{R}$, where*

$$\left\lceil \frac{r}{2} \right\rceil = \begin{cases} \dfrac{r}{2} & \text{if } r \text{ is even} \\ \dfrac{r+1}{2} & \text{if } r \text{ is odd.} \end{cases}$$

Proof. Assume that no rank n form $\lambda f + \mu g$ in $\mathcal{P}_{\mathbb{R}}(f, g)$ splits off $\lceil \frac{r}{2} \rceil$ hyperbolic planes. Let $h \in \mathcal{P}_{\mathbb{R}}(f, g)$ be any form having rank n and suppose that h splits off exactly j hyperbolic planes where $j \leq \lceil \frac{r}{2} \rceil - 1$. Then Lemma 2.7 implies that

$$h \cong X_1X_2 + \cdots + X_{2j-1}X_{2j} + h'(X_{2j+1}, \ldots, X_n),$$

where h' is definite. Thus

$$|\operatorname{sgn}(h)| = |\operatorname{sgn}(h')| = n - 2j \geq n - 2\left(\left\lceil \frac{r}{2} \right\rceil - 1\right)$$

for any form $h \in \mathcal{P}_{\mathbb{R}}(f, g)$ having rank n. Let

$$T = \{(\lambda, \mu) \in \mathbb{S}^1 \mid \det(\lambda f + \mu g) = 0\},$$

and let $\mathcal{C}_i$, $1 \leq i \leq t$, denote the connected components in $\mathbb{S}^1 - T$. Since Sgn is an odd function, there are two adjacent connected components on $\mathbb{S}^1$ where the signature jumps from being positive to negative or vice versa. Therefore, there must be a jump of at least $2\left(n - 2\left(\left\lceil \frac{r}{2} \right\rceil - 1\right)\right)$ for the signature as (λ, μ) varies on $\mathbb{S}^1$. By Proposition 5.3, such a jump occurs only when (λ, μ) passes through a point in T, and the jump is bounded above by twice the dimension of the radical of the associated singular form. Let $\lambda_0 f + \mu_0 g$ be that singular form in $\mathcal{P}_{\mathbb{R}}(f, g)$ and let $r_0 = \operatorname{rank}(\lambda_0 f + \mu_0 g)$. Then the jump in the signature as we pass through (λ_0, μ_0) is bounded above by $2(n - r_0)$. Therefore,

$$\begin{aligned} 2\left(n - 2\left(\left\lceil \frac{r}{2} \right\rceil - 1\right)\right) &\leq 2(n - r_0) \\ -2\left(\left\lceil \frac{r}{2} \right\rceil - 1\right) &\leq -r_0 \\ r_0 \leq 2\left(\left\lceil \frac{r}{2} \right\rceil - 1\right) &< r, \end{aligned}$$

which is a contradiction because every form in $\mathcal{P}_{\mathbb{R}}(f, g)$ has rank at least r. Hence there exists $(\lambda, \mu) \in \mathbb{S}^1$ such that $\operatorname{rank}(\lambda f + \mu g) = n$ and splits off at least $\left\lceil \frac{r}{2} \right\rceil$ hyperbolic planes. $\square$

Example 6.5. Let $n \geq 2$ and let

$$\begin{aligned} f &= r_1X_1^2 + r_2X_2^2 + \cdots + r_nX_n^2 \\ g &= X_1^2 + X_2^2 + \cdots + X_n^2. \end{aligned}$$

Assume that $r_1, \ldots, r_n \in \mathbb{R}$ and $r_1 < r_2 < \cdots < r_n$. If n is even, choose $t \in \mathbb{R}$ such that $r_{\frac{n}{2}} < t < r_{\frac{n}{2}+1}$. If n is odd, choose $t \in \mathbb{R}$ such that either $r_{\frac{n-1}{2}} < t < r_{\frac{n+1}{2}}$ or $r_{\frac{n+1}{2}} < t < r_{\frac{n+1}{2}+1}$. Let $h = f - tg$. Then h has rank n and

$$\operatorname{sgn}(h) = \begin{cases} 0 & \text{if } n \text{ is even} \\ 1 \text{ or } -1 & \text{if } n \text{ is odd.} \end{cases}$$

Every form in the $\mathbb{R}$-pencil of f, g has rank $\geq n-1$ and h splits off exactly $\lceil \frac{n-1}{2} \rceil$ hyperbolic planes. This shows that the bound in Proposition 6.4 is optimal.

Example 6.6. Let K be a field with $\operatorname{char} K \neq 2$ and let $n = 2m+1$, $m \geq 1$. Let

$$\begin{aligned} f_m &= X_1X_2 + X_3X_4 + \cdots + X_{2m-1}X_{2m} \\ g_m &= \quad X_2X_3 + X_4X_5 + \quad \cdots \quad + X_{2m}X_{2m+1}. \end{aligned}$$

Every quadratic form in $\mathcal{P}_K(f_m, g_m)$ can be written

$$af_m + bg_m = X_2(aX_1 + bX_3) + \cdots + X_{2m}(aX_{2m-1} + bX_{2m+1}).$$

Thus every such form in $\mathcal{P}_K(f_m, g_m)$ has rank $2m$. Every quadratic form in $\mathcal{P}_K(f_m, g_m)$ splits off exactly $\frac{2m}{2} = m$ hyperbolic planes. Note that $D(\lambda, \mu) = \det(\lambda f_m + \mu g_m) = 0$ and that no form in $\mathcal{P}_K(f_m, g_m)$ has rank n.

Pairs of quadratic forms, such as those in Example 6.6 and Theorem 6.7, are essential for classifying pairs of quadratic forms over fields K with $\operatorname{char} K \neq 2$. We state Theorem 6.7 without proof, but the interested reader can find a proof and additional details in [5, Theorem 3.3] and [9, Theorems 3.1, 3.3].

Theorem 6.7. *Let K be an infinite field with $\operatorname{char} K \neq 2$. Let $f, g \in K[X_1, \ldots, X_n]$ be quadratic forms and assume that $\operatorname{rad}(f) \cap \operatorname{rad}(g) = 0$. Then there exist uniquely defined positive integers $m_1, \ldots, m_j$, $j \geq 0$, such that the pair f, g is equivalent to*

$$f \cong f_{m_1} \perp \cdots \perp f_{m_j} \perp q_2(X_{M+1}, \ldots, X_{M+N})$$

$$g \cong g_{m_1} \perp \cdots \perp g_{m_j} \perp q_2'(X_{M+1}, \ldots, X_{M+N}),$$

and such that the determinant polynomial $D(\lambda, \mu) = \det(\lambda q_2 + \mu q_2')$ over K is not identically zero, $M = \sum_{i=1}^{j}(2m_i + 1)$, and $M + N = n$.

Theorem 6.7 allows us to prove a stronger version of Proposition 6.4 where we can weaken the hypothesis and still conclude the same result.

Theorem 6.8. *Let $f, g \in \mathbb{R}[X_1, \ldots, X_n]$ be quadratic forms and assume that $\operatorname{rad}(f) \cap \operatorname{rad}(g) = 0$. Suppose that every form in $\mathcal{P}_{\mathbb{R}}(f, g)$ has rank at least r. Then there exists a form in $\mathcal{P}_{\mathbb{R}}(f, g)$ that splits off at least $\left\lceil \frac{r}{2} \right\rceil$ hyperbolic planes over $\mathbb{R}$, where*

$$\left\lceil \frac{r}{2} \right\rceil = \begin{cases} \dfrac{r}{2} & \textit{if } r \textit{ is even} \\ \dfrac{r+1}{2} & \textit{if } r \textit{ is odd.} \end{cases}$$

Proof. Let $q_1 = f_{m_1} \perp \cdots \perp f_{m_j}$ and $q_1' = g_{m_1} \perp \cdots \perp g_{m_j}$ in the notation of Theorem 6.7. Every form in $\mathcal{P}_K(q_1, q_1')$ has rank $2(m_1 + \cdots + m_j)$ and splits off $m_1 + \cdots + m_j$ hyperbolic planes. The determinant polynomial $\det(\lambda q_2 + \mu q_2') \neq 0$ and only finitely many forms in $\mathcal{P}_{\mathbb{R}}(q_2, q_2')$ have rank less than N. Every form in $\mathcal{P}_{\mathbb{R}}(q_2, q_2')$ has rank at least $R := r - 2(m_1 + \cdots + m_j)$. By Proposition 6.4, there exists a rank N form in $\mathcal{P}_{\mathbb{R}}(q_2, q_2')$ that splits off at least $\lceil \frac{R}{2} \rceil$ hyperbolic planes. Therefore, there exists a form in $\mathcal{P}_{\mathbb{R}}(f, g)$ that splits off at least $(m_1 + \cdots + m_j) + \lceil \frac{R}{2} \rceil = \lceil \frac{r}{2} \rceil$ hyperbolic planes. $\square$

Remark 6.9. If a quadratic form in n variables splits off k hyperbolic planes, then $2k \le n$. Thus Proposition 6.4 implies that $2\left\lceil \frac{r}{2} \right\rceil \le n$. If r is even, this gives $r \le n$. If r is odd, this gives $r+1 \le n$, so $r \le n-1$. That is, if r is odd, it is not possible that every form in $\mathcal{P}_{\mathbb{R}}(f,g)$ has rank $r=n$. Here is a way to see this directly. Suppose that $r=n$ is odd. Then $D(\lambda,\mu)=\det(\lambda f+\mu g)$ is a homogeneous form of odd degree n over $\mathbb{R}$. Since every form of odd degree over $\mathbb{R}$ has at least one nontrivial zero, there exist $\lambda_0,\mu_0 \in \mathbb{R}$, not both zero, such that $\lambda_0 f+\mu_0 g$ is singular, and so $\operatorname{rank}(\lambda_0 f+\mu_0 g)<n$. Therefore, in Proposition 6.4, if r is odd, then $r<n$.

7. Nonsingular zeros and simultaneous diagonalization

Definition 7.1 (Nonsingular Zero)**.** Let $f,g \in K[X_1,\dots,X_n]$ be quadratic forms.

(1) A vector $v \in K^n$ is a nonsingular zero of f if $f(v)=0$, and
$$\frac{\partial f}{\partial X}(v)=\left(\frac{\partial f}{\partial X_1}(v),\dots,\frac{\partial f}{\partial X_n}(v)\right)$$
is not the zero vector, and is a singular zero otherwise.

(2) A vector v is a nonsingular common zero of a pair of quadratic forms f,g if $f(v)=g(v)=0$, and the vectors
$$\frac{\partial f}{\partial X}(v),\ \frac{\partial g}{\partial X}(v)$$
are linearly independent over K, and is a singular common zero otherwise.

(3) We say that f,g is a nonsingular pair of quadratic forms if every nontrivial common zero of f,g defined over K^{alg} is a nonsingular zero.

Proposition 7.2. *Let $n \ge 2$ and let $f,\ g \in K[X_1,\dots,X_n]$ be a nonsingular pair of quadratic forms. Then every form in $\mathcal{P}_K(f,g)$ has rank at least $n-1$.*

Proof. Assume that there exists a form in $\mathcal{P}_K(f,g)$ having rank at most $n-2$. We can assume that $g=g(X_1,\dots,X_{n-2})$. Let $h(X_{n-1},X_n)=f(0,\dots,0,X_{n-1},X_n)$. There exist $a,b \in K^{alg}$, with $(a,b)\ne(0,0)$, such that $h(a,b)=0$. Then $(0,\dots,0,a,b)$ is a nontrivial singular zero of $f=g=0$. □

For quadratic forms $f,g,\in K[X_1,\dots,X_n]$, if $D(\lambda,\mu)$ is nonzero, then $D(\lambda,\mu)$ factors as a product of linear factors over K^{alg}. Then next result shows that the linear factors are distinct up to nonzero scalar factors in K^{alg} if and only if f,g is a nonsingular pair.

Proposition 7.3. *Let K be a field with $\operatorname{char} K \ne 2$ and let $f,\ g \in K[X_1,\dots,X_n]$ be quadratic forms. The following statements are equivalent.*

(1) *f,g is a nonsingular pair of quadratic forms.*

(2) *The homogeneous polynomial $D(\lambda,\mu)=\det(\lambda f+\mu g)$ has no repeated linear factors over K^{alg}.*

Proof. (2) ⇒ (1). Suppose that $a=(a_1,\dots,a_n)$ is a nontrivial singular zero of $f=g=0$ where each $a_i \in K^{alg}$. We can assume that this singular zero has coordinates $(1,0,\dots,0)$. Then
$$f=X_1L_1(X_2,\dots,X_n)+Q_1(X_2,\dots,X_n),$$
where L_1 is a linear form and Q_1 is a quadratic form. Similarly,
$$g=X_1L_2(X_2,\dots,X_n)+Q_2(X_2,\dots,X_n).$$
It follows from Definition 7.1 that L_1,L_2 are linearly dependent. Thus there exist $c,d \in K^{alg}$, not both zero, such that $cL_1+dL_2=0$. We can assume that $d\ne 0$ by interchanging f and g if necessary. Then we can assume that $L_2=0$ by replacing f,g

with $f, cf + dg$. Thus λ divides each entry of the first row and column of $\lambda M_f + \mu M_g$ and its $(1,1)$-entry is zero. It follows that $\lambda^2 \mid D(\lambda, \mu)$, a contradiction.

(1) $\Rightarrow$ (2). Suppose that $D(\lambda, \mu)$ has a linear factor of multiplicity at least 2 over K^{alg}. By choosing appropriate linear combinations of f, g in place of f, g, we can assume that $\lambda^2 \mid D(\lambda, \mu)$. Then the coefficient of μ^n in $D(\lambda, \mu)$ is zero, and thus $\det(M_g) = 0$, which implies that g has rank at most $n-1$. If g has rank at most $n-2$, then $f = g = 0$ has a singular zero over K^{alg} by Proposition 7.2. Thus g has rank $n-1$. We can assume that $g = g(X_2, \ldots, X_n)$. Then $g(1, 0, \ldots, 0) = 0$ and $(1, 0, \ldots, 0) \in \mathrm{rad}(g)$, where $(1, 0, \ldots, 0)$ denotes the point where $X_1 = 1$ and $X_i = 0$ for $i \geq 2$. Since the first row and column of M_g are both zero, the coefficient of $\lambda\mu^{n-1}$ in $D(\lambda, \mu)$ is given by the $(1,1)$-entry of M_f times the determinant of the lower right $(n-1) \times (n-1)$ submatrix of M_g. The coefficient of $\lambda\mu^{n-1}$ in $D(\lambda, \mu)$ is zero because λ^2 divides $D(\lambda, \mu)$. Since $\mathrm{rank}(g) = n-1$, it follows that the $(1,1)$-entry of M_f is zero and thus $f(1, 0, \ldots, 0) = 0$. Therefore, $(1, 0, \ldots, 0)$ is a singular zero of $f = g = 0$ because $(1, 0, \ldots, 0) \in \mathrm{rad}(g)$. $\square$

The converse of Proposition 7.2 does not hold in general, as shown in the next example, but we show in Proposition 7.8 that the converse does hold if f and g are simultaneously diagonalized as above.

Example 7.4. Let $f = 2X_1X_2$, $g = X_2^2$. Then every form in $\mathcal{P}_{\mathbb{R}}(f, g)$ has rank at least 1 but $(1, 0)$ is a singular zero of the pair f, g. We have $D(\lambda, \mu) = \det(\lambda f + \mu g) = -\lambda^2$, and so $D(\lambda, \mu)$ does not have distinct linear factors, as predicted by Proposition 7.3.

Lemma 7.5. *Let $f, g \in K[X_1, \ldots, X_n]$ be quadratic forms and suppose that* $\mathrm{rank}(g) <$ *n. If either f, g have no nontrivial common zero over K or $\lambda^2 \nmid D(\lambda, \mu)$, then the pair f, g is equivalent over K to*

$$\begin{aligned} f &= a_1X_1^2 + f_1(X_2, \ldots, X_n) \\ g &= \qquad\quad\; g_1(X_2, \ldots, X_n), \end{aligned}$$

where $a_1 \neq 0$.

Proof. There is an invertible linear change of variables over K that lets us assume that $g = g(X_2, \ldots, X_n)$. Let $P = (1, 0, \ldots, 0)$. We can write

$$f = a_1X_1^2 + X_1L(X_2 \ldots, X_n) + Q(X_2, \ldots, X_n)$$

where $L, Q \in K[X_2, \ldots, X_n]$ with L a linear form and Q a quadratic form.

Suppose that $a_1 = 0$. Then $f(P) = a_1 = 0$ and $g(P) = 0$. In addition, a straightforward computation shows that $\lambda^2 \mid D(\lambda, \mu)$. One of these statements contradicts our hypotheses, and thus $a_1 \neq 0$.

We have

$$f = a_1\left(X_1 + \frac{1}{2a_1}L\right)^2 + Q - \frac{1}{4a_1}L^2.$$

Let $X_1' = X_1 + \frac{1}{2a_1}L$ and $f_1 = Q - \frac{1}{4a_1}L^2$. Then $f = a_1(X_1')^2 + f_1$ and so the pair f, g is equivalent to

$$\begin{aligned} f &= a_1X_1^2 + f_1(X_2, \ldots, X_n) \\ g &= \qquad\quad\; g_1(X_2, \ldots, X_n), \end{aligned}$$

where $f_1, g_1 \in K[X_2, \ldots, X_n]$ are quadratic forms. $\square$

Theorem 7.6. *Let $f, g \in K[X_1, \ldots, X_n]$ be quadratic forms and assume that $D(\lambda, \mu)$ is a product of linear factors defined over K. If either f, g have no nontrivial common zero over K or $D(\lambda, \mu)$ has no repeated linear factors, then f, g can be simultaneously diagonalized over K.*

Proof. Since $D(\lambda, \mu) = \det(\lambda f + \mu g)$ has a linear factor defined over K, some form in $\mathcal{P}_K(f,g)$ has rank at most $n-1$. By choosing appropriate generators of $\mathcal{P}_K(f,g)$, we can assume that $\operatorname{rank}(g) < n$. Since either f, g have no nontrivial common zero over K or $\lambda^2 \nmid D(\lambda, \mu)$, Lemma 7.5 implies that the pair f, g is equivalent over K to

$$\begin{aligned} f &= a_1X_1^2 + f_1(X_2, \ldots, X_n) \\ g &= \qquad\quad\; g_1(X_2, \ldots, X_n), \end{aligned}$$

where $a_1 \neq 0$.

Suppose that m is maximal such that f, g are equivalent over K to

$$\begin{aligned} f &= a_1X_1^2 + \cdots + a_mX_m^2 + q(X_{m+1}, \ldots, X_n) \\ g &= b_1X_1^2 + \cdots + b_mX_m^2 + q'(X_{m+1}, \ldots, X_n), \end{aligned}$$

where $q, q' \in K[X_{m+1}, \ldots, X_n]$ are quadratic forms. Then $m \geq 1$. Suppose that $m < n$.

By unique factorization in $K[\lambda, \mu]$ it follows that $\det(\lambda q + \mu q')$ is a product of linear factors defined over K. In addition, either q, q' have no nontrivial common zero defined over K or $\det(\lambda q + \mu q')$ has no repeated linear factors. Repeating the argument at the beginning of this proof gives a contradiction to the maximality of m. Therefore, $m = n$, as desired. □

The next result shows that a nonsingular pair of quadratic forms can always be simultaneously diagonalized over an algebraically closed field.

Proposition 7.7. *Let $f, g \in K[X_1, \ldots, X_n]$ be a nonsingular pair of quadratic forms. Then f, g can be simultaneously diagonalized over K^{alg}.*

Proof. Proposition 7.3 implies that $D(\lambda, \mu) = \det(\lambda f + \mu g)$ is a product of distinct linear factors over K^{alg} up to nonzero scalar factors in K. The previous theorem implies that f, g can be simultaneously diagonalized over K^{alg}. □

Suppose that $f, g \in K[X_1, \ldots, X_n]$ are simultaneously diagonalized quadratic forms. Then

$$\begin{aligned} f &= a_1X_1^2 + a_2X_2^2 + \cdots + a_nX_n^2 \\ g &= b_1X_1^2 + b_2X_2^2 + \cdots + b_nX_n^2, \end{aligned}$$

where each $a_i, b_j \in K$. Assume that $(a_i, b_i) \neq (0,0)$ for each i. Then

$$D(\lambda, \mu) = \det(\lambda f + \mu g) = \prod_{i=1}^{n} (\lambda a_i + \mu b_i)$$

is a nonzero homogeneous form of degree n.

The next result gives additional characterizations of a nonsingular pair of quadratic forms f, g in the case that f, g are simultaneously diagonalized.

Proposition 7.8. *Suppose that $f, g \in K[X_1, \ldots, X_n]$ are simultaneously diagonalized quadratic forms, as above, with $(a_i, b_i) \neq (0,0)$ for each i. Then the following statements are equivalent.*

(1) *f, g is a nonsingular pair of quadratic forms.*
(2) *$D(\lambda, \mu)$ has no repeated linear factors over K^{alg}.*
(3) *Every form in $\mathcal{P}_K(f,g)$ has rank at least $n-1$.*
(4) *$a_ib_j - a_jb_i \neq 0$ for every $i \neq j$.*

Proof. We have already seen in Proposition 7.3 that (1) and (2) are equivalent, and in Proposition 7.2 that (1) $\Rightarrow$ (3). We will prove that the negations of (2), (3), (4) are equivalent. Statement (2) is false $\Leftrightarrow$ there exists $c \in K^\times$ and $i \neq j$ such that $c(\lambda a_i + \mu b_i) = \lambda a_j + \mu b_j \Leftrightarrow$ there exists $c \in K^\times$ and $i \neq j$ such that $a_j = ca_i$ and $b_j = cb_i$

$\Leftrightarrow$ there exists $c \in K^\times$ and $i \neq j$ such that $\begin{pmatrix} a_i & a_j \\ b_i & b_j \end{pmatrix} \begin{pmatrix} c \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\Leftrightarrow$ statement (4) is false. The last equivalence uses the assumption that each $(a_i, b_i) \neq (0,0)$.

Statement (3) is false $\Leftrightarrow$ there exist $r, s \in K$, not both zero, such that $rf + sg$ has rank at most $n-2$ $\Leftrightarrow$ there exist $r, s \in K$, not both zero, and $i \neq j$ such that $ra_i + sb_i = ra_j + sb_j = 0$ $\Leftrightarrow$ there exist $r, s \in K$, not both zero, and $i \neq j$ such that $\begin{pmatrix} a_i & b_i \\ a_j & b_j \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\Leftrightarrow$ statement (4) is false. $\square$

Let $A = \begin{pmatrix} a_i & a_j \\ b_i & b_j \end{pmatrix}$. Note that the proof of the equivalence of (2) and (4) used the matrix A, and the proof of the equivalence of (3) and (4) used the matrix A^t.

If $\operatorname{char} K = 2$, then the characterization of nonsingular pairs of quadratic forms is much more difficult. See [6].

8. Simultaneous diagonalization over the real numbers and the Spectral Theorem

In this section, we include additional results on nonsingular pairs of quadratic forms and simultaneous diagonalization that are specific to $\mathbb{R}$.

Proposition 8.1. *Let $f, g \in \mathbb{R}[X_1, \ldots, X_n]$ be a nonsingular pair of quadratic forms. Then there exists a nonsingular form in $\mathcal{P}_{\mathbb{R}}(f,g)$ that splits off $\lceil \frac{n-1}{2} \rceil$ hyperbolic planes over $\mathbb{R}$.*

Proof. Since every nontrivial zero of $f = g = 0$ is nonsingular, Proposition 7.2 implies that each form in $\mathcal{P}_{\mathbb{R}}(f,g)$ has rank at least $n-1$. Proposition 6.4 implies that there exists a nonsingular form in $\mathcal{P}_{\mathbb{R}}(f,g)$ that splits off at least $\lceil \frac{n-1}{2} \rceil$ hyperbolic planes over $\mathbb{R}$. Since $2(\lceil \frac{n-1}{2} \rceil + 1) > n$, it follows that this nonsingular form splits off exactly $\lceil \frac{n-1}{2} \rceil$ hyperbolic planes over $\mathbb{R}$. $\square$

Remark 8.2. If $n = 8$, Proposition 8.1 implies that there is a nonsingular form in $\mathcal{P}_{\mathbb{R}}(f,g)$ that splits off 4 hyperbolic planes. This strengthens a result in [3, Lemma 12.1], where it is proved that there exists a nonzero form in $\mathcal{P}_{\mathbb{R}}(f,g)$ that splits off at least 3 hyperbolic planes.

Lemma 8.3. *Let $f, g \in \mathbb{R}[X_1, \ldots, X_n]$ be quadratic forms. Suppose that $D(\lambda, \mu) = \det(\lambda f + \mu g)$ has no linear factors defined over $\mathbb{R}$. Then every form $h \in \mathcal{P}_{\mathbb{R}}(f,g)$ has rank n, n is even, and $\operatorname{sgn}(h) = 0$.*

Proof. The hypothesis implies that $n \geq 2$. Proposition 5.2 implies that $\operatorname{sgn}(h)$ is the same for all $h \in \mathcal{P}_{\mathbb{R}}(f,g)$ because T (as defined prior to Lemma 5.1) is the empty set. Since $\operatorname{sgn}(-h) = -\operatorname{sgn}(h)$, it follows that each $h \in \mathcal{P}_{\mathbb{R}}(f,g)$ has $\operatorname{sgn}(h) = 0$. The hypothesis implies that every form in $\mathcal{P}_{\mathbb{R}}(f,g)$ has rank n. Thus n is even. $\square$

Theorem 8.4. *If $f, g \in \mathbb{R}[X_1, \ldots, X_n]$ are quadratic forms and f is definite, then $D(\lambda, \mu)$ is a product of linear factors defined over $\mathbb{R}$ and f, g can be simultaneously diagonalized over $\mathbb{R}$.*

Proof. The proof is by induction on $n \geq 1$, with the case $n = 1$ being obvious. Assume that $n \geq 2$ and that the result is true for fewer than n variables.

First suppose that $D(\lambda, \mu)$ has no linear factor defined over $\mathbb{R}$. Then Lemma 8.3 implies that each $h \in \mathcal{P}_{\mathbb{R}}(f,g)$ has rank n and signature 0. This is a contradiction because f is definite and thus has signature $\pm n$. Then some form in $\mathcal{P}_{\mathbb{R}}(f,g)$ has rank at most $n-1$. By choosing appropriate generators of $\mathcal{P}_{\mathbb{R}}(f,g)$, we can assume that $\operatorname{rank}(g) < n$. Lemma 7.5 implies that f,g is equivalent to the pair

$$\begin{aligned} f &= a_1X_1^2 + f_1(X_2,\ldots,X_n) \\ g &= \qquad\quad g_1(X_2,\ldots,X_n), \end{aligned}$$

where $a_1 \neq 0$. The pair f_1, g_1 satisfies the hypotheses because f_1 is definite. By induction it follows that $\det(\lambda f_1 + \mu g_1)$ is a product of linear factors defined over $\mathbb{R}$ and f_1, g_1 can be simultaneously diagonalized over $\mathbb{R}$. Then the same holds for f, g. □

Corollary 8.5. *If $n \geq 3$, and $f, g \in \mathbb{R}[X_1,\ldots,X_n]$ are quadratic forms having no nontrivial common zero over $\mathbb{R}$, then f, g can be simultaneously diagonalized over $\mathbb{R}$.*

Proof. Since $n \geq 3$ and f, g have no nontrivial common zero defined over $\mathbb{R}$, Proposition 4.4 implies that there is a form $h \in \mathcal{P}_{\mathbb{R}}(f,g)$ that is definite. The result now follows from the previous theorem. □

Example 8.6. This example shows that the previous corollary is false when $n = 2$. Let $f = 2X_1X_2$ and $g = X_1^2 - X_2^2$. The pair f, g is an anisotropic pair because there is no nontrivial common zero of f, g with either $X_1 = 0$ or $X_2 = 0$. The pair cannot be simultaneously diagonalized because $D(\lambda,\mu) = \det(\lambda f + \mu g) = -(\lambda^2+\mu^2)$ has no linear factors defined over $\mathbb{R}$.

These results lead to a proof of the Spectral Theorem. Much of the proof below is standard linear algebra but is included for the sake of completeness.

Theorem 8.7 (Spectral Theorem)**.** *Let A be an $n \times n$ symmetric matrix with entries in $\mathbb{R}$. Then the following statements hold.*

(1) *Every eigenvalue of A is real.*
(2) *There exists an $n \times n$ orthogonal matrix P with entries in $\mathbb{R}$ such that P^tAP is a diagonal matrix.*
(3) *A has n pairwise orthogonal (and thus linearly independent) eigenvectors in $\mathbb{R}^n$.*

Proof. Let $g(X_1,\ldots,X_n) = X^tAX$ where X is the column vector $(X_1,\ldots,X_n)^t$ and let $f(X_1,\ldots,X_n) = X^tI_nX$ where I_n is the $n \times n$ identity matrix. Thus $f = X_1^2+\cdots+X_n^2$. Since f is positive definite, the pair f, g can be simultaneously diagonalized over $\mathbb{R}$ by Theorem 8.4. Thus there exists an $n \times n$ invertible matrix M with entries in $\mathbb{R}$ such that $M^tAM = D_1$ and $M^tI_nM = D_2$ are both diagonal matrices. Since D_2 must be positive definite, each entry on the main diagonal of D_2 is positive. Then $D_2 = D_3^2$ for some invertible diagonal matrix D_3 with entries in $\mathbb{R}$. Let $P = MD_3^{-1}$. Then $P^tAP = (D_3^{-1})^tM^tAMD_3^{-1} = D_3^{-1}D_1D_3^{-1} = D_2^{-1}D_1$ is a diagonal matrix, and similarly, $P^tI_nP = D_3^{-1}D_2D_3^{-1} = I_n$. Thus $P^tP = I_n$ and so P is an orthogonal matrix. It follows that the columns of P are pairwise orthogonal and thus linearly independent.

We have $P^{-1}AP = P^tAP = D_2^{-1}D_1$, which gives $AP = P(D_2^{-1}D_1)$. Let $D_2^{-1}D_1 = \operatorname{diag}(d_1,\ldots,d_n)$. Let $v_1,\ldots,v_n$ be the columns of P. The equation $AP = P(D_2^{-1}D_1)$ implies that $Av_i = d_iv_i$ for $1 \leq i \leq n$. Thus the columns of P are pairwise orthogonal eigenvectors of A with eigenvalues $d_1,\ldots,d_n$. These eigenvalues of A are real because the diagonal matrices D_1 and D_2 have real entries. □

References

[1] Calabi, E. (1964). Linear systems of real quadratic forms. *Proc. Amer. Math. Soc.* **15**: 844–846.

[2] Finsler, P. (1936). Über das Vorkommen definiter und semidefiniter Formen in Scharen quadratischer Formen. *Comment. Math. Helv.* **9**(1): 188–192.

[3] Heath-Brown, D. R. (2018). Zeros of pairs of quadratic forms. *J. Reine Angew. Math.* **739**: 41–80.

[4] Hoffman, K., Kunze, R. (1972). *Linear Algebra,* second ed. Prentice-Hall, Inc., Englewood Cliffs, N.J.

[5] Leep, D. B., Schueller, L. M. (1999). Classification of pairs of symmetric and alternating bilinear forms. *Exposition. Math.* **17**(5): 385–414.
[6] Leep, D. B., Schueller, L. M. (2002). A characterization of nonsingular pairs of quadratic forms. *J. Algebra Appl.* **1**(4): 391–412.
[7] Swinnerton-Dyer, H. P. F. (1964). Rational zeros of two quadratic forms. *Acta Arith.* **9**: 261–270.
[8] Uhlig, F. (1979). A recurring theorem about pairs of quadratic forms and extensions: a survey. *Linear Algebra Appl.* **25**: 219–237.
[9] Waterhouse, W. C. (1976). Pairs of quadratic forms. *Invent. Math.* **37**(2): 157–164.

David B. Leep is a Professor of Mathematics at the University of Kentucky. He received his PhD at the University of Michigan in 1980. His research interests include the algebraic theory of quadratic forms, forms of higher degrees, and equations over finite fields and p-adic fields.

Nandita Sahajpal is an Assistant Professor of Mathematics at Nevada State College. She received her PhD at the University of Kentucky in 2020. Her research interests include the algebraic theory of quadratic forms and elliptic curves over number fields and local fields. In addition to her research, she is also passionate about inclusive and equitable teaching methods.

(David B. Leep) Department of Mathematics, University of Kentucky, Lexington, KY 40506, USA.

(Nandita Sahajpal) Department of Data, Media and Design, Nevada State College, Henderson, NV 89002, USA.

E-mail address, D. Leep: `leep@uky.edu`

E-mail address, N. Sahajpal: `nandita.sahajpal@nsc.edu`

Irish Math. Soc. Bulletin
Number 91, Summer 2023, 73–76
ISSN 0791-5578

There are infinitely many primes: two ring-theoretic variations on Euclid

ALAN ROCHE

Abstract. Using elementary ring theory, we present two proofs in the mode of Euclid that there are infinitely many primes.

1. Introduction

Euclid's proof of the infinitude of primes is a paragon of incisive mathematical reasoning. It's the first entry—deservedly—in Aigner and Ziegler's compilation, their terrestrial approximation to the celestial BOOK [1, p. 3]. The result (infinitude of primes) has been re-proved over and over. Aigner and Ziegler, for example, discuss six proofs in their first chapter and infinitely many more (in a sense) in an appendix.

We use elementary ring theory to show, yet again, that there are infinitely many primes. The argument's strategy is simple: if $p_1, \ldots, p_n$ is the complete list of primes, then the ring of rational numbers $\mathbb{Q}$ is obtained from the ring of integers $\mathbb{Z}$ by adjoining the single element $1/p_1 \cdots p_n$. The task then is to show that this is an untenable structure for $\mathbb{Q}$ which we do in two overlapping ways. In each case, the proof makes use of the key Euclidean manoeuvre: given the list of primes $p_1, \ldots, p_n$, consider $p_1 \cdots p_n + 1$.

We conclude with some comments on Euclid's classic argument.

2. First Proof

Given nonzero integers $a_1, \ldots, a_n$, we write $\mathbb{Z}[1/a_1, \ldots, 1/a_n]$ for the smallest subring of $\mathbb{Q}$ containing $\mathbb{Z}$ and each $1/a_i$. Equivalently, it's the smallest subring of $\mathbb{Q}$ with identity in which $a_1, \ldots, a_n$ are invertible. As the notation suggests, it consists of all $f(1/a_1, \ldots, 1/a_n)$ for $f(\mathrm{X}_1, \ldots, \mathrm{X}_n) \in \mathbb{Z}[\mathrm{X}_1, \ldots, \mathrm{X}_n]$.

Note that

$$\mathbb{Z}[1/a_1, \ldots, 1/a_n] = \mathbb{Z}[1/a_1 \cdots a_n]. \tag{1}$$

Indeed, $a_1 \cdots a_n$ is invertible (in a subring of $\mathbb{Q}$ with identity) if and only if each a_i is invertible (in that subring), and so the two rings coincide.

Suppose now that there are only finitely many primes, say $p_1, \ldots, p_n$. Since each positive integer m is a product of primes, our supposition implies that $1/m$ is in $\mathbb{Z}[1/p_1, \ldots, 1/p_n]$, and therefore

$$\mathbb{Q} = \mathbb{Z}[1/p_1, \ldots, 1/p_n].$$

Equivalently, by (1), $\mathbb{Q} = \mathbb{Z}[1/p_1 \cdots p_n]$. To simplify the notation, we set $a = p_1 \cdots p_n$, so that $\mathbb{Q} = \mathbb{Z}[1/a]$.

2020 *Mathematics Subject Classification.* 11A41, 13A05.
Key words and phrases. primes, polynomials, maximal ideal.
Received on 4-4–2023; revised 15-5-2023.
DOI:10.33232/BIMS.0091.73.76.

In particular, $1/(a+1) \in \mathbb{Z}[1/a]$. This means there exist integers $c_0, c_1, \dots, c_m$ such that

$$\frac{1}{a+1} = c_0 + c_1\frac{1}{a} + \cdots + c_m\frac{1}{a^m}.$$

Multiplying through by a^m, we have

$$\frac{a^m}{a+1} = c_0a^m + c_1a^{m-1} + \cdots + c_m \in \mathbb{Z}.$$

That is, $a+1$ divides a^m. Now $1 = [(a+1) - a]^m$. Expanding the right side, we see that

$$1 = A(a+1) + (-1)^m a^m,$$

for some integer A. Since $a+1$ divides a^m, it follows that $a+1$ divides 1 which is absurd. We've proved that there are infinitely many primes. □

3. Second Proof

Assume once more that there are only finitely many primes $p_1, \dots, p_n$. As above, it follows that $\mathbb{Q} = \mathbb{Z}[1/a]$ for $a = p_1 \cdots p_n$. In other words, the homomorphism of rings

$$f(\mathrm{X}) \mapsto f(1/a) : \mathbb{Z}[\mathrm{X}] \to \mathbb{Q} \tag{2}$$

is surjective. We write I_a for its kernel, so that (2) induces an isomorphism of rings

$$\overline{f(\mathrm{X})} \longmapsto f(1/a) : \mathbb{Z}[\mathrm{X}]/I_a \xrightarrow{\simeq} \mathbb{Q}. \tag{3}$$

In particular, $\mathbb{Z}[\mathrm{X}]/I_a$ is a field, or equivalently I_a is a maximal ideal in $\mathbb{Z}[\mathrm{X}]$.

To finish the argument, we could appeal to a property of maximal ideals in $\mathbb{Z}[\mathrm{X}]$—that each such ideal contains some nonzero constant polynomial. Indeed, as I_a contains no nonzero constants, we see that I_a cannot be maximal, a contradiction.

This approach, however, is unsatisfying: the property that maximal ideals in $\mathbb{Z}[\mathrm{X}]$ contain nonzero constants lies deeper than the existence of infinitely many primes. Instead, we'll use only our bare hands to prove the following: if $\mathbb{Z}[\mathrm{X}]/I_a$ is a field then $a+1$ must divide 1 (as in the first proof). Our path to this absurdity rests on identifying the structure of the ideal I_a.

Lemma. *We have $I_a = (a\mathrm{X} - 1)$, the principal ideal generated by $a\mathrm{X} - 1$.*

The ideal of elements of $\mathbb{Q}[\mathrm{X}]$ that vanish at $1/a$ is generated by $\mathrm{X} - 1/a$ and so also by $a\mathrm{X} - 1$. The proof that I_a is generated by $a\mathrm{X} - 1$ is then a short exercise using Gauss's Lemma—a product of primitive polynomials is primitive. (Recall an element of $\mathbb{Z}[\mathrm{X}]$ is primitive if the greatest common divisor of its coefficients is 1.) We prefer, however, a still more elementary, albeit ad hoc approach. We want to avoid all tools beyond the most basic properties of polynomials, even one as fundamental as Gauss's Lemma.

Proof. Let $f(\mathrm{X}) = c_0 + c_1\mathrm{X} + \cdots + c_m\mathrm{X}^m \in \mathbb{Z}[\mathrm{X}]$ with $c_m \neq 0$, so $f(\mathrm{X})$ has degree m. We have

$$c_0 + c_1\frac{1}{a} + \cdots + c_m\frac{1}{a^m} = \frac{c_0a^m + c_1a^{m-1} + \cdots + c_m}{a^m}.$$

Thus $f(1/a) = 0$ if and only if $\widetilde{f}(a) = 0$ where

$$\begin{aligned} \widetilde{f}(\mathrm{X}) &= \mathrm{X}^m f(1/\mathrm{X}) \\ &= c_0\mathrm{X}^m + c_1\mathrm{X}^{m-1} + \cdots + c_m. \end{aligned} \tag{4}$$

We call $\widetilde{f}(\mathrm{X})$ the *reverse* of $f(\mathrm{X})$ and going from $f(\mathrm{X})$ to $\widetilde{f}(\mathrm{X})$ *reversing*. Visibly, the reverse of the reverse of $f(\mathrm{X})$ is $f(\mathrm{X})$: reversing is an involution on the set of nonzero elements of $\mathbb{Z}[X]$. Moreover, it follows readily from (4) that reversing is multiplicative: that is, $\widetilde{f_1 f_2}(X) = \widetilde{f_1}(X)\widetilde{f_2}(X)$ for nonzero $f_i(X) \in \mathbb{Z}[X]$ $(i = 1, 2)$.

Remember the division algorithm for polynomials applies to monic elements of $\mathbb{Z}[X]$. Hence, for $g(X) \in \mathbb{Z}[X]$, we have $g(a) = 0$ if and only if $X - a$ divides $g(X)$ in $\mathbb{Z}[X]$. In particular,

$$\widetilde{f}(a) = 0 \iff \widetilde{f}(X) = (X - a)\, h(X),$$

for some $h(X)$. Reversing the polynomial equation and noting that the reverse of $X - a$ is $-(aX - 1)$, we see that

$$\widetilde{f}(a) = 0 \iff f(X) = (aX - 1)\left(-\widetilde{h}(X)\right).$$

Thus $f(1/a) = 0$ if and only if $aX - 1$ divides $f(X)$. We've proved the lemma. □

Now, since $a + 1 \notin I_a$, the coset $(a + 1) + I_a$ is invertible in the field $\mathbb{Z}[X]/I_a$. Hence there is an $h(X) \in \mathbb{Z}[X]$ such that $(a + 1)\, h(X) + I_a = 1 + I_a$. Using the lemma, it follows that

$$(a + 1)\, h(X) = 1 + (aX - 1)\, k(X), \tag{5}$$

for some $k(X)$. Substituting $X = a$, we obtain

$$(a + 1)\, h(a) = 1 + \left(a^2 - 1\right) k(a),$$

and so

$$(a + 1)\, [h(a) - (a - 1)\, k(a)] = 1.$$

Again, we've reached the absurdity that $a + 1$ divides 1. We've proved once more that there are infinitely many primes. □

4. Comments on Euclid's Proof

First, let's recast Euclid's argument in the language of ring theory.

Proof. Let a be a nonunit in $\mathbb{Z}$, that is, $a \neq \pm 1$. Then a has a prime divisor p, or equivalently $a \in (p)$ for some prime p. We assume that there are only finitely many primes, say $p_1, \ldots, p_n$. It follows that each nonunit in $\mathbb{Z}$ is contained in some (p_i), and therefore

$$\mathbb{Z} \setminus \{\pm 1\} = \bigcup_{i=1}^{n} (p_i). \tag{6}$$

Now $p_1 \cdots p_n + 1$ is not divisible by p_i, for $i = 1, \ldots, n$. That is,

$$p_1 \cdots p_n + 1 \notin \bigcup_{i=1}^{n} (p_i).$$

Using (6), we have $p_1 \cdots p_n + 1 = \pm 1$. Nonsense! We conclude that there are infinitely many primes. □

Remark 1. We've presented our variants of Euclid's argument in terms of contradiction. In this form, they give the *existence* of infinitely many primes. As many have noted, however, Euclid's reasoning is *constructive* (see, for example, [2, p. 31]): given a finite list of primes $p_1, \ldots, p_n$, Euclid gives a way (an inefficient way) of adjoining a new prime to the list—namely, any prime factor of $p_1 \cdots p_n + 1$.

Having dressed Euclid's proof in ring-theoretic garb, we can use some set theory to obtain a small generalization. First, some notation. For R a ring with identity, we write $R^\times$ for the group of units of R.

Proposition. *Let R be a PID that is not a field and suppose the cardinality of $R^\times$ is strictly smaller than that of R. Then R contains infinitely many irreducible elements (up to multiplication by units).*

The result applies, in particular, if $R^\times$ is finite.

Proof. We assume that R has only finitely many irreducible elements $\varpi_1, \ldots, \varpi_n$ (up to multiplication by units) and will show that $R^\times$ and R have the same cardinality.

By hypothesis, each nonunit in R is divisible by some ϖ_i. Therefore

$$R \setminus R^\times = \bigcup_{i=1}^{n} (\varpi_i).$$

Now, for $r \in R$, the element $1 + r\varpi_1 \cdots \varpi_n$ is not contained in any (ϖ_i), and so belongs to $R^\times$. Hence we have a map

$$r \mapsto 1 + r\varpi_1 \cdots \varpi_n : R \to R^\times$$

which is injective (as R is a domain). By the Schröder-Bernstein Theorem, $R^\times$ and R have the same cardinality. $\square$

Remark 2. The proposition is not sharp—it was too easy to prove to expect it to be sharp! That is, there are PIDs R with infinitely many irreducible elements (up to multiplication by units) for which $R^\times$ has the same cardinality as R. Example: $R = \mathbb{Z}[\sqrt{2}]$. Indeed, as $(\sqrt{2}+1)(\sqrt{2}-1) = 1$, we see that $R^\times$ contains the infinite cyclic group generated by $\sqrt{2}+1$, and so is countably infinite.

Remark 3. Which PIDs R contain infinitely many irreducible elements (up to multiplication by units)? The note [3] gives a characterization in terms of the polynomial ring $R[\mathrm{X}]$: a PID R has the given property if and only if each maximal chain of prime ideals in $R[\mathrm{X}]$ has length two, that is, has the form $\{0\} \subsetneq \mathfrak{p}_1 \subsetneq \mathfrak{p}_2$, for prime ideals $\mathfrak{p}_i$ in $R[\mathrm{X}]$ $(i = 1, 2)$.

References

[1] M. Aigner and G. M. Ziegler, *Proofs from THE BOOK*, 6th ed. Springer, 2018.

[2] J. Stillwell, *The Story of Proof—Logic and the History of Mathematics.* Princeton University Press, 2022.

[3] F. Zanello, *When are there infinitely many irreducible elements in a principal ideal domain?* Amer. Math. Monthly. 111(2) (2004), 150–152.

Alan Roche holds degrees in mathematics from University College Dublin and the University of Chicago. He has taught and thought at the University of Oklahoma since 2001.

Department of Mathematics, University of Oklahoma, Norman, OK 73019, USA.
E-mail address: aroche@ou.edu

Irish Math. Soc. Bulletin
Number 91, Summer 2023, 77–78
ISSN 0791-5578

Des MacHale: Comic Sections Plus: The Book of Mathematical Jokes, Humour, Wit and Wisdom, Logic Press, 2022. ISBN:978-1-4717-6147-8, EUR19.99 , 266+viii pp.

REVIEWED BY RÓISÍN & AOIFE HILL

Prior to reaching the preface, it is written that "nothing in this book should be taken too seriously". It's a book filled with 'anecjokes' (anecdotes and jokes) about mathematics or mathematicians and includes limericks, riddles, endless wit, and even an exam at the end. This is an extension to the Des MacHale's earlier book, Comic Sections [1], now including an additional thirty years worth of witty mathematical material.

The preface sets the scene particularly well - MacHale's passion for this project has at its core his desire not only to humanise mathematics and mathematicians but to use this humorous content to aid in the study and understanding of mathematics. As in every good joke book, logic is spun on its head time and again; when paradoxically coupled with the ever-logical mathematics, wit shines through. To emphasise this, MacHale cleverly quotes Einstein: "As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality".

The book is filled with many 'inside' jokes that non-mathematicians may (i) be puzzled by and (ii) regret letting this be known to a mathematician after this leads to a lecture about what Euler's number and imaginary numbers are (see: "Old MacDonald was not a great mathematician. He couldn't work out the value of $e^{i^{e^{i^{0}}}}$."). However, this is what made the book feel special; it felt like a place of community. It's the first joke book I've read where my academic education has been a blessing rather than a curse. As a maths book, it's as rare as a four-leaf clover - it may not quite satisfy uniqueness, but it certainly fits into the small subset of maths books that can be easily picked up by anyone in, or even on the cusp of, the field (although, MacHale jibes that some branches of mathematics are less susceptible to humour).

This book is an excellent aid for the classroom. MacHale explains the importance of humour in this setting and how best to use it, recommending it to be used as "dessert" or "seasoning". He reasons that a paradoxical proof that $2 = 1$ and the resulting discussion and analysis will provide elementary students with a greater understanding and interest than lengthy axioms or theorems. Such humour can not only be insightful but also be an excellent memory aid for students.

The book is split into twelve chapters, ranging from a dictionary of mathematical terms ("clearly" et al.) to riddles to humorous journal reviews that will be sure to hit at least one nerve. A "Questioner's Handbook" is included to help eliminate the uncomfortable silence accompanied by "any questions?" after a seminar that failed to resonate with the audience. "Mathematical Wit and Wisdom", a chapter devoted to what has previously been said about mathematics both by those within and outside the field, may be my favourite, containing a remarkable collection of insight into the many

Received on 03-02-2023; revised 09-02-2023.
DOI:10.33232/BIMS.0091.77.78.
Go to `https://www.logicpress.ie/LS/2022-1` for purchase options

attitudes towards mathematics. Chapter 11, "Those magnificent men on their Turing machines", sums up the tongue in cheek nature of the book, reminding us of the all too familiar funding struggles imposed by those without an understanding of what is truly essential in mathematics. The book ends with "The Final Examination", where you're reminded not to "attempt to write on both sides of the paper at the same time" and an extra credit question tells you to "persuade the first passer-by you meet to accompany you through life, using irony where necessary".

Ultimately, this is an excellent concept well-executed, reinforcing that humour can be found in every level of Mathematics. It is a brilliant compendium whether you're looking for a page-turner to dip into during a coffee break or looking to bring wit into the classroom in a way that remains thought-provoking.

References

[1] D. MacHale: *Comic Sections: The Book of Mathematical Jokes, Humour, Wit and Wisdom*, Boole Press, Dublin, 1993.

Róisín & Aoife Hill are a mother-daughter duo. Aoife pursued a BSc in Applied Mathematics from University of Galway in 2012, and intrigued by this, Róisín followed suit the following year. They both completed computational PhDs from University of Galway in 2022 in the fields of Applied Mathematics and Biomedical Engineering, respectively. Róisín is currently an Irish Research Council Postdoctoral Fellow at University of Limerick while Aoife is working on a Disruptive Technologies Innovation Fund project at University of Galway.

Department of Mathematics and Statistics, University of Limerick; Biomechanics Research Centre, University of Galway

E-mail address: roisin.hill@ul.ie;aoife.hill@universityofgalway.ie

Irish Math. Soc. Bulletin
Number 91, Summer 2023, 79–80
ISSN 0791-5578

J. Stillwell: The Story of Proof, Princeton University Press, 2022. ISBN:978-0691234366, USD 45.00, 456 pp.

REVIEWED BY TOMMY MURPHY

The physicist Freeman Dyson opined that all mathematicians are by temperament either eagles or frogs. The eagle soars over the mathematical landscape, observing connections between apparently disparate fields and generalising theorems whilst staying light on details, whereas the frog stays in a small locality of the world of mathematics, delving into the intricacies and focused on a deep understanding of specific questions. Dyson's thesis is that the academy needs both viewpoints, and collaborations are enriched when these two tribes work together. Poincaré is widely believed to be the last person who could soar high enough to see the entire mathematical landscape; due to the specialised nature and enormous volume of mathematical research nowadays even our eagles tend to roam within one or two well-trodden research areas.

In "The Story of Proof", John Stillwell has written a well-crafted, thought-provoking meditation on the concept of proof in mathematics, which is used as an organising principle to explain, in broad brushstrokes, how disparate fields of mathematics emerged and lay bare the origins of some of major problems in mathematics. If you are interested in pure mathematics, you should buy and/or read this book. In the spirit of Dyson's dichotomy, it is enlightening and satisfying to learn more about how different areas of research are connected. To give one such example, I had never known how knot theory arose from the study of singularities of algebraic curves. Before reading this book I also did not fully appreciate that Dedekind and Kronecker developed field extensions in an attempt to abstract of the concept of dimension for vector spaces, and used this to answer the ancient Greek problem of duplicating a cube. The book is full of such insights. I learnt field extensions and applications as an undergraduate, and have never really thought about them since or why they were important as it is not relevant to my research. The point is that in today's competitive race to the coal-face of research, we tend to focus on specific topics and questions and much context is lost without an appreciation of how and why these questions arose. As undergraduates we learn about the insolvability of the quintic: we perhaps do not appreciate the reason that surds are of interest in this context is because of the connection with compass-and-straightedge constructions arising from ancient Greek mathematics.

A fascinating theme of the book is the connection between logic and computation. As Stillwell points out, in Greek times logic was comparatively strong but the theory of computation was weak. Thus was born the great glories of Greek mathematics, as opposed to other civilisations of the time who could compute but did not truly understand. Computational techniques advanced tremendously with the advent of calculus, and logic and proof had to catch up. This of course leads us to the natural development of analysis in the eighteenth century. As such, the book begins by focusing on a (somewhat perfunctory) survey of Euclid, before jumping to Hilbert's axiomisation of geometry. At times the pace is too fast and material is not fully explained: we find the

Received on 19-05-2023; revised 24-05-2023.
DOI:10.33232/BIMS.0091.79.80.

question of proving consistency of an axiomatic system by constructing a model being mentioned, without motivation, on page 48. Nevertheless, there is much to glean from the text even if some statements wash over the reader initially. The gradual realisation that logic and computation were closely related mathematical concepts is explained well in this book, culminating in a very satisfactory survey of the work of Gödel and Turing. By this stage questions of consistency make more sense to the reader.

Stillwell has many interesting examples explaining how various proofs of a theorem evolved which will enrich the reader's appreciation and understanding. A striking example of this is in the discussion of various failed proofs of the Fundamental Theorem of Algebra. Trying to battle through the maze of details here is what led Bolzano to realise the importance of trying to establish the Intermediate Value Theorem via the least upper bound property: a crucial component in the development of real analysis and closely connected to Dedekind cuts. Another instance, though certainly not novel to this book, is the emphasis placed on explaining the origins of algebraic geometry and placing them in the context of Greek work on conics.

I know of no comparable book on the market. It is not suitable as a textbook for an undergraduate or postgraduate course on the history of mathematics, owing to its dizzying pace and the deliberate choice to omit details and proofs in many places and to emphasise connections. It is not a comprehensive overview of every aspect of the history of mathematics, rather a discourse on what we are really doing in mathematics and how our understanding of proof, computation, and logic have evolved and intertwined over two millennia of human thought, and how remarkably interconnected mathematics is. It is a substantive book that deserves to be read and reflected upon.

Tommy Murphy is Associate Professor at Cal State Fullerton, having completed his Ph.D. under Jürgen Berndt at UCC and postdoctoral fellowships at the Université Libre de Bruxelles (Belgium) and McMaster University (Canada). His research interests span Riemannian and Kähler geometry, focused specifically on Einstein manifolds and symmetric spaces. He also maintains an active interest in the history of mathematics and collaborating with undergraduate students.
Website:`https://www.fullerton.edu/math/faculty/tmurphy/`

Dept. of Mathematics, California State University Fullerton, 800 N. State College Blvd., Fullerton 92831, CA, USA.
E-mail address: `tmurphy@fullerton.edu`

Irish Math. Soc. Bulletin
Number 91, Summer 2023, 81–82
ISSN 0791-5578

Michael Z. Spivey: The Art of Proving Binomial Identities, CRC Press, 2019. ISBN:978-0-8153-7942-3, USD 88.00, 368+xiv pp.

REVIEWED BY HENRY RICARDO

Probably all readers of this journal have encountered binomial coefficients many times in their teaching and/or research. Even Gilbert and Sullivan's Modern Major General declared "About binomial theorem I'm teeming with a lot o' news." As someone who is not a combinatorialist (or a military officer) but who haunts the problem sections of various mathematical journals, I have encountered many challenging problems involving binomial coefficients. Often my instinct is to check Gould's impressive compendium [1] of over 500 binomial coefficient identities—but, unfortunately, there are no proofs.

In ten chapters, Michael Spivey succeeds in bringing together in a systematic way the many methods used in dealing with binomial coefficients. The techniques covered in this book consist of algebra (including finite difference methods and complex numbers), calculus, linear algebra, and combinatorics/probability. Anyone absorbing the techniques in Spivey's book should be prepared to understand Gould's collection and tackle deeper works such as Riordan's classic monograph [2]. This is not an encyclopedia, although the convenient Index of Identities and Theorems makes it easy to find alternative proofs or other uses of particular identities. Because of the way Spivey's book ties together several key courses in the undergraduate mathematics curriculum, a university maths department could base a senior seminar or capstone course/project on this book.

The author starts by proving the equivalence of four definitions of the binomial coefficient for integers n and k with $n \geq k \geq 0$:

(1) The number of subsets of size k formed from a set of n elements;
(2) $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$ when $n-1 \geq k \geq 1$ with boundary conditions $\binom{n}{0} = \binom{n}{n} = 1$;
(3) the coefficient of x^k in the expansion of $(x+1)^n$ in powers of x;
(4) $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Later, Spivey generalises to $n \in \mathbb{R}$ and $k \in \mathbb{Z}$.

Two of my favourite topics are given good introductory treatments: the central binomial coefficient (CBC) and reciprocal binomial coefficients (RBCs). One worked-out example yields this gem:

$$\sum_{n=0}^{\infty} \frac{1}{\binom{2n}{n}} = \frac{4}{3} + \frac{2\pi\sqrt{3}}{27}.$$

The CBC is developed further in other sections of the book, including discussions of Catalan numbers and lattice paths. A number of other sums (finite and infinite) of RBCs are discussed and are listed in the very useful Index to Identities and Theorems, but the term reciprocal binomial coefficient is missing from the main index.

Received on 24-04-2023; revised 02-05-2023.
DOI:10.33232/BIMS.0091.81.82.

Some of the over 300 numbered identities appear as examples, while others are used as exercises. Throughout the book there are historical inserts, presenting snippets of information about various mathematicians: Eric Temple Bell, the Bernoulli family, . . . , Newton, Pascal, Stirling, . . . , Vandermonde, Zeilberger. There are also end-of-chapter notes that provide context for the chapter's results and/or recommendations for further reading. Appendices include an 82-item Bibliography and Hints and Solutions to Exercises—the latter particularly useful for self-study.

In the final chapter ("Mechanical Summation"), art gives way to science. Here the author explains the powerful Gosper-Zeilberger algorithm, both Gosper's original form and Zeilberger's extension. Hypergeometric series are introduced early. The treatment contains "very heavy algebra", in the author's words, but is a tour de force of exposition.

In summary, this is a delightful and useful book: A readable introduction to binomial coefficients and many of their applications for the advanced undergraduate or graduate student, an aid to those mathematically mature individuals who are not combinatorialists, an inspiration for those who attempt to solve problems involving binomial coefficients.

References

[1] H. W. Gould: *Combinatorial Identities* (*Revised Edition*), Morgantown Printing, Morgantown, West Virginia, 1972.

[2] J. Riordan: *Combinatorial Identities*, Robert E. Krieger Publishing Company, New York, 1979.

Henry Ricardo received his Ph.D. from Yeshiva University in New York. Since retiring from The City University of New York as Professor of Mathematics, he has been a volunteer with the Westchester Area Math Circle (Purchase, NY). He is an avid problemist and regularly attempts problems in more than a dozen journals worldwide.

The City University of New York and the Westchester Area Math Circle
E-mail address: odedude@yahoo.com

Irish Math. Soc. Bulletin
Number 91, Summer 2023, 83–86
ISSN 0791-5578

PROBLEMS

IAN SHORT

Problems

The first problem this issue was proposed by Toyesh Prakash Sharma of Agra College, India.

Problem 91.1. Prove that the angles α, β and γ of a triangle satisfy

$$\big(1+\sec^2\alpha+\sec^2\beta+\sec^2\gamma\big)(1-\cos\alpha\cos\beta\cos\gamma)\geqslant 8$$

The second problem is from Des MacHale of University College Cork.

Problem 91.2. Prove that the perimeter P and area A of a cyclic quadrilateral satisfy

$$P^2\geqslant 16A,$$

with equality if and only if the cyclic quadrilateral is a square.

The third problem was proposed by Tran Quang Hung of the Vietnam National University at Hanoi, Vietnam.

Problem 91.3. Let $A_0,A_1,\dots,A_n$ be the vertices of a simplex in n-dimensional Euclidean space for which the edges $A_0A_1,A_0A_2,\dots,A_0A_n$ are mutually perpendicular. Let B_i be the centroid of the set of points $\{A_0,A_1,\dots,A_n\}\setminus\{A_i\}$, for $i=0,1,\dots,n$. Consider any point C other than A_0 for which the line through A_0 and C is perpendicular to the hyperplane spanned by $A_1,A_2,\dots,A_n$, and let P be the midpoint of the segment B_0C. Prove that all distances PB_i are equal, for $i=1,2,\dots,n$.

Solutions

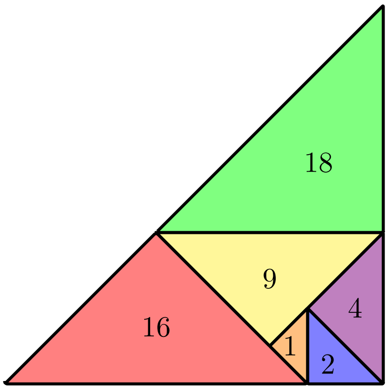
Here are solutions to the problems from *Bulletin* Number 89.

The first problem was solved by Ryan Quinn, Seán Stewart of the King Abdullah University of Science and Technology, Saudi Arabia, the North Kildare Mathematics Problem Club, and the proposer Des MacHale. We present the solution and figures of Seán Stewart.

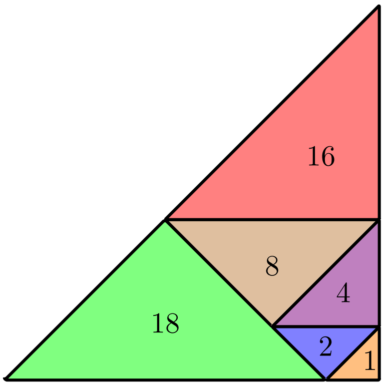
*Problem 89.1***.** It is well known that it is possible to dissect a square into a finite number of different squares, but that it is not possible to dissect an equilateral triangle into a finite number of different equilateral triangles. Determine whether it is possible to dissect an isosceles right-angled triangle into a finite number of different isosceles right-angled triangles.

Solution 89.1. One example is shown below. Here an isosceles right-angled triangle with common side length of 10 units is dissected into 6 different isosceles right-angled triangles. The numbers in the triangles record areas.

Received 17-6-2023
DOI:10.33232/BIMS.0091.83.86.



We also give a second example. Here an isosceles right-angled triangle with common side length of $7\sqrt{2}$ units is again dissected into 6 different isosceles right-angled triangles.



$\square$

Stewart references a paper by Skinner, Smith, and Tutte (*Journal of Combinatorial Theory*, Series B **80**, 2000) for related results, and there are further publications and internet resources available. He also asks what is the minimum number of triangles in a dissection that answers Problem 89.1.

The second problem was solved by Daniel Văcaru of Pitești, Romania, Brian Bradie of Christopher Newport University, USA, Seán Stewart, Henry Ricardo of the Westchester Area Math Circle, USA, the North Kildare Mathematics Problem Club, and the proposer, Toyesh Prakash. We present the solution of Brian Bradie (other solutions were similar).

Problem 89.2. Prove that

$$\int_{-\pi/2}^{\pi/2} \cos^2(\tan x)\, dx = \frac{\pi}{2}(1+e^{-2}).$$

Solution 89.2. With the substitution $u = \tan x$ and the half-angle identity

$$\cos^2 u = \frac{1}{2} + \frac{1}{2}\cos 2u,$$

we have

$$\begin{aligned}
\int_{-\pi/2}^{\pi/2} \cos^2(\tan x)\, dx &= \int_{-\infty}^{\infty} \frac{\cos^2 u}{1+u^2}\, du \\
&= \frac{1}{2}\int_{-\infty}^{\infty} \frac{1}{1+u^2}\, du + \frac{1}{2}\int_{-\infty}^{\infty} \frac{\cos 2u}{1+u^2}\, du \\
&= \frac{\pi}{2} + \frac{1}{2}\int_{-\infty}^{\infty} \frac{\cos 2u}{1+u^2}\, du.
\end{aligned}$$

Now,

$$\int_{-\infty}^{\infty} \frac{\cos 2u}{1+u^2}\, du = \operatorname{Re} \int_{-\infty}^{\infty} \frac{e^{2ui}}{1+u^2}\, du,$$

and

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{e^{2ui}}{1+u^2}\, du &= \lim_{R\to\infty} \int_{C_R} \frac{e^{2ui}}{1+u^2}\, du \\ &= 2\pi i \ \operatorname{Res}\left(\frac{e^{2ui}}{1+u^2}; u=i\right) \\ &= 2\pi i \times \frac{e^{-2}}{2i} = \frac{\pi}{e^2},\end{aligned}$$

where C_R is the contour $(-R,R) \cup \{Re^{i\theta} : 0 \leqslant \theta \leqslant \pi\}$. Thus,

$$\int_{-\pi/2}^{\pi/2} \cos^2(\tan x)\, dx = \frac{\pi}{2} + \frac{\pi}{2e^2} = \frac{\pi}{2}(1+e^{-2}). \qquad \square$$

The third problem comes from Finbarr Holland of University College Cork. It was solved by Seán Stewart, Daniel Văcaru, the North Kildare Mathematics Problem Club, and the proposer. We present the solution of the problem club.

Problem 89.3. Let a_k and b_k be positive real numbers with $a_k < b_k$, for $k = 1, 2, \ldots, n$, and let

$$r_n(z) = \prod_{k=1}^{n} \frac{b_k + z}{a_k + z}.$$

Prove that

$$\int_{-\infty}^{\infty} \log|r_n(ix)|\, dx = \pi \sum_{k=1}^{n} (b_k - a_k).$$

Solution 89.3. We will prove the result for $n = 1$; the general result then follows because $\log|x|$ converts products into sums. Let $r(x) = r_1(x)$, $a = a_1$, and $b = b_1$.

Notice that

$$\log|r(ix)| = \log\left|\frac{b+ix}{a+ix}\right| = \log\sqrt{\frac{b^2+x^2}{a^2+x^2}} = \frac{1}{2}\left(\log(b^2+x^2) - \log(a^2+x^2)\right).$$

Integration by parts shows that the antiderivitive of $\log(a^2+x^2)$ is

$$x\log(a^2+x^2) + 2a\tan^{-1}(x/a) - 2x.$$

Hence

$$\int_{-\infty}^{\infty} \log|r(ix)|\, dx = \frac{1}{2} \lim_{R\to\infty} \left[x\log\left(\frac{b^2+x^2}{a^2+x^2}\right) + 2b\tan^{-1}(x/b) - 2a\tan^{-1}(x/a)\right]_{-R}^{R}.$$

The term involving the logarithm tends to zero at both limits, by L'Hôpital's rule, so we are left with

$$\int_{-\infty}^{\infty} \log|r(ix)|\, dx = \frac{1}{2}\Big((2b-2a) \times \frac{\pi}{2} - (2b-2a) \times \Big(-\frac{\pi}{2}\Big)\Big) = \pi(b-a). \qquad \square$$

Editor's remark: I added the assumption that a_k and b_k are positive to the problem, which is necessary if the solution of the integral is to retain its present form. The requirement that $a_k < b_k$ is redundant. Finbarr points out that for real numbers a_k and b_k (not necessarily positive), we have

$$\int_{-\infty}^{\infty} \log|r_n(ix)|\, dx = \pi \sum_{k=1}^{n} (|b_k| - |a_k|).$$

We invite readers to submit problems and solutions. Please email submissions to `imsproblems@gmail.com` in any format (we prefer LaTeX). Submissions for the summer Bulletin should arrive before the end of April, and submissions for the winter Bulletin should arrive by October. The solution to a problem is published two issues after the issue in which the problem first appeared. Please include solutions to any problems you submit, if you have them.

School of Mathematics and Statistics, The Open University, Milton Keynes MK7 6AA, United Kingdom

Instructions to Authors

Papers should be submitted by email to the address:

`ims.bulletin@gmail.com`

In the first instance, authors may submit a pdf version of their paper. Other formats such as MS/Word or RTF are not acceptable. The *Bulletin* is typeset using PDF files produced from LaTeX source; therefore, authors must be ready to supply LaTeX source files (and any ancillary graphics or other files needed) should their paper be accepted. Source files should be suitable for processing using `pdflatex`.

Once a paper is accepted in final form, the author(s) will be responsible for producing an output according to the *Bulletin's* standard layout. Standard template files for articles, abstracts and reviews, and the necessary class and style files may be downloaded from the IMS website `http://www.irishmathsoc.org`, or obtained from the editor in case of difficulty.

Since normally no proofs are sent out before publication, it is the author's responsibility to check carefully for any misprints or other errors.

The submission of a paper carries with it the author's assurance that the text has not been copyrighted or published elsewhere (except in the form of an abstract or as part of a published lecture or thesis); that it is not under consideration for publication elsewhere; that its submission has been approved by all coauthors and that, should it be accepted by the *Bulletin*, it will not be published in another journal. After publication, copyright in the article belongs to the IMS. The IMS will make the pdf file of the article freely available online. The Society grants authors free use of this pdf file; hence they may post it on personal websites or electronic archives. They may reuse the content in other publications, provided they follow academic codes of best practice as these are commonly understood, and provided they explicitly acknowledge that this is being done with the permission of the IMS.