European Mathematical Society Magazine

ISSN 2747-7894 / Issue 119 / March 2021

EMS Magazine

Pedro J. Freitas
The geometric theorems of Almada Negreiros

Karine Chemla
All roads come from China – For a theoretical approach to the history of mathematics

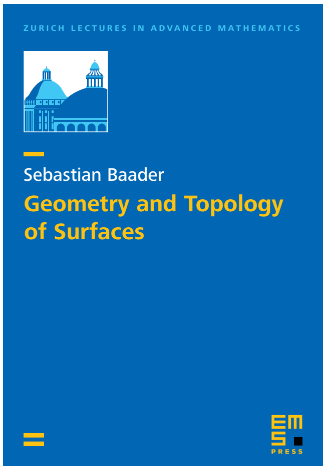
Irene Fonseca and Giovanni Leoni
Surface evolution of elastically stressed films

Franka Miriam Brueckler
Croatian Mathematical Society

EUROPEAN MATHEMATICAL SOCIETY

New EMS Press books

20% discount on any book purchases for individual members of the EMS, member societies or societies with a reciprocity agreement when ordering directly from EMS Press.

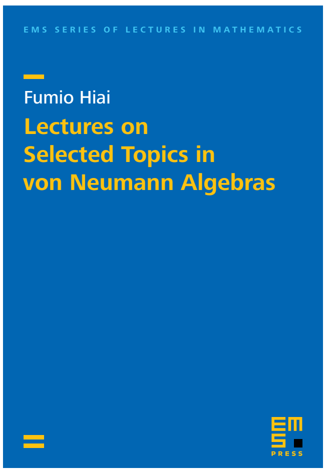


Sebastian Baader (Universität Bern, Switzerland)
Geometry and Topology of Surfaces

Zurich Lectures in Advanced Mathematics
ISBN print 978-3-98547-000-6, ISBN online 978-3-98547-500-1
March 2021. Softcover. 86 pages. €29.00

These lecture notes cover the classification of hyperbolic structures and measured foliations on surfaces in a minimalist way. While the inspiration is obviously taken from the excellent books *Primer on mapping class groups* and *Travaux de Thurston sur les surfaces*, we tried to include a little bit more of hyperbolic trigonometry, including a proof of Basmajian's identity on the orthogeodesic spectrum, while keeping the rest short.

The book is written for students and researchers working in low-dimensional topology and geometry.



Fumio Hiai (Tohoku University, Japan)
Lectures on Selected Topics in von Neumann Algebras

EMS Series of Lectures in Mathematics
ISBN print 978-3-98547-004-4, ISBN online 978-3-98547-504-9
March 2021. Softcover. 250 pages. €39.00

The theory of von Neumann algebras, originating with the work of F. J. Murray and J. von Neumann in the late 1930s, has grown into a rich discipline with connections to different branches of mathematics and physics. Following the breakthrough of Tomita–Takesaki theory, many great advances were made throughout the 1970s by H. Araki, A. Connes, U. Haagerup, M. Takesaki and others.

These lecture notes aim to present a fast-track study of some important topics in classical parts of von Neumann algebra theory that were developed in the 1970s. Starting with Tomita–Takesaki theory, this book covers topics such as the standard form, Connes' co-cycle derivatives, operator-valued weights, type III structure theory and non-commutative integration theory.

The self-contained presentation of the material makes this book useful not only to graduate students and researchers who want to know the fundamentals of von Neumann algebras, but also to interested undergraduates who have a basic knowledge of functional analysis and measure theory.

EMS Press
European Mathematical Society – EMS – Publishing House GmbH
Institut für Mathematik, Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin, Germany

Website
https://ems.press

Contact
orders@ems.press

EMS Magazine 119

European Mathematical Society Magazine

Published by EMS Press, an imprint of the
European Mathematical Society – EMS – Publishing House GmbH
Technische Universität Berlin, Institut für Mathematik,
Straße des 17. Juni 136, 10623 Berlin, Germany

https://ems.press

Graphic design and typesetting: Christoph Eyrich, Berlin
Printing: Beltz Bad Langensalza GmbH, Bad Langensalza, Germany

ISSN (Print) 2747-7894
ISSN (Online) 2747-7908

The illustration on the cover makes reference to the work of the Portuguese artist Almada Negreiros (1893–1970), whose geometric constructions are the subject of an article by Pedro Freitas in the present issue. The background figure comes from the geometric construction of the painting *Porta da Harmonia* (oil on canvas, 1957), which translates in English as "door of harmony". This is the first in a series of four geometric works by the artist, the other three being *O Ponto de Bauhütte*, *Quadrante I* and *Relação 9/10*, all currently part of the collection of the Calouste Gulbenkian Museum in Lisbon, Portugal. The foreground figure is a play on the iconic photo of Almada Negreiros in the "flight suit" that he wore at the *Sessão Futurista* in April 1917, the first futurist conference in Portugal. *António B. Araújo*

A message from the president

Dear members of the EMS,

Welcome to the first issue of the new *EMS Magazine*. It features new content directions and introduces for the first time the new logo of the EMS. The magazine is planned to appear 'online first' and then will be combined into quarterly issues.

As the last year of living with a pandemic has shown, our usual procedures on how to perform research and communication in the mathematics community are very vulnerable. This means, in particular, that further conferences and meetings have to be cancelled, postponed or held virtually.

This concerns the EMS 30th anniversary which we want to organize in presence and so we have postponed it further. The joint congress of AMS–EMS–SMF in Grenoble that was planned for 2021 has been postponed to 2022 as well. It will take place right after the ICM, in order to create synergies and to avoid too much extra international travel.

We have also just discussed the fate of the European Congress in Portorož and decided that it will be an online-only conference. We truly hope that the mathematical community will still join the conference as if it was a real conference in presence and show its solidarity with the local organisors who put so much work and effort into this.

After many struggles with data protection regulations we have decided to implement a completely new internet performance for the EMS in which we also move to the new layout and new logo, the prototype version is up and running, see euromathsoc.org.

Another very positive development is that our publishing house EMS Press has moved to the subscribe-to-open model and was successful in obtaining sufficiently many subscriptions so that all ten EMS journals could be turned open for the year 2021, see ems.press/updates/2021-02-01-10-titles-announced-oa-s2o.

Finally, as was previously announced, Zentralblatt became an open-access platform from 1st January 2021. The mathematical community is invited to participate in its further development, see zbmath.org.

I wish all of you a healthy spring.

Volker Mehrmann
President of the EMS

Brief words from the editor-in-chief

Dear readers of the EMS Magazine,

I am sure you are excited to browse and read this first issue of the new *EMS Magazine*, so I will not take too long in these brief words. It is always a big challenge to improve on something that is already well done, as the *Newsletter* was. So I sincerely hope you find that in this new Magazine format, the challenge has been successfully met.

This first issue of the *Magazine* is also the first one to have, as editors, three new colleagues whose names and biographical notes you can read on page 67. We continue the publication of articles by the 2020 EMS Prize winners, this time with articles by Ana Caraiani and Simion Filip, and also by the 2020 Otto Neugebauer Prize winner Karine Chemla. And there is also the first article in a new section about Art and Mathematics.

But it is better for me to stop writing and let you start reading your Magazine. We all hope you enjoy it!

Fernando Pestana da Costa
Editor-in-chief

The geometric theorems of Almada Negreiros

Pedro J. Freitas

In the 20th century, some artists took to using mathematical concepts (such as the golden ratio) in their works, in the belief that these would encapsulate a certain form of universal beauty. Portuguese artist Almada Negreiros was among them. However, he did more than absorb mathematical elements, he actually proved some mathematical results about them. This paper addresses some of these discoveries, setting them in the context of the author's views on mathematics and art.

José de Almada Negreiros (São Tomé and Príncipe, 1893 – Lisbon, 1970) was a key figure of 20th century Portuguese culture, in both visual arts and literature. His visual work went through several stages: starting with mostly figurative work, he became increasingly closer to geometric abstractionism, which he came to adopt completely in 1957, in four works displaying simple geometric figures in black and white.

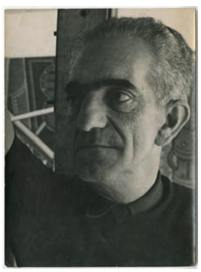


Figure 1. Almada Negreiros in the late 1940s

This progressive change in style was not just the result of an aesthetic choice, but also the consequence of a way of thinking about the relationship between art and geometry. In fact, Almada – this was his own choice of name – believed in a universal geometric system, underlying all visual art, throughout time. He called his system "The Canon".

This belief in the universality of mathematics as a foundation for art may remind us of Le Corbusier's belief that his Modulor system, based on three simple concepts – unit, the double and the golden section – would tap into an abstract and universal form of beauty. Other authors of his time, such as Matila Ghyka, also developed similar lines of thought.

Almada's system involved several geometrical elements, such as rectangles with known proportions – such as $\sqrt{2}, \varphi$ (the golden rectangle), $\sqrt{3}, 2, \sqrt{5}, \sqrt{\varphi}$ – divisions of the circle into equal parts, and the golden angle. These were used by Almada to describe and understand artistic artifacts, and were seen as a sign that there was a collection of such constructions that was used, consciously or unconsciously, by artists of all styles and origins.

However, also like Le Corbusier, Almada also actually proved mathematical statements. In his book *Le Modulor* [2, p. 37] Le Corbusier suggests a construction for a right angle, placed within a rectangle, according to certain rules. The construction (which is not presented in full clarity and detail) actually does lead to an approximate right angle.

Almada goes much farther in his speculative geometry. In two collections of drawings as well as some artist's notebooks, comprising more than a hundred completed works as well as many additional sketches, he presents constructions of a geometric nature which can rightly be regarded as artworks, but which are at the same time geometric results related to the elements of the Canon, showing their intrinsic proximity. Figures 2 and 3 present two examples from the collection "Language of the square", a set of about forty-five finished drawings on paper, 50 × 70 cm.

Both drawings can be regarded as protocols for geometric constructions, both starting with a square divided in two equal parts by a horizontal line, with an inscribed quarter circle. The construction in Figure 2 leads to two red lines, marked $\sqrt{5}$ and φ, indicating that these lines are diagonals of rectangles with those proportions, and sides parallel to the sides of the square.

It is not difficult to verify the correctness of the proportions. If we take half of the side of the square as our unit measure, then, by the Pythagorean Theorem, the green diagonal measures $\sqrt{5}$, and as this measure is transported, by compass, to the top side of the square, we do obtain a rectangle with this proportion.

The small green horizontal line would then measure $\sqrt{5}-2$, and as we add it to half the square, we get a length of $1+\sqrt{5}-2 = \sqrt{5}-1$. The proportion of the rectangle having as diagonal the red line marked φ would then be

$$\frac{2}{\sqrt{5}-1} = \frac{2(\sqrt{5}+1)}{4} = \frac{1+\sqrt{5}}{2} = \varphi.$$

The numbers 9 and 10 on the right refer to the divisions of the circle into 9 and 10 parts, which are achieved by the intersection

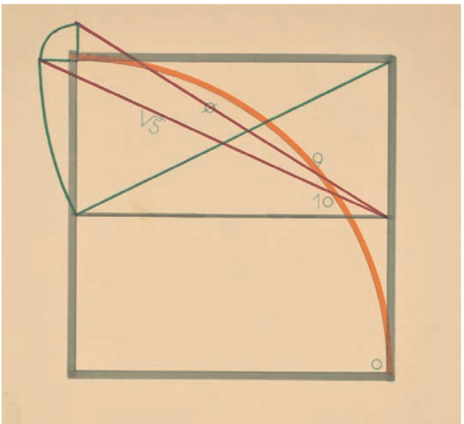


Figure 2

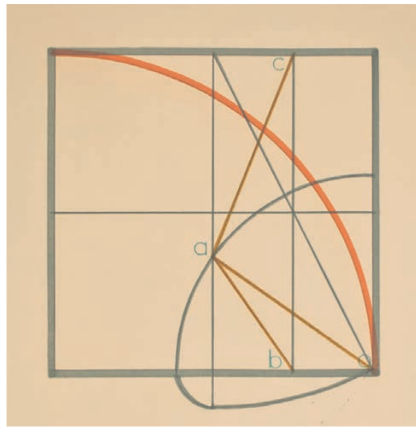


Figure 3

of the red lines with the circle arc. As it happens, the proportions of the rectangles are precise, but the divisions of the circle are not: they represent very good approximations, each having an error of about 0.7 %. In the book [4], the authors present a few more analyses of this type concerning about 30 of Almada's drawings.

In Figure 3, a similar but more intricate construction leads to four points marked a, b, c and o, and three lengths, which Almada interprets as follows:

$$ab = \frac{\overline{\odot}}{14}, \qquad ao = \frac{\overline{\odot}}{10}, \qquad ac = \frac{\overline{\odot}}{9}.$$

This is again a reference to the division of the circle into equal parts. In this case, the 10th part is exact, the 14th part has an error of 1 % and the 9th part has an astonishingly tiny error of 0.001 %.[1]

In the history of mathematics, the problem of dividing the circle into n equal parts with straightedge and compass has a respectable place. Thinking about prime values of n, it has known since ancient Greece that it is possible to divide the circle into 3 and 5 parts, but no method was found for 7, 11 or 13 parts. It was Gauss, in 1796, who proved that it was possible to divide the circle into 17 parts, a result that came to be included in Section VII of *Disquisitiones Arithmeticae*.[2] Eventually, he found a sufficient condition for the division into n parts, and stated that this condition should also be necessary. In 1837, Pierre Wantzel proved the necessity, leading to the result that became known as the Gauss–Wantzel theorem.

Gauss–Wantzel's theorem. *A circle can be divided into n parts with straightedge and compass if and only if*

$$n = 2^k p_1 \cdots p_t$$

where $p_1, \ldots, p_t$ *are distinct* Fermat primes, *that is, primes of the form* $2^{2^r} + 1$. *The only known Fermat primes are* $3, 5, 17, 257$ *and* 62237.

Going back to Almada's constructions, we note that 7 is not a Fermat prime, and $9 = 3 \times 3$ (3 is a Fermat prime that occurs twice in the factorization). Therefore, a circle cannot be divided precisely into 9 and 14 parts using only straightedge and compass.

The purpose of these constructions is mainly to show that the various elements of Almada's Canon have a natural and harmonious relationship amongst themselves, which is revealed by the elegance and simplicity of the geometric constructions he presents. Thus, the aim is primarily symbolic and philosophical rather than mathematical. Nevertheless, these drawings also present original constructions for the divisions of the circle and for producing rectangles with a given proportion, so in fact, they represent original mathematical results, even though some of them are approximate.

Many more drawings exist, some of them leading to general statements which one might regard as theorems, if it weren't for the fact that they represent approximations. One of these statements is

$$2\frac{\overline{\odot}}{9} + \frac{\overline{\odot}}{10} = 2r.$$

The meaning of this equality is that 2 chords of the 9th part of the circle plus a chord of the 10th part will equal the diameter of the circle (two times the radius). The first side of the equation is actually equal to $1.986r$, an error of 0.7 %.

The presence of the numbers 9 and 10 in this equality (and in many geometric constructions) is not fortuitous. Almada believed

[1] The article [3] compares the approximation for the 9th part of the circle achieved by this construction with that of other constructions and concludes this is one is the best of all those analysed.

[2] There is an anecdote that Gauss was so pleased by this result that he requested that a regular heptadecagon be inscribed on his tombstone. The stonemason declined, stating that the difficult construction would essentially look like a circle.

there was a special connection between these two numbers, which he called the "Relation 9/10" and strove to seek for them and to connect them with other elements of his Canon. This is also the reason for the presence of these numbers in Figures 2 and 3.

The statement above is used in an approximate construction for the pentagram (another of Almada's favourite figures, because of its relation to the Pythagoreans). The construction starts with a circle with two elements marked: a diameter and a 9th part, measured from one of the extremes of the diameter. In Figure 4, the 9th part is the arc AP_9, on the left, and the diameter is marked AB.

The construction now goes as follows. The arc of a circle centred at A is drawn from the point P_9 to the diameter, marking point D, through which one of the lines of the pentagram is drawn: CE, perpendicular to AB. This point D is now the centre of the half circle AF. Point F now determines a new arc of a circle with centre B, yielding points G and H, through which the remaining lines of the pentagram are defined.

According to this construction, lines AD and DF are chords of the 9th part of the circle, and line FB is the chord of the 10th part (which is the chord of both arcs BG and GH). These three lines add up to a diameter, which illustrates the previous equation, connecting it to the pentagram.

We emphasize that this is not an accurate construction – it is actually impossible for it to be accurate, according to Gauss–Wantzel's theorem, since otherwise the 9th part of the circle would be constructible, if we could start with a pentagon and extract point P_9 from it.

The fact that some of these results are approximations, and that these appear among exact results with no distinction, is actually quite revealing of Almada's methods. We have reason to believe that Almada didn't actually compute the exact measures of the elements he claimed to produce, but probably only checked them visually, using instruments of measure or other elements of comparison. So, his way of establishing geometric results is not the same as the one used by mathematicians. And even though Almada was aware that some of his results were approximate, he stuck to them and never distinguished between exact and approximate ones. This was probably because he was more interested in the visual aspects of such results, and for this effect, some approximations are acceptable.

One of his last artworks, which can be considered his geometric legacy and a summary of many of his statements on this subject, is the mural *Começar* (*To begin*, 1968), which is located in the main hall of the building of the Gulbenkian Foundation in Lisbon. It is a very intricate collection of lines and circles, inscribed in stone, with dimensions 12.87×2.31 m (see Figure 5).

The mural is usually divided into five parts. The first one displays the pentagram with the construction presented in Figure 4 as the main motif. Then, we find a 16-point star, which is an allusion to a drawing by Leonardo da Vinci appearing (apparently by mistake) in Geneva's codex of Luca Pacioli's *Divina Proportione*. The central element is again a pentagram, which Almada associates to a coin, minted by Portugal's first king, which is set in the midst of many other constructions that Almada used to study some 15th century Portuguese paintings. The right part of the mural presents references to the Minoan civilization and to a medieval poem associated with a guild of cathedral stonemasons. There is a virtual guided tour at gulbenkian.pt/almada-comecar/en/ where the reader can find more detailed information about this mural. However, what we wish to point out with these brief remarks is that Almada's geometry is truly an effort to unite all art and all cultures.

A thorough study of the geometric works in Almada's estate has been undertaken, in a collaboration between Simão Palmeirim and the author; some of its results can be found in [4]. We hope that this study can bring to light not only the remarkable visual

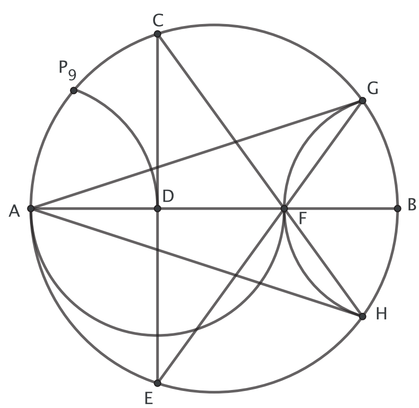


Figure 4. A construction of the pentagram from the 9th part of the circle

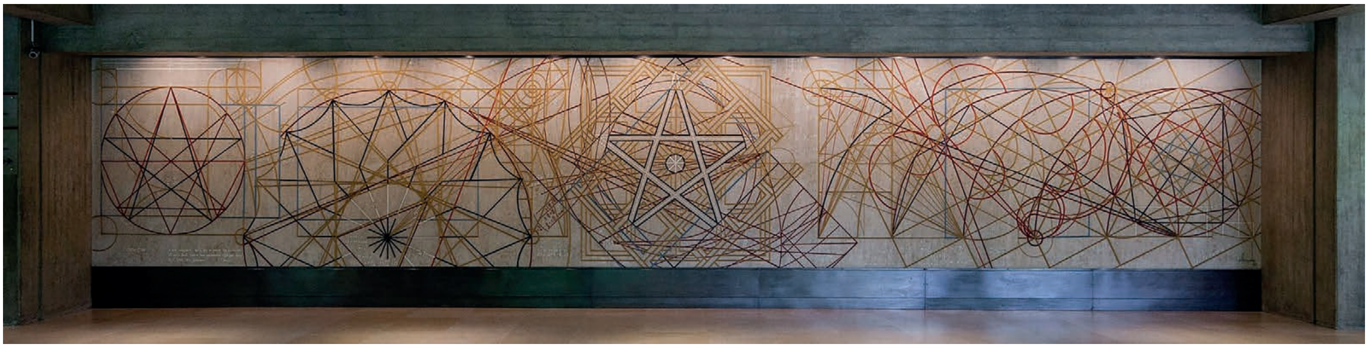


Figure 5. Mural *Começar* by Almada Negreiros (1968)

aspects of Almada's geometric work, but also the mathematics behind it, which not only yield new geometric results, but also represent a powerful statement about the author's thought regarding geometry and art.

Acknowledgements. We wish to thank the family of Almada Negreiros for allowing us to study his estate and for giving permission to use the reproductions in Figures 1, 2 and 3. This study was done within the scope of project *Modernismo Online*, Universidade Nova de Lisboa (www.modernismo.pt), which aims to collect in digital form the heritage of Portuguese Modernism.

Figure 5 is by Manuel V. Botelho (commons.wikimedia.org/w/index.php?curid=35070536), CC BY-SA 4.0 (creativecommons.org/licenses/by-sa/4.0/).

The author is supported by FCT, I.P., through Project UID/HIS/00286/2019.

References

[1] S. P. Costa and P. J. Freitas, *Livro de Problemas de Almada Negreiros*, Leituras em Matemática, 14. Portuguese Mathematical Society (2015)

[2] Le Corbusier, *Le Modulor*. Birkhäuser Architecture, 2000.

[3] P. J. Freitas, Almada Negreiros and the regular nonagon. *Recreat. Math. Mag.* 39–51 (2015)

[4] P. J. Freitas and S. P. Costa, Almada Negreiros and the geometric canon. *J. Math. Arts* **9**, 27–36 (2015)

Pedro J. Freitas has a PhD in mathematics and works in the Department of History and Philosophy of Sciences in the University of Lisbon. He is a member of the Interuniversity Center for the History and Philosophy of Sciences. His teaching and research are mainly related to the history of mathematics, recreational mathematics and relations between mathematics and art. He is also involved in mathematical outreach.

pjfreitas@fc.ul.pt

New frontiers in Langlands reciprocity

Ana Caraiani

In this survey, I discuss some recent developments at the crossroads of arithmetic geometry and the Langlands programme. The emphasis is on recent progress on the Ramanujan–Petersson and Sato–Tate conjectures. This relies on new results about Shimura varieties and torsion in the cohomology of locally symmetric spaces.

The Langlands programme is a "grand unified theory" of mathematics: a vast network of conjectures that connect number theory to other areas of pure mathematics, such as representation theory, algebraic geometry, and harmonic analysis.

One of the fundamental principles underlying the Langlands conjectures is *reciprocity*, which can be thought of as a magical bridge that connects different mathematical worlds. This principle goes back centuries to the foundational work of Euler, Legendre and Gauss on the law of quadratic reciprocity. A celebrated modern instance of reciprocity is the correspondence between modular forms and rational elliptic curves, which played a key role in Wiles's proof of Fermat's Last Theorem [46] and which relied on the famous Taylor–Wiles method for proving modularity [43]. Recently, the search for new reciprocity laws has begun to expand the scope of the Langlands programme.

The *Ramanujan–Petersson conjecture* is an important consequence of the Langlands programme, which goes back to a prediction Ramanujan made a century ago about the size of the Fourier coefficients of a certain modular form Δ, a highly symmetric function on the upper half plane. The *Sato–Tate conjecture* is an equidistribution result about the number of points of a given elliptic curve modulo varying primes, formulated half a century ago. It is also a consequence of the Langlands programme. In Section 1, I survey progress on these conjectures in two fundamentally different settings: one setting in which there is a direct connection to algebraic geometry (*modular curves*) and one setting in which such a connection is missing (arithmetic hyperbolic 3-manifolds, or *Bianchi manifolds*).

Shimura varieties are certain highly symmetric algebraic varieties that generalise modular curves and that provide, in many cases, a geometric realisation of Langlands reciprocity. In Section 2, I explain a new tool for understanding Shimura varieties called the *Hodge–Tate period morphism*. This was introduced by Scholze in [35] and refined in my joint work with Scholze [16]. I then discuss vanishing theorems for the cohomology of Shimura varieties proved using the geometry of the Hodge–Tate period morphism [16, 17].

The *Calegari–Geraghty method* [11] vastly extends the scope of the Taylor–Wiles method, though it is conjectural on an extension of the Langlands programme to incorporate torsion in the cohomology of locally symmetric spaces. In Section 3, I discuss joint work with Allen, Calegari, Gee, Helm, Le Hung, Newton, Scholze, Taylor, and Thorne [1], where we implement the Calegari–Geraghty method unconditionally over *CM fields*, an important class of number fields that contains imaginary quadratic fields as well as cyclotomic fields. This work relies crucially on one of the vanishing theorems mentioned above [17], and has applications to both the Ramanujan–Petersson and the Sato–Tate conjectures over CM fields.

Remark 1. The Langlands programme is a beautiful but technical subject, with roots in many different areas of mathematics. For a general mathematician, Section 1 is the most accessible, as it highlights two concrete consequences of the Langlands conjectures. The later Sections 2 and 3 assume more background in algebraic geometry and number theory.

I have prioritised references to well-written surveys above references to the original papers. I particularly recommend [21] for a historical account of Langlands reciprocity, [41] for more background on the Langlands correspondence, and [36] for a cutting-edge account of the deep connections between arithmetic geometry and the Langlands programme.

1 The Ramanujan and Sato–Tate conjectures

1.1 Modular curves and Bianchi manifolds

The goal of this section is to discuss two fundamental examples of locally symmetric spaces: *modular curves*, which have an algebraic structure, and *Bianchi manifolds*, which do not. This dichotomy underlies the fundamental difference between reciprocity laws over

EMS MAGAZINE 119 (2021) — DOI 10.4171/MAG-3

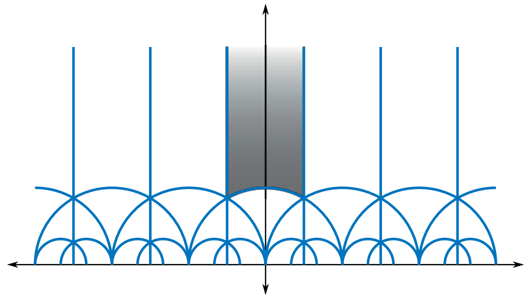


Figure 1. A fundamental domain for $SL_2(\mathbb{Z})$ acting on $\mathbb{H}^2$

the field of rational numbers $\mathbb{Q}$ (and over real quadratic fields such as $\mathbb{Q}(\sqrt{5})$), and reciprocity laws over imaginary quadratic fields such as $\mathbb{Q}(i)$.

Let G be a connected reductive group defined over $\mathbb{Q}$, for example SL_n, GL_n or Sp_{2n}. We can then consider an associated *symmetric space* X, endowed with an action of the real points $G(\mathbb{R})$. This is roughly identified with $G(\mathbb{R})/K_\infty$, where $K_\infty \subset G(\mathbb{R})$ is a maximal compact subgroup. We then want to consider the action of certain arithmetic groups on X: more precisely we want to restrict to finite index subgroups $\Gamma \subset G(\mathbb{Z})$ cut out by congruence conditions. If Γ is sufficiently small, we can form the quotient $\Gamma\backslash X$ and obtain a smooth orientable Riemannian manifold, which is a *locally symmetric space* for G.

Example 2. If $G = SL_2/\mathbb{Q}$, the corresponding symmetric space is the upper-half plane

$$SL_2(\mathbb{R})/SO_2(\mathbb{R}) \simeq \mathbb{H}^2 := \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$$

endowed with the hyperbolic metric. The action of $SL_2(\mathbb{R})$ on $\mathbb{H}^2$ is by Möbius transformations:

$$z \mapsto \frac{az+b}{cz+d} \quad \text{for} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R}).$$

For $\Gamma \subset SL_2(\mathbb{Z})$ a finite index congruence subgroup (that is assumed sufficiently small), the quotients $\Gamma\backslash\mathbb{H}^2$ are Riemann surfaces. These Riemann surfaces come from algebraic curves X_Γ defined over $\mathbb{Q}$ (or over finite extensions of $\mathbb{Q}$) called *modular curves*. A fundamental domain for a proper subgroup $\Gamma \subset SL_2(\mathbb{Z})$ acting on $\mathbb{H}^2$ is a finite union of translates of the fundamental domain in Figure 1.

Example 3. If $G = SL_2/F$, where F is an imaginary quadratic field[1], the corresponding symmetric space is 3-dimensional hyperbolic space

$$SL_2(\mathbb{C})/SU_2(\mathbb{R}) \simeq \mathbb{H}^3$$

and the locally symmetric spaces are called *Bianchi manifolds*. These are arithmetic hyperbolic 3-manifolds and, since their real dimension is odd, they do not admit a complex or algebraic structure.

The locally symmetric spaces for a group G are important in what follows because they give a way to access *automorphic representations* of G, the central objects of study in the Langlands programme. This is explained more in Section 2. For example, *modular forms*[2], which are holomorphic functions on $\mathbb{H}^2$ that satisfy a transformation relation under some Γ, contribute to the first Betti cohomology of modular curves (with possibly twisted coefficients).

Some locally symmetric spaces have an algebraic structure. If this happens, they in fact come from smooth, quasi-projective varieties X_Γ defined over number fields, which are called *Shimura varieties*. The geometry of Shimura varieties is a rich and fascinating subject in itself, that we discuss more in Section 2. On the other hand, the Langlands programme is much more mysterious beyond the setting of Shimura varieties, because there is no obvious connection to algebraic geometry or arithmetic. We discuss this more in Section 3.

1.2 *The Ramanujan conjecture*

A famous example of a modular form is Ramanujan's Δ function. If z is the variable on the upper-half plane $\mathbb{H}^2$ and $q = e^{2\pi i z}$, Δ is given by the Fourier series expansion

$$\Delta(z) = q\prod_{n=1}^{\infty}(1-q^n)^{24} = \sum_{n>0}\tau(n)q^n.$$

[1] This can be viewed as a connected reductive group over $\mathbb{Q}$ using a technical notion called the Weil restriction of scalars.

[2] These give rise to automorphic representations for the group $SL_2/\mathbb{Q}$.

In 1916, Ramanujan made three predictions about the behaviour of the Fourier coefficients $\tau(n)$. The first two of these were immediately proved by Mordell by studying the action on Δ of certain *Hecke operators*, that we return to in Section 2. The *Ramanujan conjecture*, which resisted attempts at proof for much longer, bounds the absolute value of the Fourier coefficients: it states that $|\tau(p)| \leq 2p^{11/2}$ for all primes p.

Deligne finally established this bound in the early 1970's, and this was one of the reasons for which he was awarded a Fields Medal in 1978. While the bound on the Fourier coefficients is purely a statement within harmonic analysis, the proof used the bridge of Langlands reciprocity and was ultimately obtained from a statement in arithmetic geometry. More precisely, Deligne's proof of the Ramanujan conjecture went via the étale cohomology of modular curves, obtaining the desired bound as a consequence of his proof of the Weil conjectures for smooth projective varieties over finite fields.

The generalised *Ramanujan–Petersson conjecture* is a vast extension of the above statement, with numerous applications across mathematics and computer science. See, for example, the survey [31] for its applications to extremal combinatorial objects called Ramanujan graphs. This more general conjecture, which is part of Arthur's conjectures on the automorphic spectrum of GL_n (see also the survey [34]), predicts that the local components at finite places of cuspidal automorphic representations of GL_n are *tempered*.

Temperedness means roughly that the matrix coefficients of the representation are in $L^{2+\epsilon}$ for all $\epsilon > 0$. This singles out the building blocks of the category of irreducible admissible representations of p-adic groups, such as $\mathrm{GL}_n(\mathbb{Q}_p)$, in the sense that everything else can be constructed from tempered representations of smaller groups. Tempered representations also play an important role in the *local Langlands conjecture*, which relates them to arithmetic objects, essentially representations of local Galois groups. For the group GL_n, local Langlands is a theorem, proved by Harris–Taylor and Henniart in the early 2000's, and later reproved by Scholze.

For certain cuspidal automorphic representations of GL_n, which are global objects built from the irreducible admissible representations mentioned above, one can try to follow Deligne's approach to the Ramanujan conjecture using the étale cohomology of higher-dimensional Shimura varieties. When these varieties have singular reduction, the arithmetic counterpart of the Ramanujan–Petersson conjecture is Deligne's *weight-monodromy conjecture*. This goes beyond the Weil conjectures to predict that the étale cohomology of smooth projective varieties over p-adic fields has a remarkably elegant shape, even in the case of singular reduction.

In [12], building on [18, 39, 44] and [25], I follow Deligne's approach and complete the proof of the following result.

Theorem 4. *Let F be a CM field and let π be a regular algebraic, self-dual cuspidal automorphic representation of GL_n/F. Then π satisfies the generalised Ramanujan–Petersson conjecture.*

The global Langlands correspondence relates automorphic representations to global Galois representations. The direction from automorphic to Galois is best understood in the setting of Theorem 4, which is the so-called "self-dual case". This has been a milestone achievement in the field: it required the combined effort of many people over several decades, including Kottwitz, Clozel, Harris, Taylor, Shin, and Chenevier, and was built on fundamental contributions by Arthur, Laumon, Ngô and Waldspurger. In [12, 13], I also complete the proof that the associated Galois representations are compatible with local Langlands[3], by establishing new instances of the weight-monodromy conjecture for Shimura varieties.

More recently, in joint work with Allen, Calegari, Gee, Helm, Le Hung, Newton, Scholze, Taylor, and Thorne, I obtained an application to the Ramanujan–Petersson conjecture beyond the self-dual case. This is the first instance where this conjecture is not deduced from the Weil conjectures, but rather by an approximation of the very different strategy outlined by Langlands in [30].

Theorem 5 ([1]). *Let F be a CM field and π be a cuspidal automorphic representation of GL_2/F of parallel weight 2. Then π satisfies the generalised Ramanujan–Petersson conjecture.*

The condition on the weight means that π contributes to the Betti cohomology with constant coefficients of the relevant locally symmetric space, which is for example a Bianchi manifold. These locally symmetric spaces do not have an algebraic structure, so one cannot appeal directly to arithmetic geometry. We come back to discuss the strategy for the proof of Theorem 5 in Section 3.

1.3 The Sato–Tate conjecture

An *elliptic curve* is a smooth, projective curve of genus one together with a specified point. If F is a number field, an elliptic curve defined over F can be described as a plane curve, given by (the homogenisation of) a cubic equation of the form $y^2 = x^3 + ax + b$ with $a, b \in F$.

Such an elliptic curve E/F, if it does not have complex multiplication, is expected to satisfy the *Sato–Tate conjecture*. When $\mathfrak{p}$ is a prime of F over which E has good reduction, the number

$$\frac{1 + q_{\mathfrak{p}} - \#E(k(\mathfrak{p}))}{2\sqrt{q_{\mathfrak{p}}}}$$

[3] Local-global compatibility is a crucial property one expects from the Langlands correspondence, which generalises the compatibility between local and global class field theory.

(where $k(\mathfrak{p})$ denotes the residue field at $\mathfrak{p}$, of cardinality $q_\mathfrak{p}$) is contained in the interval $[-1, 1]$ by a result of Hasse; this is also a special case of Deligne's result on the Weil conjectures. The Sato–Tate conjecture, formulated in the 1960's, states that, as $\mathfrak{p}$ runs over all the primes of F over which E has good reduction, these numbers become equidistributed in $[-1, 1]$ with respect to the semicircle probability measure $\frac{2}{\pi}\sqrt{1-x^2}dx$.

Remark 6. The condition for an elliptic curve to have complex multiplication is very special, and in that case the probability distribution is different and well-understood. See [40] for a survey on Sato–Tate-type conjectures, which explains the expected distributions, and [26] for the more general conceptual framework that underlies this conjecture.

According to the Langlands reciprocity conjecture, any elliptic curve E/F is also expected to come from an automorphic representation of GL_2 over F. If this is the case, we say that E is *automorphic*. The precise relationship between elliptic curves and automorphic representations can be expressed as an equality of the two *L-functions* associated to them. *L*-functions are complex analytic functions that generalise the Riemann zeta function and that remember deep arithmetic information about the original objects.

For example, the *L*-functions of all elliptic curves defined over $\mathbb{Q}$ are known to come from modular forms, by work of Breuil–Conrad–Diamond–Taylor [9] building on [46] and [43]. The analogous result for elliptic curves defined over real quadratic fields was later proved by Freitas–Le Hung–Siksek [22]. The *L*-functions of elliptic curves over imaginary quadratic fields are expected to come from classes in the cohomology of Bianchi manifolds, but this case has historically been much more mysterious.

Soon after the Sato–Tate conjecture was formulated, Serre and Tate discovered that the correct distribution would follow from the expected analytic properties of the symmetric power *L*-functions of E. In turn, these analytic properties would follow if one knew the automorphy of E and all its symmetric powers. This argument is explained in [37] and uses Tauberian theorems in analytic number theory: the techniques are essentially those that led to the proof of the prime number theorem. In fact, to establish the correct distribution, it suffices to know that E and its symmetric powers are *potentially automorphic*: this means they become automorphic after base change to some Galois field extension F' of F.

The Sato–Tate conjecture for elliptic curves defined over totally real fields was proved in most cases by Clozel, Harris, Shepherd-Barron, and Taylor [19, 24, 42], and completed in work of Barnet-Lamb–Geraghty–Harris–Taylor around 2010 [4]. This relied on the potential automorphy of symmetric powers, which could be established in the self-dual setting using a generalisation of the Taylor–Wiles method. However, the method broke down for elliptic curves defined over imaginary quadratic fields or more general CM fields. In Section 3, we explain how to overcome the barrier to treating elliptic curves defined over CM fields and obtain the following result.

Theorem 7 ([1]). *Let F be a CM field and E/F be an elliptic curve that does not have complex multiplication. Then E is potentially automorphic and satisfies the Sato–Tate conjecture.*

Remark 8. Both Theorems 5 and 7 rely crucially on the vanishing theorem for Shimura varieties proved in [17], which is discussed in Section 2.

Remark 9. The beautiful work of Boxer–Calegari–Gee–Pilloni [7], completed at the same time as [1], proves the potential automorphy of elliptic curves in Theorem 7 independently, and they are even able to show the potential automorphy of abelian surfaces over totally real fields. Moreover, in the recent paper [2], Allen–Khare–Thorne establish actual automorphy of elliptic curves in certain cases (rather than potential automorphy). All of this is hopefully only the beginning of a fascinating story over CM fields!

2 Vanishing theorems for Shimura varieties with torsion coefficients

2.1 Shimura varieties

Recall that, if the locally symmetric spaces for a group $G/\mathbb{Q}$ have an algebraic structure, they in fact come from smooth, quasi-projective varieties X_Γ defined over number fields, which are called Shimura varieties.

The pair (G, X) must satisfy certain axioms in order for the corresponding locally symmetric spaces to come from Shimura varieties. The key point is for the symmetric space X to be a Hermitian symmetric domain (or a finite disjoint union thereof). There is a complete classification of groups G for which this holds. For example, the symplectic group Sp_{2n} and the unitary group $\mathrm{U}(n, n)$ give rise to Shimura varieties, which can be described in terms of moduli spaces of abelian varieties equipped with additional structures.

Remark 10. Some locally symmetric spaces that are not Shimura varieties can still be studied by relating them to Shimura varieties. For example, Bianchi manifolds can be realised in the boundary of certain compactifications of Shimura varieties attached to the unitary group $\mathrm{U}(2, 2)$. We come back to this in Section 3.

Recall also that the locally symmetric spaces for a group G give a way to access automorphic representations of G. More precisely, as the congruence subgroup $\Gamma \subset G(\mathbb{Z})$ varies, we have a tower of locally symmetric spaces. The symmetries of this tower induce correspondences on each individual space $\Gamma\backslash X$ called Hecke operators[4]. Keeping track of the various Hecke operators, we obtain an action of a commutative Hecke algebra $\mathbb{T}$ on the Betti cohomology $H^i(\Gamma\backslash X, \mathbb{C})$. The work of Matsushima, Franke and others shows that the systems of eigenvalues of $\mathbb{T}$ that occur in $H^i(\Gamma\backslash X, \mathbb{C})$ come from certain automorphic representations of G.

In addition to the Hecke symmetry, the cohomology of Shimura varieties also has a Galois symmetry, because Shimura varieties are defined over number fields. Because of these two kinds of symmetries, Shimura varieties give, in many cases, a geometric realisation of the global Langlands correspondence between automorphic and Galois representations.

One can ask a more precise question, about the range of degrees of cohomology to which any particular automorphic representation can contribute. Assume, for simplicity, that $X_\Gamma(\mathbb{C})$ is a compact Shimura variety. Then Borel–Wallach [6] show that, if π is an automorphic representation whose component at ∞ is a tempered representation of $G(\mathbb{R})$, then π can only contribute to $H^i(X_\Gamma(\mathbb{C}), \mathbb{C})$ in the middle degree $i = \dim_{\mathbb{C}} X_\Gamma$. This result, like the Ramanujan–Petersson conjecture, also fits within the framework of Arthur's conjectures [3].

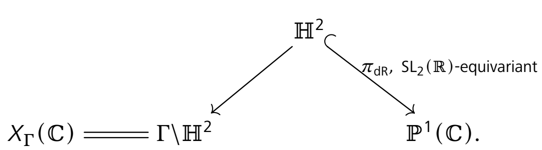
Question 11. The upshot of the Borel–Wallach result is that the cohomology of a Shimura variety X_Γ with $\mathbb{C}$-coefficients is somehow degenerate outside the middle degree. Can we extend this to torsion coefficients, such as $H^i(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)$?

More precise versions of this question are formulated as conjectures in [11] and [20]. These are motivated by the Calegari–Geraghty method, which is discussed in Section 3, and by the search for a mod ℓ analogue of Arthur's conjectures. In the next two subsections, we explain a new tool that can be used to compute $H^i(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)$ and discuss our results towards Question 11.

2.2 The Hodge–Tate period morphism

This morphism was introduced by Scholze in his breakthrough paper [35] and gives a completely new way to access the geometry and cohomology of Shimura varieties.

In the case of the modular curve, the Hodge–Tate period morphism is a p-adic analogue of the following complex picture, where the map on the right is the standard holomorphic embedding of the upper-half plane $\mathbb{H}^2$ into the Riemann sphere $\mathbb{P}^1(\mathbb{C})$:



This picture has the following moduli interpretation. First, X_Γ is a moduli space of elliptic curves equipped with some additional structures (determined by Γ). The upper-half plane $\mathbb{H}^2$ is the universal cover of $X_\Gamma(\mathbb{C}) = \Gamma\backslash\mathbb{H}^2$; it parametrises (positive) complex structures one can put on a two-dimensional real vector space. This amounts to parameterising Hodge structures of elliptic curves, i.e. direct sum decompositions:

$$\mathbb{C}^2 = H^1(E(\mathbb{C}), \mathbb{C}) \simeq H^0(E, \Omega_E^1) \oplus H^1(E, \mathcal{O}_E)$$

with $H^1(E, \mathcal{O}_E) = \overline{H^0(E, \Omega_E^1)}$. The morphism π_{dR} sends the Hodge decomposition to the associated Hodge filtration

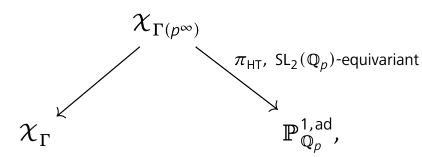
$$H^0(E, \Omega_E^1) \subset H^1(E(\mathbb{C}), \mathbb{C}) = \mathbb{C}^2.$$

This is an example of a *period morphism*. One can construct such a diagram for higher-dimensional Shimura varieties as well, and this has played an important role in studying automorphic forms on Shimura varieties from a geometric point of view.

The Hodge–Tate period morphism is based on the Hodge–Tate filtration on étale cohomology, tracing back to foundational work in p-adic Hodge theory by Tate and Faltings. Let p be a prime and let C be the p-adic completion of an algebraic closure of $\mathbb{Q}_p$, which will play a role analogous to that of $\mathbb{C}$ in what follows. If E/C is an elliptic curve, its étale cohomology admits a Hodge–Tate filtration:

$$0 \to H^1(E, \mathcal{O}_E) \to H^1_{et}(E, \mathbb{Z}_p) \otimes_{\mathbb{Z}_p} C \to H^0(E, \Omega_E^1)(-1) \to 0.$$

See Bhatt's article in [5] for an excellent survey on p-adic Hodge theory and more details on the Hodge–Tate filtration. Instead of viewing the curve X_Γ as a Riemann surface, we view it as an adic space $\mathcal{X}_\Gamma$, a kind of p-adic analytic space introduced by Huber. Then there exists a diagram



where $\mathcal{X}_\Gamma(p^\infty)$, which is roughly the inverse limit of modular curves $\mathcal{X}_{\Gamma(p^n)}$ with increasing level at p, is a perfectoid space. Over a point of $\mathcal{X}_{\Gamma(p^\infty)}$ corresponding to an elliptic curve E/C, we have a trivialisation of $H^1_{et}(E, \mathbb{Z}_p) \simeq \mathbb{Z}_p^2$. This point gets sent under π_{HT} to the line

$$H^1(E, \mathcal{O}_E) \subset H^1_{et}(E, \mathbb{Z}_p) \otimes_{\mathbb{Z}_p} C \simeq C^2.$$

For higher-dimensional Shimura varieties, the following result describes the geometry of the Hodge–Tate period morphism in detail. While the statement of Theorem 12 involves much non-trivial arithmetic geometry, it has applications to Theorems 16 and 17 below, whose statements are substantially more elementary.

[4] To discuss Hecke operators rigorously, we should use the adelic perspective on locally symmetric spaces and Shimura varieties. The resulting spaces would be disjoint unions of finitely many copies of $\Gamma\backslash X$. We ignore this subtlety here and later on in the text.

Theorem 12 ([35, 16]). *Let X_Γ be a Shimura variety of Hodge type associated to a connected reductive group G. Let μ denote the conjugacy class of Hodge cocharacters and let $\mathcal{F}\ell_{G,\mu} := G/P_\mu$ denote the corresponding flag variety, considered as an adic space over a p-adic completion of the reflex field.*

1. *There exists a unique perfectoid space $\mathcal{X}_{\Gamma(p^\infty)}$ which can be identified with the inverse limit of the adic spaces $(\mathcal{X}_{\Gamma(p^n)})_n$.*
2. *There exists a Hodge–Tate period morphism*
$$\pi_{\mathrm{HT}} : \mathcal{X}_{\Gamma(p^\infty)} \to \mathcal{F}\ell_{G,\mu},$$
which is $G(\mathbb{Q}_p)$-equivariant.
3. *There exists a Newton stratification*
$$\mathcal{F}\ell_{G,\mu} = \bigsqcup_{b \in B(G,\mu)} \mathcal{F}\ell^b_{G,\mu}$$
into locally closed strata.
4. *If $\mathcal{X}_\Gamma$ is compact and of PEL type, and $\bar{x}$ is a geometric point of the Newton stratum $\mathcal{F}\ell^b_{G,\mu}$, we identify the fiber $\pi^{-1}_{\mathrm{HT}}(\bar{x})$ with a "perfectoid" version of an Igusa variety Ig^b.*

Remark 13. The first two parts of Theorem 12 are due to Scholze[5] and play the lead role in his breakthrough construction of Galois representations for torsion in the cohomology of locally symmetric spaces. There are many surveys of this result; see for example [33] or [45]. For more details on the Hodge–Tate period morphism, see also the last article in [5].

Remark 14. Igusa varieties were introduced by Harris–Taylor as part of their proof of local Langlands for GL_n, and generalised by Mantovan. Rapoport–Zink spaces are local analogues of Shimura varieties, which provide a geometric realisation of the local Langlands correspondence. The computation of the fibers of π_{HT} suffices for applications to Theorems 16 and 17 below, but in [16], we go further and prove a conceptually cleaner version of Mantovan's product formula [32], which relates Shimura varieties, Igusa varieties and Rapoport–Zink spaces.

Remark 15. In [17] we extend part (4) of Theorem 12 to $\mathrm{U}(n,n)$-Shimura varieties, which are non-compact. We compute the fibers of π_{HT} for both the minimal and toroidal compactifications of these Shimura varieties, and relate them to partial minimal and toroidal compactifications of Igusa varieties.

2.3 Vanishing theorems

In order to address Question 11, we would like to compute the localisation $H^*(\mathcal{X}_\Gamma, \mathbb{F}_\ell)_\mathfrak{m}$, where the maximal ideal $\mathfrak{m} \subset \mathbb{T}$ is equivalent to a mod ℓ system of Hecke eigenvalues. Using the Hodge–Tate period morphism at an auxiliary prime $p \neq \ell$[6], we obtain an action of $\mathbb{T}$ on the complex of sheaves $R\pi_{\mathrm{HT}*}\mathbb{F}_\ell$ living over $\mathcal{F}\ell_{G,\mu}$, and we are reduced to understanding the localisation $(R\pi_{\mathrm{HT}*}\mathbb{F}_\ell)_\mathfrak{m}$. By the properties of π_{HT}, this behaves similarly to a perverse sheaf, which is the key to controlling the degrees in which $(R\pi_{\mathrm{HT}*}\mathbb{F}_\ell)_\mathfrak{m}$ can have non-zero cohomology. We make these ideas rigorous in [16, 17] for unitary Shimura varieties, under some mild technical assumptions.

Let $F = F^+ \cdot E$ be a CM field, with maximal totally real field $F^+ \neq \mathbb{Q}$ and E imaginary quadratic. Let G be a unitary group preserving a skew-Hermitian form on F^m. Assume that G is quasi-split at all finite places. Let $\mathfrak{m} \subset \mathbb{T}$ be a system of Hecke eigenvalues that occurs in $H^i(X_\Gamma, \mathbb{F}_\ell)$. Assume $\mathfrak{m}$ is *generic* at an auxiliary prime $p \neq \ell$[7]. This condition guarantees that all lifts of $\mathfrak{m}$ to characteristic 0 are as simple as possible at p, from a representation-theoretic point of view: they are generic principal series representations of $G(\mathbb{Q}_p)$.

Theorem 16 ([16]). *If $\mathcal{X}_\Gamma$ is compact and $\mathfrak{m}$ is generic, then $H^i(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)_\mathfrak{m}$ is concentrated in the middle degree $i = \dim_\mathbb{C} X_\Gamma$.*

In the non-compact case, genericity, which is a local condition at an auxiliary prime $p \neq \ell$, is not enough. We also need a global condition to control the boundary of the Shimura variety. To formulate the global condition, we consider the semi-simple Galois representation $\bar{\rho}_\mathfrak{m}$ associated to the system of eigenvalues $\mathfrak{m}$ by [35]; the existence of $\bar{\rho}_\mathfrak{m}$ is an instance of the global Langlands correspondence in the torsion setting. We want to assume that $\bar{\rho}_\mathfrak{m}$ is not too degenerate; this amounts to bounding the number of its absolutely irreducible factors.

Theorem 17 ([17]). *If X_Γ is a $\mathrm{U}(n,n)$-Shimura variety (so m is even and G is quasi-split at the infinite places as well), $\mathfrak{m}$ is generic, and $\bar{\rho}_\mathfrak{m}$ has at most two absolutely irreducible factors, then:*

1. *$H^i_c(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)_\mathfrak{m}$ is concentrated in degrees $i \leq \dim_\mathbb{C} X_\Gamma$, and*
2. *$H^i(X_\Gamma(\mathbb{C}), \mathbb{F}_\ell)_\mathfrak{m}$ is concentrated in degrees $i \geq \dim_\mathbb{C} X_\Gamma$.*

Remark 18. There are previous results in this direction, due to Dimitrov, Shin, Emerton–Gee, and especially Lan–Suh [28, 29]. Compared to previous work, our result is sharper and better adapted to applications. There is also intriguing ongoing work of Boyer [8], which proves a stronger result in the special case of Harris–Taylor Shimura varieties: he goes beyond genericity and investigates the distribution of non-generic systems of Hecke eigenvalues.

[5] Up to the precise identification of the target of the Hodge–Tate period morphism as the flag variety $\mathcal{F}\ell_{G,\mu}$ in all cases, which is done in [16].
[6] Here, we assume that the Hecke operators in $\mathbb{T}$ are all supported at primes different from p.
[7] See [17, Theorem 1.1] for the precise condition, which is technical, but explicit. This condition should be thought of as a mod ℓ analogue of temperedness.

Remark 19. The idea of the proof in the compact case is the following: start with a top-dimensional Newton stratum $\mathcal{F}\ell^b_{G,\mu} \subset \mathcal{F}\ell_{G,\mu}$ in the support of $(R\pi_{\mathrm{HT},*}\mathbb{F}_\ell)_\mathfrak{m}$. Since the complex $(R\pi_{\mathrm{HT},*}\mathbb{F}_\ell)_\mathfrak{m}$ behaves like a perverse sheaf, its restriction to $\mathcal{F}\ell^b_{G,\mu}$ is concentrated in one degree. Therefore, $(R\pi_{\mathrm{HT},*}\mathbb{Q}_\ell)_\mathfrak{m}$ is also concentrated in one degree over $\mathcal{F}\ell^b_{G,\mu}$. On the other hand, we can compute the alternating sum of cohomology groups of Ig^b with $\mathbb{Q}_\ell$-coefficients, using the trace formula and work of Shin [38]. In the end, the genericity condition is contradicted unless b corresponds to the zero-dimensional ordinary stratum. The upshot is that $(R\pi_{\mathrm{HT},*}\mathbb{F}_\ell)_\mathfrak{m}$ is concentrated in one degree over a zero-dimensional stratum!

Remark 20. In parallel to Question 11, one can also study the cohomology of locally symmetric spaces with torsion coefficients and with increasing level at p. The resulting structure is called *completed cohomology* and was introduced by Emerton as a general framework for studying congruences modulo p^k between automorphic forms. Motivated by heuristics coming from the p-adic Langlands programme, Calegari–Emerton [10] formulated a vanishing conjecture for completed cohomology. For most Shimura varieties, the Calegari–Emerton conjecture is now a theorem due to Scholze and Hansen–Johansson.

In [14, 15], we prove a vanishing result for the compactly supported cohomology of Shimura varieties of Hodge type with unipotent level at p. The only assumption is that the group G giving rise to the Shimura variety is split over $\mathbb{Q}_p$. This result is stronger than what Calegari–Emerton conjectured, and it also points towards analogues of Theorems 16 and 17 for $\ell = p$, with *generic* replaced by *ordinary* in the sense of Hida.

3 Potential automorphy over CM fields

Theorem 5 on the Ramanujan–Petersson conjecture and Theorem 7 on the Sato–Tate conjecture would follow if we knew that all the symmetric powers of the associated Galois representations were automorphic, or even just potentially automorphic. The original method developed by Taylor–Wiles is a powerful technique for proving automorphy, but it is restricted to settings where a certain numerical criterion holds: these are roughly the settings where the objects on the automorphic side arise from the middle degree cohomology of a Shimura variety.

When F is a number field, the locally symmetric spaces for GL_n/F, such as the Bianchi manifolds discussed in Example 3, do not have an algebraic structure (outside very special cases). Calegari–Geraghty [11] proposed an extension of the Taylor–Wiles method to general number fields F, conjectural on a precise understanding of the cohomology of locally symmetric spaces for GL_n/F. Part of their insight was to realise the central role played by torsion classes in the cohomology of these locally symmetric spaces, which should be thought of as modulo p^k versions of automorphic forms and treated on equal footing with their characteristic 0 counterparts. Another part of their insight was to reinterpret the failure of the Taylor–Wiles numerical criterion in terms of certain non-negative integers q_0, l_0 seen on the automorphic side.

The Calegari–Geraghty method gives an automorphy lifting result for GL_n/F as long as the following prerequisites are in place:

1. The construction of Galois representations associated to classes in the cohomology with $\mathbb{Z}_p$ coefficients of the locally symmetric spaces for GL_n/F.
2. Local-global compatibility for these Galois representations at all primes of F, including at primes above p.
3. A vanishing conjecture for the cohomology with $\mathbb{Z}_p$ coefficients outside the range of degrees $[q_0, q_0 + l_0]$, under an appropriate non-degeneracy condition.

Remark 21. For Shimura varieties, the third problem is closely related to Theorems 16 and 17, since in that case q_0 is the middle degree of cohomology and $l_0 = 0$. For 3-dimensional Bianchi manifolds, the third problem says that the non-degenerate part of cohomology is concentrated in degrees 1 and 2; this can be checked by hand. For general locally symmetric spaces that do not have an algebraic structure, this problem most likely lies deeper than the first two.

When F is a CM field, the first problem was solved by Scholze in [35], strengthening previous results of Harris–Lan–Taylor–Thorne [23] for characteristic 0 coefficients. After completing [16], it became clear to Scholze and me that a non-compact version of Theorem 16 would give a strategy to attack the second (rather than the third!) problem over CM fields. In joint work with Scholze, I set out to prove Theorem 17 and, in November 2016, I co-organised with Taylor an "emerging topics" working group at the IAS, whose goal was to explore this strategy and its consequences. The working group was a resounding success and it led to the paper [1], where we implement the Calegari–Geraghty method in arbitrary dimension for the first time and obtain as consequences Theorems 5 and 7.

The solution to the first problem above, i.e., the construction of Galois representations, is much more subtle than in the self-dual case, because one cannot directly use the étale cohomology of Shimura varieties. Instead, the starting point for both [23] and [35] is to realise the locally symmetric spaces for GL_n/F in the boundary of the *Borel–Serre compactification* of $\mathrm{U}(n,n)$-Shimura varieties. The Borel–Serre compactification is a real manifold with corners, which is homotopy equivalent to the original $\mathrm{U}(n,n)$-Shimura variety. In the torsion setting, Scholze constructs the desired Galois representations by congruences, using the Hodge–Tate period morphism for the $\mathrm{U}(n,n)$-Shimura variety. This increases the level at primes of F dividing p, and makes the second problem, local-global compatibility, particularly tricky at these primes.

In [1], we begin to solve the second problem, by establishing the first instances of local-global compatibility at primes of F dividing p. We need a delicate argument to understand the boundary of the Borel–Serre compactification, which combines algebraic topology and modular representation theory. In addition, Theorem 17 is the crucial new ingredient: in the middle degree, it implies that classes from the boundary lift to the cohomology of a $\mathrm{U}(n,n)$-Shimura variety with $\mathbb{Q}_p$-coefficients, while remembering the level and weight at primes of F dividing p.

The proofs of Theorems 5 and 7 use the Calegari–Geraghty method, together with solutions to the first two problems discussed above. The third problem was not solved with $\mathbb{Z}_p$ coefficients. By an insight of Khare–Thorne [27], this problem could be replaced by its $\mathbb{Q}_p$ coefficient analogue in certain settings. One of the main challenges in [1] was to make this insight compatible with other techniques in automorphy lifting, which rely on reduction modulo p. We resolve this challenge by considering reduction modulo p from a derived perspective. Outside low-dimensional cases, such as Bianchi manifolds, or Shimura varieties, the third problem remains open for $\mathbb{Z}_p$ coefficients.

Acknowledgements. This article was written in relation to my being awarded one of the 2020 Prizes of the European Mathematical Society. I wish to dedicate this article to my father, Cornel Caraiani (1954–2020), who inspired my love of mathematics.

I have been lucky to have many wonderful mentors and collaborators and I am grateful to all of them for the mathematics they have taught me. In addition, I especially want to thank Matthew Emerton, Toby Gee, Sophie Morel, James Newton, Peter Scholze, and Richard Taylor for generously sharing their ideas with me over the years, and for their substantial moral and professional support.

I am also grateful to Toby Gee, James Newton, Steven Sivek, and Matteo Tamiozzo for comments on an earlier version of this article.

References

[1] P. B. Allen, F. Calegari, A. Caraiani, T. Gee, D. Helm, B. V. Le Hung, J. Newton, P. Scholze, R. Taylor, and J. A. Thorne. Potential automorphy over CM fields. arXiv:1812.09999 (December 2018)

[2] P. B. B. Allen, C. Khare, and J. A. Thorne. Modularity of $\mathrm{GL}_2(\mathbb{F}_p)$-representations over CM fields. arXiv:1910.12986 (October 2019)

[3] J. Arthur, Unipotent automorphic representations: Conjectures. 171–172, 13–71 (1989)

[4] T. Barnet-Lamb, D. Geraghty, M. Harris and R. Taylor, A family of Calabi-Yau varieties and potential automorphy II. *Publ. Res. Inst. Math. Sci.* **47**, 29–98 (2011)

[5] B. Bhatt, A. Caraiani, K. S. Kedlaya and J. Weinstein, *Perfectoid spaces*. Mathematical Surveys and Monographs 242, American Mathematical Society, Providence, RI (2019)

[6] A. Borel and N. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*. Second ed., Mathematical Surveys and Monographs 67, American Mathematical Society, Providence, RI (2000)

[7] G. Boxer, F. Calegari, T. Gee, and V. Pilloni. Abelian Surfaces over totally real fields are Potentially Modular. arXiv:1812.09269 (December 2018)

[8] P. Boyer, Sur la torsion dans la cohomologie des variétés de Shimura de Kottwitz–Harris–Taylor. *J. Inst. Math. Jussieu* **18**, 499–517 (2019)

[9] C. Breuil, B. Conrad, F. Diamond and R. Taylor, On the modularity of elliptic curves over $\mathbf{Q}$: Wild 3-adic exercises. *J. Amer. Math. Soc.* **14**, 843–939 (2001)

[10] F. Calegari and M. Emerton, Completed cohomology – A survey. In *Non-abelian fundamental groups and Iwasawa theory*, London Math. Soc. Lecture Note Ser. 393, Cambridge Univ. Press, Cambridge, 239–257 (2012)

[11] F. Calegari and D. Geraghty, Modularity lifting beyond the Taylor–Wiles method. *Invent. Math.* **211**, 297–433 (2018)

[12] A. Caraiani, Local-global compatibility and the action of monodromy on nearby cycles. *Duke Math. J.* **161**, 2311–2413 (2012)

[13] A. Caraiani, Monodromy and local-global compatibility for $l = p$. *Algebra Number Theory* **8**, 1597–1646 (2014)

[14] A. Caraiani, D. R. Gulotta, C.-Y. Hsu, C. Johansson, L. Mocz, E. Reinecke and S.-C. Shih, Shimura varieties at level $\Gamma_1(p^\infty)$ and Galois representations. *Compos. Math.* **156**, 1152–1230 (2020)

[15] A. Caraiani, D. R. Gulotta, and C. Johansson. Vanishing theorems for Shimura varieties at unipotent level. arXiv:1910.09214 (October 2019)

[16] A. Caraiani and P. Scholze, On the generic part of the cohomology of compact unitary Shimura varieties. *Ann. of Math. (2)* **186**, 649–766 (2017)

[17] A. Caraiani and P. Scholze, On the generic part of the cohomology of compact unitary Shimura varieties. arXiv:1909.01898 (September 2019)

[18] L. Clozel, Purity reigns supreme. *Int. Math. Res. Not. IMRN* 328–346 (2013)

[19] L. Clozel, M. Harris and R. Taylor, Automorphy for some l-adic lifts of automorphic mod l Galois representations. *Publ. Math. Inst. Hautes Études Sci.* 1–181 (2008)

[20] M. Emerton, Completed cohomology and the p-adic Langlands program. In *Proceedings of the International Congress of Mathematicians – Seoul 2014. Vol. II*, Kyung Moon Sa, Seoul, 319–342 (2014)

[21] M. Emerton. Langlands reciprocity: L-functions, automorphic forms, and Diophantine equations. To appear in *The Genesis of the Langlands program* (2020)

[22] N. Freitas, B. V. Le Hung and S. Siksek, Elliptic curves over real quadratic fields are modular. *Invent. Math.* **201**, 159–206 (2015)

[23] M. Harris, K.-W. Lan, R. Taylor and J. Thorne, On the rigid cohomology of certain Shimura varieties. *Res. Math. Sci.* **3**, Paper No. 37, 308 (2016)

[24] M. Harris, N. Shepherd-Barron and R. Taylor, A family of Calabi-Yau varieties and potential automorphy. *Ann. of Math. (2)* **171**, 779–813 (2010)

[25] M. Harris and R. Taylor, *The geometry and cohomology of some simple Shimura varieties*. Annals of Mathematics Studies 151, Princeton University Press, Princeton, NJ (2001)

[26] N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*. American Mathematical Society Colloquium Publications 45, American Mathematical Society, Providence, RI (1999)

[27] C. B. Khare and J. A. Thorne, Potential automorphy and the Leopoldt conjecture. *Amer. J. Math.* **139**, 1205–1273 (2017)

[28] K.-W. Lan and J. Suh, Vanishing theorems for torsion automorphic sheaves on compact PEL-type Shimura varieties. *Duke Math. J.* **161**, 1113–1170 (2012)

[29] K.-W. Lan and J. Suh, Vanishing theorems for torsion automorphic sheaves on general PEL-type Shimura varieties. *Adv. Math.* **242**, 228–286 (2013)

[30] R. P. Langlands, Problems in the theory of automorphic forms. In *Lectures in modern analysis and applications, III*, 18–61. Lecture Notes in Math., Vol. 170 (1970)

[31] W.-C. W. Li, The Ramanujan conjecture and its applications. *Philos. Trans. Roy. Soc. A* **378**, 20180441, 14 (2020)

[32] E. Mantovan, On the cohomology of certain PEL-type Shimura varieties. *Duke Math. J.* **129**, 573–610 (2005)

[33] S. Morel, Construction de représentations galoisiennes de torsion [d'après Peter Scholze]. *Astérisque* Exp. No. 1102, 449–473 (2016)

[34] P. Sarnak, Notes on the generalized Ramanujan conjectures. In *Harmonic analysis, the trace formula, and Shimura varieties*, Clay Math. Proc. 4, Amer. Math. Soc., Providence, RI, 659–685 (2005)

[35] P. Scholze, On torsion in the cohomology of locally symmetric varieties. *Ann. of Math. (2)* **182**, 945–1066 (2015)

[36] P. Scholze, *p*-adic geometry. In *Proceedings of the International Congress of Mathematicians – Rio de Janeiro 2018. Vol. I. Plenary lectures*, World Sci. Publ., Hackensack, NJ, 899–933 (2018)

[37] J.-P. Serre, *Abelian l-adic representations and elliptic curves*. McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute, W. A. Benjamin, Inc., New York-Amsterdam (1968)

[38] S. W. Shin, A stable trace formula for Igusa varieties. *J. Inst. Math. Jussieu* **9**, 847–895 (2010)

[39] S. W. Shin, Galois representations arising from some compact Shimura varieties. *Ann. of Math. (2)* **173**, 1645–1741 (2011)

[40] A. V. Sutherland, Sato–Tate distributions. In *Analytic methods in arithmetic geometry*, Contemp. Math. 740, Amer. Math. Soc., Providence, RI, 197–248 (2019)

[41] R. Taylor, Galois representations. *Ann. Fac. Sci. Toulouse Math. (6)* **13**, 73–119 (2004)

[42] R. Taylor, Automorphy for some *l*-adic lifts of automorphic mod *l* Galois representations. II. *Publ. Math. Inst. Hautes Études Sci.* 183–239 (2008)

[43] R. Taylor and A. Wiles, Ring-theoretic properties of certain Hecke algebras. *Ann. of Math. (2)* **141**, 553–572 (1995)

[44] R. Taylor and T. Yoshida, Compatibility of local and global Langlands correspondences. *J. Amer. Math. Soc.* **20**, 467–493 (2007)

[45] J. Weinstein, Reciprocity laws and Galois representations: Recent breakthroughs. *Bull. Amer. Math. Soc. (N.S.)* **53**, 1–39 (2016)

[46] A. Wiles, Modular elliptic curves and Fermat's last theorem. *Ann. of Math. (2)* **141**, 443–551 (1995)

Ana Caraiani is a Royal Society University Research Fellow and Reader in Pure Mathematics at Imperial College London.

a.caraiani@imperial.ac.uk

Geometry and dynamics on Riemann and K3 surfaces

Simion Filip

Surfaces are some of the simplest yet geometrically rich manifolds. Geometric structures on surfaces illuminate their topology and are useful for studying dynamical systems on surfaces. We illustrate below how some of these concepts blend together, and relate them to algebraic geometry.

1 Geometry of surfaces

In this section we give a brief overview of some geometric facts regarding Riemann surfaces and K3 surfaces. Although both are called "surfaces", Riemann surfaces are examples of algebraic curves, while K3 surfaces are genuine algebraic surfaces. This means that considering the complex points, Riemann surfaces are complex 1-dimensional while K3 surfaces are complex 2-dimensional. We will give some explicit examples and then describe geometric structures that live on these surfaces. In the Riemann case we are concerned with flat geometry (with singularities) while in the K3 case we consider Ricci-flat Kähler metrics. Moduli spaces of these geometric structures play an important role in the results of Section 2 below.

Riemann surfaces. One can describe a compact Riemann surface by giving the algebraic equations that cut it out in some ambient space. For example, consider

$$X: y^2 = x(x^5 - 1), \quad \Omega = \frac{dx}{y}. \tag{1}$$

The locus $X(\mathbb{C})$ of points satisfying this equation in $\mathbb{C}^2$ is a real 2-dimensional surface of genus 2 (with two points at infinity added to X). The 1-form Ω from (1) is the unique (up to scale) holomorphic 1-form vanishing at the two points at infinity.

Flat geometry. There is an alternative way to describe the pair (X, Ω). Take the regular decagon in the plane and glue its opposite and parallel edges to form a surface of genus 2, with two marked points given by the vertices. If we identify the plane with $\mathbb{C}$, then the 1-form $\tilde{\Omega} = dz$ will be invariant under the translations used to glue opposite edges and will descend to a 1-form Ω on the new surface. This construction gives back the same pair (X, Ω) as described in (1), although this is by no means obvious. For more on the algebraic curve in (1), see [26, §5]. (A question for experts: what is the area of the decagon under this identification?)

This construction is quite general: starting from a pair (X, Ω) consisting of a compact Riemann surface and a holomorphic 1-form, one can associate to it a polygon in the plane by cutting the surface X and mapping it to the plane in such a way that in local charts the 1-form Ω becomes dz. Equivalently, one can give charts to $\mathbb{C}$ near a point $p_0 \in X$ by $p \mapsto \int_{p_0}^{p} \Omega$, and the transition maps between charts are translations in $\mathbb{C}$. Conversely, given a polygon in the plane (possibly disconnected), with side identifications given by translations, one can reconstruct a Riemann surface with a holomorphic 1-form using the converse to the above recipe.

Action of $\mathbf{GL}_2(\mathbb{R})$. A polygon is determined by its sides, which are vectors in $\mathbb{R}^2 \xrightarrow{\sim} \mathbb{C}$. The group $\mathbf{GL}_2(\mathbb{R})$ acts on polygons, keeping parallel sides parallel, so if we have a polygonal description of (X, Ω) then we obtain a new pair $g \cdot (X, \Omega) = (X', \Omega')$. One can also express the action intrinsically on the surface, by letting a matrix act on the real and imaginary parts of Ω, viewed as differential 1-forms on X:

$$\begin{bmatrix} \operatorname{Re}\Omega' \\ \operatorname{Im}\Omega' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} \operatorname{Re}\Omega \\ \operatorname{Im}\Omega \end{bmatrix}.$$

Note that the holomorphic structure on X' is usually different from the one on X. Furthermore, even if explicit algebraic equations are given for X, it is typically not possible to describe the equations cutting out X'. The sides of the polygons describing the surfaces are computed by taking integrals of Ω on paths in X, and the passage from algebraic equations to integrals and back is by no means explicit.

Moduli spaces of Riemann surfaces. Because algebraic equations have finitely many coefficients, one can consider full parameter spaces, or *moduli spaces*, of algebraic manifolds defined by the same type of equations. For Riemann surfaces we will be interested in the space $\mathcal{H}(\varkappa)$ of pairs (X, Ω) where X is a compact

Riemann surface and Ω is a holomorphic 1-form with $\varkappa$ describing the multiplicities of the zeros of Ω. The process described above of obtaining a new pair (X, Ω) using a real matrix gives an action of $\mathbf{GL}_2(\mathbb{R})$ on $\mathcal{H}(\varkappa)$. This action, however, is not via polynomial or even holomorphic automorphisms. In Section 2 we will see, however, that there are some relations between the $\mathbf{GL}_2(\mathbb{R})$-action and the algebraic equations defining Riemann surfaces.

For more on the $\mathbf{GL}_2(\mathbb{R})$-action on $\mathcal{H}(\varkappa)$, see the surveys of Zorich [36] for an introduction as well as numerous motivations and applications, as well as the more recent surveys of Forni–Matheus [20] and Wright [33].

K3 surfaces. We now switch gears and consider *algebraic surfaces*, such as those given by the equation

$$X = \{(1+x^2)(1+y^2)(1+z^2) - 16xyz = 4\}, \tag{2}$$

$$\Omega = \frac{dx \wedge dy}{z(1+x^2)(1+y^2) - 8xy}. \tag{3}$$

The 2-form Ω is nowhere-vanishing on X and is computed via a residue construction (take the residue of $\frac{dx \wedge dy \wedge dz}{F}$ along $F = 0$). It is interesting to consider both the complex and the real solutions of this equation, denoted $X(\mathbb{C})$ and $X(\mathbb{R})$ respectively. See Figure 2 for an example of real solutions. Algebraic curves (i.e., Riemann surfaces), such as those described in (1), have only finite automorphism groups as soon as the genus is at least two, but algebraic surfaces such as the one in (2) have dynamically interesting automorphisms such as

$$(x, y, z) \xrightarrow{\sigma_x} \left(\frac{16yz}{(1+y^2)(1+z^2)} - x, y, z\right) \tag{4}$$

as well as their analogues σ_y, σ_z in which the roles of the coordinates are permuted. The formula for the automorphism σ_x is obtained by "freezing" the y and z variables, viewing (2) as a quadratic equation for x, and exchanging the two solutions of the quadratic. In particular applying σ_x twice gives back the identity transformation: $\sigma_x \circ \sigma_x = \mathbf{1}_X$.

Kähler geometry. The complex solutions of algebraic equations such as the ones above yield projective algebraic manifolds. These admit Kähler metrics, special kinds of Riemannian metrics adapted to the complex structure. For Kähler metrics, the Riemannian volume of an algebraic submanifold is determined by its homology class alone.

Specializing further, if the algebraic manifold admits an algebraic volume form, such as Ω from (3), then Yau's solution of the Calabi conjecture [34] gives canonical Kähler metrics whose Ricci curvature vanishes. To construct such metrics, one needs to solve a nonlinear PDE of Monge–Ampère type and there is no "hands-on" description of such metrics as in Figure 1.

Moduli spaces of K3 surfaces. Moving to algebraic manifolds of higher dimensions, additional data needs to be specified in order to have well-behaved moduli spaces. For us, the most relevant will be the space $\mathcal{KE}$ of Ricci-flat metrics on the manifold underlying a complex K3 surface. The abbreviation $\mathcal{KE}$ is for Kähler–Einstein, since Ricci-flat metrics satisfy the Einstein equation $\mathrm{Ric}_{ij} = \lambda g_{ij}$ with $\lambda = 0$. These moduli spaces play an essential role in the study of K3 surfaces and have the remarkable feature that they are (essentially) homogeneous spaces for appropriate Lie groups. For an introduction to the geometry of K3 surfaces, see the collection of notes [2] and the more recent monograph of Huybrechts [22].

2 Dynamics on moduli spaces

This section describes results on the dynamics of group actions in the moduli space of Riemann and K3 surfaces equipped with appropriate flat, or Ricci-flat, metrics.

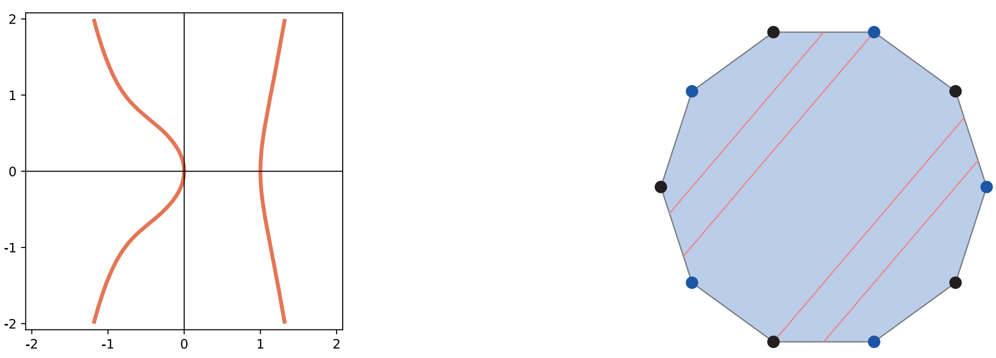


Figure 1. Left: Real solutions of $y^2 = x(x^5 - 1)$. Right: Decagon with opposite sides identified, and a straight line on the surface connecting the two singularities

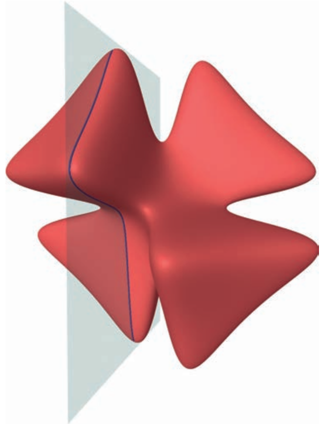


Figure 2. The real solutions of (2), intersected with the plane $x = 1$ to obtain an elliptic curve

Dynamics on moduli spaces of flat surfaces. The action of the group $\mathbf{GL}_2(\mathbb{R})$ on the moduli space $\mathcal{H}(\kappa)$ of Riemann surfaces with a holomorphic 1-form satisfies rigidity properties akin to those for unipotent flows in homogeneous dynamics developed by Ratner [30, 31], Margulis [24] and many others. The following results were established by Eskin, Mirzakhani, and Mohammadi:

Theorem 1 ([10, 11]). *For any pair $(X,\Omega) \in \mathcal{H}(\kappa)$, the orbit closure $\mathcal{M} := \overline{\mathbf{GL}_2(\mathbb{R}) \cdot (X,\Omega)}$ is a submanifold of $\mathcal{H}(\kappa)$, described in local coordinates by linear relations among the sides of the polygons used to parametrize surfaces. Furthermore,[1] any $\mathbf{SL}_2(\mathbb{R})$-invariant ergodic probability measure must be Lebesgue supported on such a manifold.*

Going back to the algebraic description of Riemann surfaces, we saw that except in special symmetric situations, it is not possible in general to relate the algebraic equations to the polygonal description of the surface. In the case of $\mathbf{GL}_2(\mathbb{R})$-orbit closures, it is possible to give an alternative, purely algebraic description of their geometry. Specifically, recall that the Jacobian $\mathrm{Jac}(X)$ associated to a genus g Riemann surface X is the complex torus defined as $H^0(X;K_X)^\vee/H_1(X;\mathbb{Z})$, where $H^0(X;K_X)$ denotes the complex g-dimensional space of holomorphic 1-forms on X, $\vee$ denotes the dual, and the first homology group $H_1(X;\mathbb{Z})$ embeds in the dual by integration along cycles. Alternatively, the Jacobian is the moduli space of holomorphic degree 0 line bundles on X and this description provides a link between the algebraic and analytic structures on a Riemann surface. Although the automorphism group of a genus $g \geq 2$ Riemann surface is finite, the endomorphism group of its Jacobian can be much larger (real or complex multiplication give examples of such symmetries).

Theorem 2 ([14, 13]). *Orbit closures $\mathcal{M}$ as in Theorem 1 parametrize Riemann surfaces whose Jacobians have additional endomorphisms a specific kind (such as real multiplication). Furthermore, specific combinations of the zeros of the distinguished 1-form yield torsion points on the Jacobian.*

These conditions characterize $\mathcal{M}$ as a locus inside $\mathcal{H}(\kappa)$.

Additional finiteness results for orbit closures are established in [6], jointly with Eskin and Wright.

The relation between the $\mathbf{GL}_2(\mathbb{R})$-action and real multiplication on Jacobians was discovered by McMullen [25], who also established most of the above-mentioned results in the case of genus 2 Riemann surfaces [27]. Möller introduced the tools of Hodge theory to the subject [29, 28] which were used to connect the algebraic and combinatorial descriptions of holomorphic 1-forms on Riemann surfaces.

Billiards. Fix a polygon and consider the dynamical system consisting of a billiard ball bouncing off the sides in the customary way, with the angle of incidence equal to the angle of reflection. By studying billiards in regular n-gons, Veech discovered the first instances of nontrivial orbit closures for the $\mathbf{GL}_2(\mathbb{R})$-action and established along the way:

Theorem 3 ([32, Thm. 1.5]). *For a regular n-gon, the number of closed billiard trajectories of length at most L grows like c_nL^2 for a constant c_n.*

[1] The switch from $\mathbf{GL}_2$ to $\mathbf{SL}_2$ is done to exclude the scaling action.

This is in analogy with the Gauss circle problem of counting lattice points in the plane, which corresponds to playing billiards on a square. The deeper study of the dynamics of billiards on surfaces, and polygons with rational angles, ties in with the study of the $\mathbf{GL}_2(\mathbb{R})$-action on the moduli space $\mathcal{H}(\kappa)$, and this is the key to Veech's result and many others.

An analogue for K3 surfaces. Billiard trajectories are locally given by straight lines. Besides the characterization of straight lines as giving the shortest path between points, they have the following alternative description. Take the 1-form $\Omega = dz = dx + \sqrt{-1}dy$ in the plane. A straight line is a curve γ such that $l(\gamma) = \left|\int_\gamma \Omega\right|$ where $l(\gamma)$ is the Euclidean length of γ. Note that for an arbitrary curve γ we have the inequality

$$l(\gamma) \geq \left|\int_\gamma \Omega\right|$$

which, in differential-geometric language, says that the 1-form Ω calibrates the straight lines.

This last point of view generalizes to K3 surfaces, where the analogue of closed billiard trajectories are special Lagrangian tori. These are real 2-dimensional tori inside a K3 surface with a Ricci-flat Kähler metric which, among many other properties, minimize volume in their homology class.

Theorem 4 ([17, Thm. C]). *Under appropriate assumptions on the Ricci-flat metric on a K3 surface, the number of such special Lagrangian tori, of volume bounded by V, is asymptotic to cV^{20}, for an explicit constant $c > 0$.*

It is possible to make the above counting effective and give an error term of order $V^{20-\varepsilon}$, for $\varepsilon > 0$, which was estimated effectively by Bergeron–Matheus in the appendix to [17]. Analogously to the counting result for Riemann surfaces, this one is established by studying the dynamics in the full moduli space $\mathcal{KE}$ of Ricci-flat metrics. Although we are asking a question about a specific one, it proves useful to study the space of all possible metrics. The idea of using dynamics on homogeneous spaces for counting results goes back to Eskin and McMullen [9].

3 Dynamics on K3 surfaces

In this section we describe some results on individual automorphisms of K3 surfaces. Again, a key role in the proofs is played by Ricci-flat metrics and their moduli space on a fixed K3 surface.

Entropy. Suppose for a moment that (X, d) is a compact metric space and $f : X \to X$ is a continuous map. Define a new distance function by $d_n(x, y) := \max_{0 \leq i \leq n} d(f^i(x), f^i(y))$, so two points are at d_n-distance at least ε if along their f-orbits, they have separated at some time at distance ε. Let now $S(d_n, \varepsilon)$ be the maximal number of ε-separated points in X, i.e., any two are at d_n-distance at least ε. This is the number of essentially distinct trajectories, up to time n, when observing the system with accuracy ε. The topological entropy $h^{top}(f)$ is the exponential growth rate in n of $S(d_n, \varepsilon)$:

$$h^{top}(f) := \lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n} \log S(d_n, \varepsilon).$$

There is also an associated notion of measure-theoretic entropy. Recall that an f-invariant measure μ must satisfy $\mu(f^{-1}(A)) = \mu(A)$ for any measurable subset $A \subset X$. To define the entropy of μ, partition X into disjoint measurable sets $X_1, \ldots, X_k$. Then an orbit of a point gives rise to a sequence of elements that it visits, encoded as a sequence in $\{1, \ldots, k\}^{\mathbb{N}}$. The number of such distinct sequences, weighted appropriately by μ, grows exponentially, and the exponential growth rate is called the entropy (after taking a supremum over all finite partitions of X).

Yomdin and Gromov theorems. Suppose now that X is a smooth manifold and f is a smooth diffeomorphism. Then the pullback f^* acts on the cohomology groups $H^\bullet(X; \mathbb{R})$ and we consider its spectral radius $\rho(f)$ (viewed as a linear transformation). Settling a conjecture of Shub, Yomdin proved the following:

Theorem 5 ([35]). *The topological entropy of f satisfies*

$$h^{top}(f) \geq \log \rho(f).$$

Thus topological complexity implies dynamical complexity. When X is a complex manifold admitting a Kähler metric, and f is a holomorphic automorphism, Gromov [21] established the reverse upper bound

$$h^{top}(f) \leq \rho(f) \quad \text{and so} \quad h^{top}(f) = \rho(f).$$

Gromov's proof is based on the special feature of Kähler metrics that the Riemannian volume of complex submanifolds is determined by their homology class (they are calibrated submanifolds, just like the special Lagrangians in Theorem 4).

Measures on K3 surfaces. Suppose now that X is a K3 surface and f is an automorphism of positive topological entropy. The surface in (2) works, and a composition of automorphisms like the one in (4), one for each coordinate, gives an example. Cantat [4] showed that there exists a unique f-invariant measure μ which maximizes entropy, i.e., $h_\mu(f) = \log \rho(f)$. But the holomorphic 2-form Ω induces another measure $\Omega \wedge \overline{\Omega}$ on X, which is canonical and invariant under the dynamics. It is then natural to ask: what is the relationship between the two measures? This was answered by Cantat–Dupont [5] and later, using different techniques, in [19]:

Theorem 6. *With notation as above, suppose that the measure of maximal entropy μ is absolutely continuous with respect to Lebesgue measure on X. Then X is a "Kummer surface", i.e., obtained from a complex torus $A = \mathbb{C}^2/\Lambda$ by a quotient $A \to A/\pm 1$ and desingularization, and the automorphism f comes by the same construction from a linear automorphism of the torus.*

The proof in [19] uses Ricci-flat metrics on K3 surfaces and their compatibility with the volume form $\Omega \wedge \overline{\Omega}$. Indeed, the Kähler form ω associated to a Ricci-flat metric satisfies the identity $\omega \wedge \omega = \Omega \wedge \overline{\Omega}$ (as volume forms on X) and this poses constraints on the dynamical invariants, such as Lyapunov exponents.

Rough currents. A pseudo-Anosov homeomorphism of a real 2-dimensional surface expands/contracts a pair of measured foliations on the surface. This is a basic result of Thurston's analysis of mapping class group elements. Analogous objects, called closed positive currents, have been constructed by Cantat on K3 surfaces [4], and earlier for polynomial maps of the plane by Bedford–Lyubich–Smillie [3]. Theorem 6 implies that if a K3 surface is not Kummer and admits a positive entropy automorphism, then the measure of maximal entropy is singular for Lebesgue measure. General dynamical considerations imply that its Hausdorff dimension is strictly below the maximal one, and thus the closed positive currents defined above must also have less than maximal Hausdorff dimension, see [18].

4 An overview

The geometry and topology of surfaces is a subject with a long history. Many fundamental topics have been omitted in the above discussion, yet they all play a role in motivating constructions and formulating questions in the subject. For example, although Riemann surfaces do not admit infinite-order holomorphic automorphisms in genus at least two, the study of topological automorphisms (homeomorphism and diffeomorphisms) is essential for much of low-dimensional topology in the form of the Nielsen–Thurston theory of the classification of mapping class group elements (see the monographs of Farb–Margalit [12] and the collection of articles [1]). This leads to the study of measured foliations on surfaces, and their analogues on algebraic surfaces that become closed positive currents. The geometry of these last objects is far less understood than that of surface foliations.

In all instances, moduli spaces of geometric structures play a crucial role. The Teichmüller and moduli spaces of Riemann surfaces are essential for understanding the topology of surfaces, and in the case of algebraic surfaces, moduli spaces of metrics play a similar role. In the case of K3 surfaces, the moduli spaces turn out to be locally homogeneous, and this makes available all the tools of homogeneous dynamics.

Finally, understanding the dynamics in moduli spaces requires one to understand dynamical invariants called *Lyapunov exponents*, which play a role similar to entropy. The tools of complex geometry and Hodge theory turned out to be crucial in gaining control over these otherwise elusive dynamical invariants, and these techniques

Table 1. Parallels between the geometry of Riemann and K3 surfaces

Riemann surfaces	K3 surfaces
Mapping classes of diffeomorphisms: pseudo-Anosov, reducible, periodic	Holomorphic automorphisms: hyperbolic, parabolic, elliptic
Entropy, action on curves	Entropy, action on H^2
Stable and unstable foliations	Stable and unstable currents
Teichmüller space	Period Domain(s)
Flat metrics	Ricci-flat metrics
Holomorphic 1-form	Holomorphic 2-form
Straight lines for flat metric	Special Lagrangians
Periodic trajectories	Special Lagrangian tori
Completely Periodic Foliations	Torus Fibrations
$\mathbb{S}^1$: directions for straight lines	$\mathbb{S}^2$: twistor rotation
Lyapunov exponents for families	

are behind many of the theorems described above. This connection was originally made by Kontsevich [23], see also [8, 7, 16, 15] for further developments related to Lyapunov exponents and Hodge theory.

We end with a summary of the above parallels between the geometry of Riemann and K3 surfaces in Table 1.

References

[1] A. Fathi, F. Laudenbach, and V. Poénaru, *Travaux de Thurston sur les surfaces – Séminaire Orsay*. Astérisque 66, Société Mathématique de France, Paris (1979)

[2] A. Beauville, Géométrie des surfaces $K3$: modules et périodes. Astérisque, Société Mathématique de France, Paris (1985)

[3] E. Bedford, M. Lyubich and J. Smillie, Polynomial diffeomorphisms of $\mathbf{C}^2$. IV. The measure of maximal entropy and laminar currents. *Invent. Math.* **112**, 77–125 (1993)

[4] S. Cantat, Dynamique des automorphismes des surfaces $K3$. *Acta Math.* **187**, 1–57 (2001)

[5] S. Cantat and C. Dupont, Automorphisms of surfaces: Kummer rigidity and measure of maximal entropy. *J. Eur. Math. Soc. (JEMS)* **22**, 1289–1351 (2020)

[6] A. Eskin, S. Filip and A. Wright, The algebraic hull of the Kontsevich–Zorich cocycle. *Ann. of Math. (2)* **188**, 281–313 (2018)

[7] A. Eskin, M. Kontsevich, M. Möller and A. Zorich, Lower bounds for Lyapunov exponents of flat bundles on curves. *Geom. Topol.* **22**, 2299–2338 (2018)

[8] A. Eskin, M. Kontsevich and A. Zorich, Sum of Lyapunov exponents of the Hodge bundle with respect to the Teichmüller geodesic flow. *Publ. Math. Inst. Hautes Études Sci.* **120**, 207–333 (2014)

[9] A. Eskin and C. McMullen, Mixing, counting, and equidistribution in Lie groups. *Duke Math. J.* **71**, 181–209 (1993)

[10] A. Eskin and M. Mirzakhani, Invariant and stationary measures for the $\mathrm{SL}(2,\mathbb{R})$ action on moduli space. *Publ. Math. Inst. Hautes Études Sci.* **127**, 95–324 (2018)

[11] A. Eskin, M. Mirzakhani and A. Mohammadi, Isolation, equidistribution, and orbit closures for the $\mathrm{SL}(2,\mathbb{R})$ action on moduli space. *Ann. of Math. (2)* **182**, 673–721 (2015)

[12] B. Farb and D. Margalit, *A primer on mapping class groups*. Princeton Mathematical Series 49, Princeton University Press, Princeton, NJ (2012)

[13] S. Filip, Semisimplicity and rigidity of the Kontsevich–Zorich cocycle. *Invent. Math.* **205**, 617–670 (2016)

[14] S. Filip, Splitting mixed Hodge structures over affine invariant manifolds. *Ann. of Math. (2)* **183**, 681–713 (2016)

[15] S. Filip, Zero Lyapunov exponents and monodromy of the Kontsevich–Zorich cocycle. *Duke Math. J.* **166**, 657–706 (2017)

[16] S. Filip, Families of K3 surfaces and Lyapunov exponents. *Israel J. Math.* **226**, 29–69 (2018)

[17] S. Filip, Counting special Lagrangian fibrations in twistor families of K3 surfaces. *Ann. Sci. Éc. Norm. Supér. (4)* **53**, 713–750 (2020)

[18] S. Filip and V. Tosatti, Smooth and rough positive currents. *Ann. Inst. Fourier (Grenoble)* **68**, 2981–2999 (2018)

[19] S. Filip and V. Tosatti, Kummer rigidity for K3 surface automorphisms via Ricci-flat metrics. arXiv:1808.08673 (August 2018). To appear in *Am. J. Math.*

[20] G. Forni and C. Matheus, Introduction to Teichmüller theory and its applications to dynamics of interval exchange transformations, flows on surfaces and billiards. *J. Mod. Dyn.* **8**, 271–436 (2014)

[21] M. Gromov, On the entropy of holomorphic maps. *Enseign. Math. (2)* **49**, 217–235 (2003)

[22] D. Huybrechts, *Lectures on K3 surfaces*. Cambridge Studies in Advanced Mathematics 158, Cambridge University Press, Cambridge (2016)

[23] M. Kontsevich, Lyapunov exponents and Hodge theory. In *The mathematical beauty of physics (Saclay, 1996)*, Adv. Ser. Math. Phys. 24, World Sci. Publ., River Edge, NJ, 318–332 (1997)

[24] G. A. Margulis, Formes quadratriques indéfinies et flots unipotents sur les espaces homogènes. *C. R. Acad. Sci. Paris Sér. I Math.* **304**, 249–253 (1987)

[25] C. T. McMullen, Billiards and Teichmüller curves on Hilbert modular surfaces. *J. Amer. Math. Soc.* **16**, 857–885 (2003)

[26] C. T. McMullen, Teichmüller curves in genus two: Torsion divisors and ratios of sines. *Invent. Math.* **165**, 651–672 (2006)

[27] C. T. McMullen, Dynamics of $\mathrm{SL}_2(\mathbb{R})$ over moduli space in genus two. *Ann. of Math. (2)* **165**, 397–456 (2007)

[28] M. Möller, Periodic points on Veech surfaces and the Mordell–Weil group over a Teichmüller curve. *Invent. Math.* **165**, 633–649 (2006)

[29] M. Möller, Variations of Hodge structures of a Teichmüller curve. *J. Amer. Math. Soc.* **19**, 327–344 (2006)

[30] M. Ratner, On Raghunathan's measure conjecture. *Ann. of Math. (2)* **134**, 545–607 (1991)

[31] M. Ratner, Raghunathan's topological conjecture and distributions of unipotent flows. *Duke Math. J.* **63**, 235–280 (1991)

[32] W. A. Veech, Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. *Invent. Math.* **97**, 553–583 (1989)

[33] A. Wright, From rational billiards to dynamics on moduli spaces. *Bull. Amer. Math. Soc. (N.S.)* **53**, 41–56 (2016)

[34] S. T. Yau, On the Ricci curvature of a compact Kähler manifold and the complex Monge–Ampère equation. I. *Comm. Pure Appl. Math.* **31**, 339–411 (1978)

[35] Y. Yomdin, Volume growth and entropy. *Israel J. Math.* **57**, 285–300 (1987)

[36] A. Zorich, Flat surfaces. In *Frontiers in number theory, physics, and geometry. I*, Springer, Berlin, 437–583 (2006)

Simion Filip is Associate Professor of Mathematics at the University of Chicago.

sfilip@math.uchicago.edu

All roads come from China –
For a theoretical approach to the history of mathematics

Karine Chemla

This article presents some of the theoretical issues that interest me in the history of mathematics. Each of them has its origin in the work I have done on mathematical sources in Chinese. However, they all have ramifications in other bodies of mathematical literature, and I have pursued them beyond Chinese sources.

To an outside observer, I suppose I appear to be working on the history of mathematics in ancient and medieval China. To a certain extent, this is true. However, this is also partly wrong. By this (perhaps unexpected) statement, I do not mean simply that I have also carried out research and published on the history of projective geometry and of duality more broadly, as well as on the history of medieval mathematics in Arabic, Greek, Hebrew and Sanskrit. I mean something deeper. Working on the history of mathematics in China is certainly meaningful in and of itself. However, to my eyes, it becomes all the more meaningful in that it confronts us with sources with which we are not used to thinking about mathematics, and these sources suggest interesting new issues, as well as new ways of addressing old issues. In other words, Chinese sources, like in fact any mathematical document if treated appropriately, give us resources with which to nurture a theoretical approach to the history of mathematics. This, in the end, is my main goal. In what follows, I will illustrate how this has worked for me in practice, by discussing some of the theoretical issues I have been led to address in the course of my research.

1 History of science, history of text

My first significant encounter with Chinese mathematical sources took place in 1981, as I was studying in China at the Institute for the History of Natural Sciences (Chinese Academy of Sciences), and it confronted me right away with striking phenomena, about which I still think today.

Following a suggestion that had been made to me by Leuven sinologist Ulrich Libbrecht, I started reading the book that Li Ye 李冶 (1192–1279) had published in 1248 under the title *Measuring the Circle on the Sea-Mirror* (*Ceyuan haijing* 測圓海景, hereafter *Measuring the Circle*), which was to become the subject of my dissertation. For this, I benefited from the guidance of the person in charge of organizing my study in Beijing at the time, Mei Rongzhao 梅榮照, who had already worked on Li Ye's book. I was also lucky to receive advice from the group of scholars who had been appointed to teach me during my time in China, namely: Du Shiran 杜石然, Guo Shuchun 郭書春, He Shaogeng 何紹庚 and Yan Dunjie 嚴敦傑.

Li Ye's book opens with a diagram, to which the entire book is devoted (see Figure 2, and Chemla [7] for an analysis). The diagram is followed by a set of about 700 formulas, stating relationships between its segments, and then 170 problems, which basically all share the same structure. They give two segments of the diagram and in general ask to determine the diameter of the circle. The point of the problems is thus not the answer, since it is systematically the same, but rather the method. Li Ye begins with the choice of an unknown (not always the diameter itself, but a magnitude that could easily be related to it), to which he refers as "the celestial origin". He then brings into play polynomials, written using a place-value notation, along with a geometrical reasoning that relies on the data and the unknown, in order to establish an algebraic equation, "the" root of which is the unknown sought (see Figure 1). Indeed, at the time, in China, equations were considered as having a single root. In brief, this was how the book had been understood up to then: it was the earliest extant book attesting to the algebraic method known as "the procedure of the celestial origin *tian yuan shu* 天元術."

However, something immediately struck me in the solutions Li Ye gave to the problems. Every solution had the same structure, which consisted of two parts. Each of these parts described, in a different way, how to obtain the same equation that solved the problem. The first part, called "method" (*fa* 法), described a sequence of algorithms that relied on the data to compute the successive coefficients of the sought-for equation. In this part, there were no numerical values, in contrast with the second part, called "detail of the procedure" (*cao* 草), which, starting from the data and the chosen unknown, presented two ways of reasoning to obtain the same geometrical magnitude and a numerical polynomial associated with it. The reasonings, along with the related polynomial computations, systematically followed the same pattern: each step consisted of an operation that took previously determined magnitudes and the associated polynomial as its operands, and then yielded a result in the form of another magnitude (the reasoning part) and the polynomial associated with it (the computation part). At the end of a procedure of this kind, the equation was obtained numerically, by subtracting from each other the two polynomials that corresponded to the same magnitude. Why, I wondered, should the author systematically tell the reader, twice and in two different ways, how to get the same equation? This was my first question, soon followed by a second one: taking for granted that the "method" and the "detail of the procedure" led to exactly the same equation, how were the algorithms given in the "method" obtained?

For each of the 170 problems, I made an experiment. I computed the sequences of polynomials leading to the final equation in the "details of the procedure" symbolically, and not numerically as they were presented in Li Ye's book. Although the text did not contain any computation of this kind, I established that, in each case, my computations highlighted a missing link between the algorithms of the "method" and the "details of the procedure". Indeed, every algorithm in the "method" actually described the sequence of operations that, in the symbolic computations deriving from the "details of the procedure", had been applied to coefficients of successive polynomials to shape the corresponding coefficient of the final equation [2]. In brief, using mathematical knowledge and practice that did not feature in the book, and that appeared long after the book was completed (that is, algebraic symbolism and algebraic computations), I could highlight a correlation between the two parts of every solution. The correlation was so intimate that the "method" could not have been obtained independently from the "details of the procedure". Clearly, the systematic correlation yielded a clue indicating that the description in the "method" derived from a work that Li Ye had carried out, but not recorded in his book. So the question became: what kind of mathematical work was that?

Perhaps, in the future, someone will find clues in *Measuring the Circle*, or elsewhere, to answer this question with certainty. However, as far as we know today, nothing in the book seems to indicate exactly how, for every single problem, Li Ye produced the "method" part of the solution, relying on the "details of the procedure" part. I cannot attribute to him without further ado the knowledge that I, as an observer, bring into play to establish the correlation between the "method" and the "details of the procedure". Nevertheless, my experiment sheds light on knowledge that Li Ye must have possessed, and practices that he must have used, in order to write *Measuring the Circle* as it stands, even though I cannot describe them precisely since he did not expose them himself, even indirectly. As historians of mathematics, we cannot content ourselves with a superficial reading of the book and offer a historical treatment that would ignore this new dimension that studying *Measuring the Circle* allows us to perceive. We are committed to try to account for the knowledge the actors we observe possessed and the practices they put into play, even when these were not the objects of discursive exposition.

This example illustrates why, in order to fully accomplish their task, historians must look for clues and then strive to interpret those clues as best as they can. One might of course be tempted to consider this case as an exception and an outlier; however, since I began working as a historian, my experience has convinced me of the contrary, not only because other similar phenomena occur in Li Ye's book, but because they actually occur much more broadly. In fact, as early as 1974, drawing on discussions with Igor Shafarevich, Isabella Bashmakova once showed something quite similar about the four books of Diophantus' *Arithmetics* that still exist in Greek [1]. Ten years later, Roshdi Rashed fully developed this approach and observed the same phenomena in his publication of the four other books that had just resurfaced in Arabic [17]. These historians used insights from modern algebraic geometry to analyze the procedures Diophantus followed in the *Arithmetics* to solve Diophantine problems. This reading, instrumented by a type of mathematical knowledge that Diophantus certainly did not possess, revealed something that completely contradicted previous interpretations, according to which Diophantus was fundamentally unpredictable in his approach to a problem, even after one has read dozens of his solutions to other problems. Indeed, the analysis of the *Arithmetics* using algebraic geometry showed that Diophantus' solutions systematically made use of the same methods. Exactly as was the case for Li Ye, we cannot attribute to Diophantus knowledge of the tool modern historians put into play to read the *Arithmetics*. However, this tool brings to light knowledge that Diophantus possessed and practices he used without recording them. How can we approach his knowledge and practices on the basis of these clues? This is the theoretical problem raised by these phenomena [3].

In the cases of Li Ye and Diophantus, the clues provided by a certain type of mathematical reading reveal facets of the knowledge and practice that these authors have put into play in their approaches to specific problems, without, however, writing about them. In fact, clues can do more for historians, as Anne Robadey has illustrated in her work on Henri Poincaré. For instance, Robadey [18] starts from the remark that Poincaré's publications abound in enumerations, and she sets out to analyze what these textual phenomena can teach us about the way Poincaré carried out his mathematical work. Robadey [18] establishes that these textual clues reveal an intellectual practice that Poincaré recurrently put into play in different contexts and that left traces not only in his writings but also in the type of mathematical results he formulated. Indeed, faced with certain mathematical situations, Poincaré regularly analyzes them, focusing first on the case that presents itself most often (in

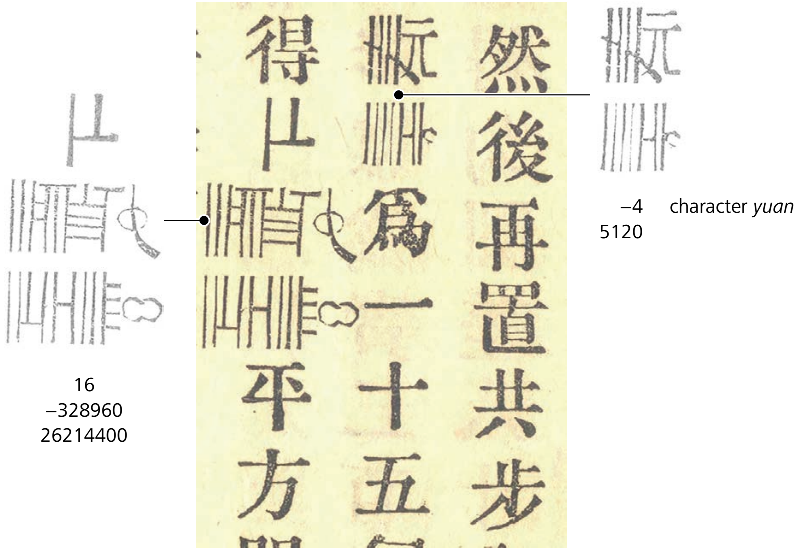


Figure 1. Li Ye, *Measuring the Circle on the Sea-mirror*
To the right and to the left, resp., a polynomial ($-4x + 5120$)
and an equation ($16x^2 - 328960x + 26214400 = 0$) are written
using a place-value notation.

a sense of the latter expression for which Poincaré puts forward a definition and an assessment), then on the second most frequent case, and so on, until reaching phenomena that he thinks he can disregard, since they "almost never" occur (on the basis of an assessment of the same type). The enumerations embody precisely this recurring intellectual procedure that Poincaré follows. Moreover, they mesh with theorems of Poincaré's in which he asserts that something holds true except for a set of situations that can be neglected. In this example, textual phenomena, mathematical practice and mathematical results appear to be closely intertwined. How, as historians, we can find clues that allow us go deeper in our analysis, and how we can use them in historical research are precisely two of the main issues that I study, not least in the context of the project "History of science, history of text", which I launched in 1995 and on which I have been working since then with a group of colleagues (see, e.g., [5]).

The result obtained by Robadey that I have summarized naturally leads us to another interesting issue: why do practitioners of mathematics not always present their knowledge and practices "explicitly" (as we would be tempted to say, but I explain below why this term is inadequate), to the extent that historians need to rely on clues to uncover part of this knowledge and these practices? The example of Poincaré's enumerations suggests a first answer to this question. If his publications yield the clues I have mentioned, the reason seems to be that Poincaré carries his analysis forward while engaging with his page. The page thereby keeps the trace of the procedure that recurringly structures his mathematical exploration. This remark explains why the writing gives us clues about his way of conducting mathematical work. Poincaré chooses to work with the textual structure of the enumeration, since it offers a support on which he can rely to unfold his reasoning.

We can observe a similar phenomenon in the prehistory and history of duality, on which I have begun to work with Serge Pahaut [13]. If we considered that duality emerged in mathematics at the point when actors first explicitly mentioned the phenomenon, we would set its beginnings in the 1820s. However, Pahaut and I noticed that starting from the 1750s, some mathematicians who published on spherical trigonometry chose new notations, and shaped types of text, both of which were appropriate to highlight a phenomenon that they had observed without thematizing it. Using new notation, Leonhard Euler, for instance, presented in 1753 a memoir about spherical trigonometry that is remarkable for the following reason: its text displayed, without any comment, a symmetry in a corpus of propositions asserted, and also in a corpus of proofs establishing these propositions. Today, we associate this symmetry with the duality that affects spherical trigonometry.

Euler did not address this phenomenon discursively. For him, as for several mathematicians who wrote about spherical trigonometry in the same way in the following decades, this was a phenomenon to be explored, and, instead of writing about it in a discursive way, they expressed what they observed using textual features of their writings: they gave it to readers to read off from the structure of the text. Should we call such a way of expressing knowledge "not explicit" simply because it is not expressed with a subject, a verb and a complement? I don't think so. This would be quite a narrow interpretation of what "explicit" might mean.

Indeed, we can establish that for these mathematicians, writing in this way was a genuine choice. The reasoning goes as follows. In a second memoir on the topic that Euler presented in 1781, Pahaut and I were able to show that he made a mistake in a proof, which was then replicated in the dual proof. This clue thus indicates that Euler relied on the notation to produce the dual theorem and the dual proof by mere rewriting of the corresponding theorem and proof, without actually redoing the computation. In other words, Euler knew that a theorem and a proof could automatically give rise to another theorem and another proof, but he chose to present both systematically and to cast light on the symmetry between them by means of the structure of his exposition. This remark allows us to establish another key point: the notation appropriate to investigate phenomena related to duality constituted a tool created by Euler to work with and to produce a text that displayed the symmetry. More generally, texts are not always merely discursive expositions of knowledge, as a modern reading all too often expects. This remark might seem obvious for rough drafts, but Euler's inquiry into duality and Poincaré's enumerations show that it also applies to texts intended for publication. We see mathematicians shaping notations and textual resources and developing practices using them, in order to work with them and explore new phenomena. As a result, these textual resources and practices present intimate correlations with the questions these actors pursue and the research they conduct. I take these textual innovations as a key dimension of their activity.

This observation highlights one of the reasons why, as a result, texts can give clues about the mathematical work that produced them and also about the knowledge that mathematicians acquired through working with them. In Euler's case, he met more than once with phenomena caused by duality, and regularly made use of similar textual resources. Interestingly, when dealing with the same topics, subsequent mathematicians used notation and textual resources that were either identical or similar to Euler's, which indicates that notation and textual resources are, like mathematical theories and concepts, products of mathematical work that get picked up and used further by others [4]. The joint production of knowledge and textual resources (in the broadest sense of this expression) is likewise one of the theoretical issues that interest me most.

If we pursue this line of thought, we see that sometimes, in order to deal with specific topics, new types of textual resources are introduced (like writing propositions and proofs in a symmetrical way), and that some among the subsequent readers will not only grasp what is being given to read in this new manner, but also then go on to reuse the new textual resources to carry on further research along the same lines. However, not all readers will notice what is given to read in this way. For instance, historians of mathematics had not underline what the structures of these texts expressed, at least for works written before Joseph Diez Gergonne's explicit introduction, in 1826, of a double-column device to display the symmetry between propositions and proofs elicited by duality [14]. My purpose is not to blame these historians, but to draw a conclusion from this observation. Obviously, we do not all read in the same way, particularly because we have not all been acculturated to reading mathematical writings in the same way. Reading (and reading mathematical texts is no exception) has a history that itself deserves to be studied in the various contexts in which it has been carried out over time, in order to better account for what our sources convey in ways that are not always obvious to us. This latter issue, and more generally those brought to light in this section, have turned out to be central in basically every single piece of research that I have conducted.

2 *The Nine Chapters*: Algorithms, proofs, and epistemological values

While working on *Measuring the Circle*, it appeared to me that this book was deeply rooted in an ancient Chinese canonical work in mathematics to which Li Ye explicitly refers, namely the first-century classic *The Nine Chapters on Mathematical Procedures* (*Jiuzhang suanshu* 九章算術, hereafter *The Nine Chapters*). In fact, most mathematical writings composed in China before the fourteenth century referred to this work and to the commentaries with which it has been handed down, that is, Liu Hui's 劉徽 commentary, completed in 263, and the subcommentary published in 656 by a team working under the supervision of Li Chunfeng 李淳風. When, as early as 1981, Guo Shuchun suggested that we could cooperate to translate *The Nine Chapters* and its commentaries into French, I thus found the project meaningful and accepted without hesitation. We agreed in 1983 that in addition to the translation, our joint book would offer a new critical edition of these texts as well as our own annotations, unaware that these tasks would take us over twenty years to complete [12].

The project was difficult not only because the Chinese text was hard and the establishment of the critical edition challenging, but also because the endeavor raised many theoretical problems that I felt we needed to address to complete our task satisfactorily. I will illustrate some of these problems while outlining some

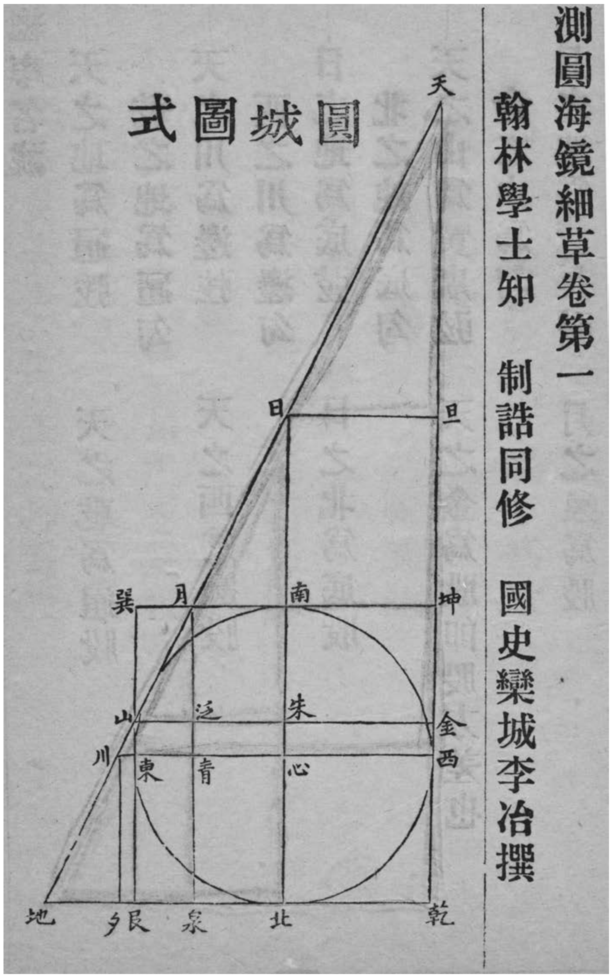


Figure 2. The diagram in *Measuring the Circle on the Sea-Mirror* (1248) *Tongwenguan* Edition 同文館, 1876

of the research directions I have followed in my research on *The Nine Chapters* and the more theoretical projects that this work has inspired me with.

The Nine Chapters was composed of 246 mathematical problems along with procedures to solve such problems. The "problem and procedure" form had led some historians to read it either as an exercise book or as a manual for bureaucrats, who would simply need to pick up instructions and follow them blindly. It seemed to me that these interpretations, which derived from a modern reading of an ancient text, could not explain why the book had been considered a classic for centuries. We thus needed to find ways of reading *The Nine Chapters* that could be less anachronistic and for which we could put forward arguments.

My efforts first bore on the procedures. In 1972, Donald Knuth, whose work on algorithms needs no introduction, published an article that had a significant impact on the history of ancient mathematics [16]. Indeed, his article was proposing a completely new way of approaching cuneiform texts of the beginning of the second millennium BCE, by reading them as they were written, i.e., as lists of operations, or "algorithms", and not by rewriting them into modern algebraic formulas, as had been the case until then. I became aware of this breakthrough in 1981, thanks to discussions with Wu Wenjun, a topologist who had turned to automated theorem proving and the history of mathematics in China during the cultural revolution [15]. Wu immediately adopted Knuth's perspective in his reading of the procedures contained in ancient Chinese mathematical texts, since they too were written in the form of lists of operations. What mattered most to him was the emphasis that Chinese texts of the past had placed on algorithms providing constructive means to solve problems. A typical example is the algorithm described by Qin Jiushao 秦九韶 in his 1247 book, in order to compute actual solutions to problems whose solution is known to exist by the Chinese remainder theorem [19]. I became more interested in the work with lists of operations to which the procedures of *The Nine Chapters* attest.

This work is eloquently illustrated by the procedures given in *The Nine Chapters* for square and cube root extraction, whose texts put into play three operations that Knuth identified as fundamental in the writing of algorithms: assignment of variables, iterations and conditionals. Before Knuth introduced the idea of reading lists of operations as algorithms, historians knew which computations these texts referred to, but I claim they were unable to understand *how* these texts in *The Nine Chapters* actually referred to these computations. Thanks to the recent development of mathematical work on algorithms, in relation to their implementation on computers, mathematicians have shaped new types of texts to write procedures, and these textual resources have given us new insights into how ancient procedural texts might have been written and consequently how we might interpret them [10]. We see again how, in different contexts, actors put into play different types of textual resources to carry out mathematical work and to write about it. In the case of the procedures of *The Nine Chapters*, we could not fully interpret their texts and understand the work involved in producing them without first asking ourselves how these texts were expected to be read and how mathematical work put them into play. With respect to, e.g., the algorithms for root extractions, this research brought to light two key points.

First, the way in which the author(s) had used conditionals and iterations to write a list of operations on the basis of which any square (resp. cube) root extraction could be performed highlighted important features of the work with operations in this context. To explain this point, let me make clear that the execution relied on a decimal place-value system and that the roots were determined digit by digit. On this basis, the list of operations used for the first digit and the one used for each digit after the first one had been shaped in such a way that they could be integrated into a single text. Without entering into details (for which I refer to the 2004 book [12]), let me simply emphasize that the integration relied on the assignment of variables. It also relied on the fact of treating operations formally and without taking their intention into account. More specifically, highlighting an operation common to the concluding part of a root extraction (when the digit of the units has been dealt with) and to the preparation of the computation for the next digit, if any, even though the purpose of using this operation differed in the two contexts, as well as placing the statement of this operation in the list of operations adequately with respect to the conditionals and the iteration, played a key part in allowing the authors to compose a single algorithm valid for all cases. In *The Nine Chapters* and their commentaries, we find more generally many indications that the authors worked on lists of operations formally, without taking into account the variety of reasons for bringing these operations into play in each of their respective contexts. As a result, we regularly see authors striving to unify lists of operations that performed different tasks, but could be made formally identical. This highlights a form of algebra specific to the work with operations, to which I have devoted some research, but on which much remains to be done.

The latter remarks lead me to the second key point. A search of the kind just described with respect to algorithms, that is, a search for lists of operations whose efficacy would extend as broadly as possible, bespeaks actors' valuing of generality. The fact of giving a single algorithm for square (resp. cube) root extraction points in the same direction. The text of the algorithm was general in the sense that for any number, an adequate circulation within it, guided by the conditionals and the iteration, would yield the list of operations required to determine the desired square root. What is more, the text added this: should the extraction not be completed when reaching the digit of the units, the result should be given as the "side of the number", i.e., as a quadratic irrational. If we considered this suggestion from the viewpoint of the discussions about irrationality by Greek authors of antiquity, we would completely miss the point – I return to this below. At stake in the interpretation of the text of the square root algorithm is thus a better appreciation of how it reflects the importance actors in this context gave to the epistemological value of generality.

In fact, generality proved to be a key value for these actors much more broadly [9]. For instance, the way in which commentators read problems in *The Nine Chapters* indicated that for them, a particular problem was to be interpreted as a paradigm. This might seem obvious: the problems from our childhood about trains passing each other were not meant to stand only for themselves, but expressed something more general. However, an observation of the commentators' way of reading problems in the classic shows that they meant something more specific. In a key case, when the procedure placed after a problem correctly solves it, but lacks generality, the third century commentator Liu Hui expresses dissatisfaction. After pointing out that the procedure of *The Nine Chapters* is based on the use of two singular characteristics of the problem, he proposes a first procedure that fixes the failure of the original procedure to solve the most general problem, distinguishing between two cases, and then a completely general and uniform procedure. In other words, for Liu Hui, the fact that a mathematical problem was not abstract did not affect the expectation he had with respect to its generality. This remark, inspired by this Chinese document, raised an important theoretical issue: it was an invitation to dissociate the values of generality and abstraction in our reflection about mathematics and to see what a focus on generality alone might show. The emergence of projective geometry in France during the first decades of the 19th century proved to be an ideal case for me to address this issue. Indeed, this new geometry took shape in the hands of Carnot, Poncelet, Chasles and others, on the basis of a comparative reflection about the different types of generality brought about by analytic and geometric approaches to geometric problems. More broadly, this direction of research proved fruitful for a group of historians and philosophers of science, as is illustrated by the collective reflection we developed on this issue [11].

The remarks that I have presented about the problems of *The Nine Chapters* illustrate a method that I have used more systematically. Indeed, if we need to restore how ancient actors used and read the texts with which they performed mathematical activity, or, in other words, if we need to develop a history of the reading and handling of ancient mathematical texts, observing how ancient readers proceeded seems to be a method that has great potential. This is precisely one of the reasons why it is so valuable to have early commentaries on *The Nine Chapters*. If, for instance, we continue to rely on them to better understand how and why ancient actors used problems in their mathematical practice, we discover something quite unexpected, which definitively rules out the interpretation of *The Nine Chapters* as an exercise book or as a manual for bureaucrats.

This facet of their practice also appears when we turn to another key point about the commentaries on *The Nine Chapters*, namely that they systematically put forward proofs of the correctness of the algorithms presented in the classic. This is quite an important fact for a history of mathematical proof, to which I return shortly. What matters here is that the commentators' way of carrying out proving brings mathematical problems into play. To put it differently, in their practice, mathematical problems appear to be tools with which to conduct proofs, and not merely statements awaiting a solution [6]. If we think that the same fact held true for the author(s) of *The Nine Chapters,* this invites a radically new interpretation of the work. The reading of problems and procedures that I have suggested might help us understand better how *The Nine Chapters* has been considered a canonical work over so many centuries. What is more, the fact that *The Nine Chapters* was handed down with these commentaries might have also played a part in giving the book its value in the eyes of its users. This remark brings us back to the proofs that Liu Hui as well as the team working with Li Chunfeng formulated.

In contrast with what we read in Euclid's *Elements* and Archimedes's writings, these proofs aim at establishing the correctness of procedures. Observing them hence gives us source material to think about this other branch of the history of proof that has so far been almost completely neglected, and, more broadly, about the various dimensions of the exercise of proving in mathematics [8]. What is essential here is that the commentators use theoretical concepts to refer to key aspects of the conduct of a proof.

To begin with, they devote a specific term, i.e., "meaning/intention" (*yi* 意), to designate the meaning of an operation that corresponds to the interpretation of its result in the context in which the operation is used. Typically, for an operation, this is the kind of meaning that the context of a problem enables a practitioner to make explicit. By extension, the term *yi* also refers to a sequence of meanings of this kind, and in the end to the reasoning from which the sequence derives. As a rule, a reasoning of this type consists in making clear the "meaning" of the successive steps of an algorithm, thereby showing why its end result corresponds to what was expected. Interestingly, we find here an echo with the type of reasoning Li Ye expounded in his "details of the procedure", which we mentioned at the beginning of this article. The only difference lies in this: in Li Ye's case, instead of yielding a meaning and a number, each operation yields a meaning and a polynomial. We nevertheless see that there might be traditions of reasoning to which Chinese writings attest, but which were not yet studied.

The second term used by commentators of *The Nine Chapters* in the context of their proofs, which I denote by *yi'* 義, refers to another type of "meaning" for procedures. It designates a fundamental procedure that underlies the procedure whose correctness must be established. As part of the proof of the correctness, this fundamental procedure highlights the strategy followed by the algorithm under consideration. At the same time, identifying it connects this algorithm with others, which follow the same formal strategy, even though the reasons for using the same operations might differ, depending on the context. The interest that commentators have for this kind of "meaning" thus appears to be connected with the formal work on operations to which the algorithms contained in *The Nine Chapters* also attest. This focus of their proofs is in fact more broadly connected with a research program for which we have evidence between the first and the thirteenth century and which aimed at identifying the least number of algorithms from which all the others derive [9].

The last set of terms that commentators use for their proofs relates to what I have called "algebraic proofs in an algorithmic context". A proof of this kind consists in establishing a list of operations that starts from the same data as the algorithm under consideration, and yields the desired result. The commentator then takes the algorithm established as correct as a basis, and operates on its list of operations to transform it, *qua* list of operations, into the algorithm whose correctness is to be established. In other words, instead of rewriting equalities, as we do in an algebraic proof, this type of proof rewrites algorithms. The meta-operations applied to the list of operations include swapping multiplication and division that follow each other, and cancelling a multiplication and a division inverse of one another. They also include inverting an algorithm known to be correct. The essential point here is that the commentators associate the correctness of these meta-operations with the fact that divisions and square root extractions are given exact results, notably through the introduction of fractions and quadratic irrationals. They thus bring the set of numbers used and the meta-operations applied to a list of operations in relation with each other. Moreover, here again, we see that these "algebraic proofs in an algorithmic context" also involve formal work on lists of operations.

What does all this tell us about the history of algebraic proof, of which we still lack a proper account? What does it tell us about the history of algebra and the part played by operations in the history of mathematics? These are some of the theoretical questions that remain on my agenda.

Acknowledgements. I have pleasure in expressing my gratitude to Fernando Manuel Pestana da Costa for his invitation to write this article and Leila Schneps for her feedback, which greatly improved it.

References

[1] I. G. Bašmakova, *Diophant und diophantische Gleichungen*. Translated from the Russian 1972 original by L. Boll, Birkhäuser, Basel-Stuttgart (1974)

[2] K. Chemla, Equations with general coefficients in the *Ce Yuan Hai Jing*. *Publications de l'Institut de Recherche Mathématique de Rennes, Fascicule II: Science, Histoire, Société*, 23–30, www.numdam.org/article/PSMIR_1985___2_23_0.pdf (1985)

[3] K. Chemla, What is the content of this book? A plea for developing history of science and history of text conjointly. *Philosophy and the History of Science: A Taiwanese Journal* **4**, 1–46 (1995) [Republished in [5]]

[4] K. Chemla, Euler's Work in spherical trigonometry: Contributions and applications. In *Euler. Opera Omnia. Commentationes physicae ad theoriam caloris, electricitatis et magnetismi pertinentes. Appendicem addidit Karine Chemla*, edited by P. Radelet-de Grave and D. Speiser, CXXV–CLXXXVII, Birkhäuser, Basel (2004)

[5] K. Chemla (ed.), *History of science, history of text*. Boston Studies in the Philosophy of Science 238, Springer, Dordrecht (2004)

[6] K. Chemla, On mathematical problems as historically determined artifacts: reflections inspired by sources from ancient China. *Historia Math.* **36**, 213–246 (2009)

[7] K. Chemla, Une figure peut en cacher une autre. Reconstituer une pratique des figures géométriques dans la Chine du XIIIe siècle. *Images des mathématiques*, images.math.cnrs.fr/Une-figure-peut-en-cacher-une.html (2011)

[8] K. Chemla (ed.), *The history of mathematical proof in ancient traditions*. Cambridge University Press, Cambridge (2012)

[9] K. Chemla, *The Motley Practices of generality in various epistemological cultures, The Hans Rausing lecture 2017*. Salvia Småskrifter, Uppsala, www.idehist.uu.se/digitalAssets/775/c_775182-l_1-k_2019motley-practices-of-generality--final-versionrausinglecture2017originalcorrected.pdf (2019)

[10] K. Chemla, From reading rules to reading algorithms. Textual anachronisms in the history of mathematics and their effects on interpretation. In *Anachronisms in the history of mathematics*, edited by N. Guicciardini, Cambridge University Press, Cambridge (to appear)

[11] K. Chemla, R. Chorlay and D. Rabouin (eds.), *The Oxford handbook of generality in mathematics and the sciences*, Oxford Univ. Press, Oxford (2016)

[12] K. Chemla and S. Guo, *Les neuf chapitres*. Dunod, Paris (2004)

[13] K. Chemla and S. Pahaut, Préhistoires de la dualité: explorations algébriques en trigonométrie sphérique (1753–1825). In *Sciences à l'époque de la Révolution Française*, edited by R. Rashed, Lib. Sci. Tech. Albert Blanchard, Paris, 151–201 (1988)

[14] J. D. Gergonne, Considérations philosophiques sur les élémens de la science de l'étendue. *Ann. Math. Pures Appl. [Ann. Gergonne]* **16**, 209–231 (1825/26)

[15] J. Hudeček, *Reviving ancient Chinese mathematics*. Needham Research Institute Studies, Routledge/Taylor & Francis Group, London (2014)

[16] D. E. Knuth, Ancient Babylonian algorithms. *Comm. ACM* **15**, 671–677 (1972)

[17] R. Rashed, *Diophante. Les Arithmétiques*, Tome 3: Livre IV. Tome 4: Livre V–VII. Les Belles Lettres, Paris (1984)

[18] A. Robadey, A work on the degree of generality revealed in the organization of enumerations: Poincaré's classification of singular points of differential equations. In *Texts, textual acts and the history of science*, edited by K. Chemla and J. Virbel, Springer, Dordrecht, 385–419 (2015)

[19] W.-T. Wu, Recent studies of the history of Chinese mathematics. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986)*, Amer. Math. Soc., Providence, RI, 1657–1667 (1987)

Karine Chemla, Senior Researcher at the French National Center for Scientific Research (CNRS), in the laboratory SPHERE (CNRS and Université de Paris), focuses, from a historical anthropology viewpoint, on the relationship between mathematics and the various cultures in the context of which it is practiced.

chemla@univ-paris-diderot.fr

Surface evolution of elastically stressed films

Irene Fonseca and Giovanni Leoni

An overview of recent analytical developments in the study of epitaxial growth is presented. Quasistatic equilibrium is established, regularity of solutions is addressed, and the evolution of epitaxially strained elastic films is treated using minimizing movements.

In this paper, we give a brief overview of recent analytical developments in the study of the deposition of a crystalline film onto a substrate, with the atoms of the film occupying the substrate's natural lattice positions. This process is called epitaxial growth. Here we are interested in heteroepitaxy, that is, epitaxy when the film and the substrate have different crystalline structures. At the onset of the deposition, the film's atoms tend to align themselves with those of the substrate because the energy gain associated with the chemical bonding effect is greater than the film's strain due to the mismatch between the lattice parameters. As the film continues to grow, the stored strain energy per unit area of the interface increases with the film thickness, rendering the film's flat layer morphologically unstable or metastable after the thickness reaches a critical value. As a result, the film's free surface becomes corrugated, and the material agglomerates into clusters or isolated islands on the substrate. The formation of islands in systems such as In-GaAs/GaAs or SiGe/Si has essential high-end technology applications, such as modern semiconductor electronic and optoelectronic devices (quantum dots laser). The Stranski–Krastanow (SK) growth mode occurs when the islands are separated by a thin wetting layer, while the Volmer–Weber (VW) growth mode refers to the case when the substrate is exposed between islands.

In what follows, we adopt the variational model considered by Spencer in [41] (see also [36, 42], and the references contained therein). To be precise, the free energy functional associated with the physical system is given by

$$\int_{\Omega_h} W(\boldsymbol{E}(\boldsymbol{u}))\,dx + \int_{\Gamma_h} \psi(\boldsymbol{\nu})\,d\mathcal{H}^2. \tag{1}$$

Here $h : Q \to [0,\infty)$ is the function whose graph Γ_h describes the profile of the film, assumed to be Q-periodic, with $Q := (0,b)^2 \subset \mathbb{R}^2$, for some $b > 0$, Ω_h is the region occupied by the film, i.e., writing $\boldsymbol{x} = (x,y,z)$,

$$\Omega_h := \{(x,y,z) \in Q \times \mathbb{R} : 0 < z < h(x,y)\},$$

$\boldsymbol{u} : \Omega_h \to \mathbb{R}^3$ is displacement of the material, $\boldsymbol{E}(\boldsymbol{u}) := \frac{1}{2}(D\boldsymbol{u}+D^T\boldsymbol{u})$ is the symmetric part of $D\boldsymbol{u}$. Also, the elastic energy density $W : \mathbb{M}^{3\times 3}_{\mathrm{sym}} \to [0,+\infty)$ is a positive definite quadratic form defined on the space of 3×3 symmetric matrices

$$W(A) := \frac{1}{2}\mathbb{C}A : A,$$

with $\mathbb{C}$ a positive definite fourth-order tensor, so that $W(A) > 0$ for all $A \in \mathbb{M}^{3\times 3}_{\mathrm{sym}} \setminus \{0\}$, $\psi : \mathbb{R}^3 \to [0,\infty)$ is an anisotropic surface energy density evaluated at the unit normal $\boldsymbol{\nu}$ to Γ_h, and $\mathcal{H}^2$ denotes the two-dimensional Hausdorff measure. We suppose that ψ is positively one-homogeneous and of class C^2 away from the origin, so that, in particular,

$$\frac{1}{c}|\boldsymbol{\xi}| \le \psi(\boldsymbol{\xi}) \le c|\boldsymbol{\xi}| \qquad \text{for all } \boldsymbol{\xi} \in \mathbb{R}^3,$$

for some constant $c > 0$.

The substrate and the film admit different natural states corresponding to the mismatch between their respective crystalline structures. To be precise, a natural state for the substrate is given by $\boldsymbol{u} \equiv \mathbf{0}$, while a natural state for the film is given by $\boldsymbol{u} \equiv \boldsymbol{A}_0\boldsymbol{x}$ for some nonzero 3×3 matrix $\boldsymbol{A}_0$. Our models will reflect this mismatch, either by setting the elastic bulk energy as $\int_{\Omega_h} W(\boldsymbol{E}(\boldsymbol{u})(\boldsymbol{x}) - \boldsymbol{E}_0(\boldsymbol{x}))\,dx$, where

$$\boldsymbol{E}_0(\boldsymbol{x}) := \begin{cases} \frac{A_0 + A_0^T}{2} & \text{if } z > 0, \\ \mathbf{0} & \text{if } z \le 0, \end{cases} \tag{2}$$

or by imposing the Dirichlet boundary condition $\boldsymbol{u}(x,y,0) \equiv \boldsymbol{A}_0(x,y,0)$.

In the two-dimensional static case, existence of equilibrium solutions and their qualitative properties, including regularity, were studied in [3, 4, 5, 15, 16, 17, 20, 24, 26, 29, 33]. The variational techniques and analytical arguments developed in these papers have been used to treat other materials phenomena, such as voids and cavities in elastic solids [9, 19].

The scaling regimes of the minimal energy in epitaxial growth were identified in [2, 30] in terms of the parameters of the problem. The shape of the islands under the constraint of faceted profiles was addressed in [25]. A variational model that takes into account the formation of misfit dislocations was introduced in [23].

The effect of atoms freely diffusing on the surface (called adatoms) was studied in [10], where the model involves only surface energies.

A discrete-to-continuum analysis for free-boundary problems related to crystalline films deposited on substrates was undertaken in [35, 38].

The three-dimensional static case was studied in [6, 12] in the case in which the symmetrized gradient $\boldsymbol{E}(\boldsymbol{u})$ is replaced by the gradient (see also [4]). More recently, new developments in the theory of *GSBD*, i.e., generalized special functions of bounded deformation (see [13, 14], and the references therein) have led to considerable progress on the relaxation of the functional (1) in the three dimensional case (see [13]). The regularity of equilibrium solutions remains an open problem. A local minimality sufficiency criterion, based on the strict positivity of the second variation, was established in [4], based on the work [29].

To study the morphological evolution of anisotropic epitaxially strained films, we assume that the surface evolves by *surface diffusion* under the influence of a chemical potential μ. To be precise, according to the Einstein–Nernst relation, the evolution is governed by the *volume preserving* equation

$$V = C\Delta_\Gamma \mu \,, \tag{3}$$

where $C > 0$, V denotes the normal velocity of the evolving interface Γ, Δ_Γ stands for the tangential laplacian, and the chemical potential μ is given by the first variation of the underlying free-energy functional. In our context, this becomes (assuming $C = 1$)

$$V = \Delta_\Gamma\big[\mathrm{div}_\Gamma\big(D\psi(\boldsymbol{\nu})\big) + W\big(\boldsymbol{E}(\boldsymbol{u})\big)\big]\,, \tag{4}$$

where div_Γ stands for the tangential divergence along $\Gamma_{h(\cdot,t)}$, and $\boldsymbol{u}(\cdot,t)$ is the elastic equilibrium in $\Omega_{h(\cdot,t)}$, i.e., the minimizer of the elastic energy under the prescribed periodicity and boundary conditions (see (7) below).

If the surface energy density ψ is highly anisotropic, there may be directions $\boldsymbol{\nu}$ for which

$$D^2\psi(\boldsymbol{\nu})[\boldsymbol{\tau},\boldsymbol{\tau}] > 0 \qquad \text{for all } \boldsymbol{\tau} \perp \boldsymbol{\nu},\ \boldsymbol{\tau} \neq \boldsymbol{0}$$

fails, see for instance [18, 40]. In this case, the evolution equation (4) is backward parabolic, and to overcome the ill-posedness of the problem we consider the following singular perturbation of the surface energy

$$\int_{\Gamma_h} \Big(\psi(\boldsymbol{\nu}) + \frac{\varepsilon}{p}|H|^p\Big)\, d\mathcal{H}^2,$$

where $p > 2$, H stands for the sum $\kappa_1 + \kappa_2$ of the principal curvatures of Γ_h, and ε is a small positive constant (see [18, 31, 32]). The restriction $p > 2$ in $\mathbb{R}^3$ is motivated by the fact that the profile h of the film will belong to $W^{2,p}(Q)$, where $Q \subset \mathbb{R}^2$, so that $W^{2,p}(Q)$ is continuously embedded into $C^{1,\frac{p-2}{p}}(Q)$. This regularity is strongly used to prove existence of solutions. In contrast, in $\mathbb{R}^2$ we can assume $p \geq 2$ since $W^{2,2}((0,b))$ is embedded in $C^{1,1}([0,b])$.

The regularized free-energy functional becomes

$$\int_{\Omega_h} W\big(\boldsymbol{E}(\boldsymbol{u})\big)\, dx + \int_{\Gamma_h} \Big(\psi(\boldsymbol{\nu}) + \frac{\varepsilon}{p}|H|^p\Big)\, d\mathcal{H}^2, \tag{5}$$

and (3) is replaced by

$$\begin{aligned} V = \Delta_\Gamma\Big[&\mathrm{div}_\Gamma\big(D\psi(\boldsymbol{\nu})\big) + W\big(\boldsymbol{E}(\boldsymbol{u})\big) \\ &- \varepsilon\Big(\Delta_\Gamma(|H|^{p-2}H) - |H|^{p-2}H\Big(\kappa_1^2 + \kappa_2^2 - \frac{1}{p}H^2\Big)\Big)\Big]. \end{aligned} \tag{6}$$

Coupling this evolution equation on the profile of the film with the elastic equilibrium elliptic system holding in the film, and parametrizing Γ using $h : \mathbb{R}^2 \times [0,T_0] \to (0,\infty)$, we obtain the following Cauchy system of equations with initial and natural boundary conditions:

$$\begin{cases} \dfrac{1}{J}\dfrac{\partial h}{\partial t} = \Delta_\Gamma\Big[\mathrm{div}_\Gamma\big(D\psi(\boldsymbol{\nu})\big) + W\big(\boldsymbol{E}(\boldsymbol{u})\big) \\ \qquad -\varepsilon\Big(\Delta_\Gamma\big(|H|^{p-2}H\big) - |H|^{p-2}H\Big(\kappa_1^2 + \kappa_2^2 - \dfrac{1}{p}H^2\Big)\Big)\Big] \\ \qquad\qquad \text{in } \mathbb{R}^2 \times (0,T_0), \\ \mathrm{div}\,\mathbb{C}\boldsymbol{E}(\boldsymbol{u}) = 0 \ \text{ in } \Omega_h, \\ \mathbb{C}\boldsymbol{E}(\boldsymbol{u})[\boldsymbol{\nu}] = 0 \ \text{ on } \Gamma_h, \quad \boldsymbol{u}(x,y,0,t) = \boldsymbol{A}_0(x,y,0), \\ h(\cdot,t) \ \text{ and } \ D\boldsymbol{u}(\cdot,t) \ \text{ are } Q\text{-periodic}, \\ h(\cdot,0) = h_0, \end{cases} \tag{7}$$

where $J := \sqrt{1+|Dh|^2}$ and $h_0 \in H^2_{\mathrm{loc}}(\mathbb{R}^2)$ is a Q-periodic function.

One can find in the literature sixth-order evolution equations of this type (see, e.g., [31] for the case without elasticity, see [40] for the evolution of voids in elastically stressed materials, and [7, 39]).

We use the gradient flow structure of (7) with respect to a suitable H^{-1}-metric (see, e.g., [8]) to solve the equation via a *minimizing movement scheme* (see [1]), i.e., we discretize the problem in time and solve suitable minimum incremental problems.

If instead of H^{-1} we used the gradient flow with respect to an L^2-metric, we would obtain a fourth order evolution equation describing motion by evaporation-condensation (see [8, 31, 37]).

The short time existence of solutions to (7) established in [22] is the first such result for geometric surface diffusion equations with elasticity in three-dimensions. In the recent paper [28] (see also [27] for the two-dimensional case), the authors proved short-time existence of a smooth solution without the additional curvature regularization. They also showed asymptotic stability of strictly stable stationary sets.

The results summarized here can be found in the papers [20, 21, 22].

1 2D quasistatic equilibrium of epitaxially strained elastic films

In the following sections we assume self-similarity with respect to a planar axis and reduce the context to a two-dimensional framework. To be precise, we suppose that the material fills the infinite strip

$$\Omega_h := \{\boldsymbol{x} = (x,y) : 0 < x < b,\ y < h(x)\}, \tag{8}$$

where $h : [0,b] \to [0,\infty)$ is a Lipschitz function representing the *free* profile of the film, which occupies the open set

$$\Omega_h^+ := \Omega_h \cap \{y > 0\}. \tag{9}$$

The line $y = 0$ corresponds to the film/substrate interface.

We assume that the mismatch strain corresponding to different natural states of the material in the substrate and in the film, respectively, is represented by

$$\boldsymbol{E}_0(y) = \begin{cases} \hat{\boldsymbol{E}}_0 & \text{if } y \geq 0, \\ 0 & \text{if } y < 0, \end{cases} \tag{10}$$

with $\hat{\boldsymbol{E}}_0 \neq \mathbf{0} > 0$. We will suppose that the film and the substrate share material properties, with homogeneous elasticity positive definite fourth-order tensor $\mathbb{C}$. Hence, bearing in mind the mismatch, the elastic energy per unit area is given by $W(\boldsymbol{E} - \boldsymbol{E}_0(y))$, where

$$W(\boldsymbol{E}) := \frac{1}{2}\boldsymbol{E} \cdot \mathbb{C}[\boldsymbol{E}] \tag{11}$$

for all symmetric matrices $\boldsymbol{E} \neq \mathbf{0}$.

In turn, the interfacial energy density ψ has a step discontinuity at $y = 0$, i.e.,

$$\psi(y) := \begin{cases} \gamma_{\text{film}} & \text{if } y > 0, \\ \gamma_{\text{sub}} & \text{if } y = 0, \end{cases} \tag{12}$$

where the property

$$\gamma_{\text{sub}} \geq \gamma_{\text{film}} > 0 \tag{13}$$

will favor the SK growth mode over the VW mode. For the case $\gamma_{\text{sub}} < \gamma_{\text{film}}$, and for different crystalline materials stress tensors $\mathbb{C}$ for the substrate and for the film, we refer to [15, 16].

The total energy of the system is given by

$$\mathcal{F}(\boldsymbol{u},h) := \int_{\Omega_h} W(\boldsymbol{E}(\boldsymbol{u}) - \boldsymbol{E}_0)\,d\boldsymbol{x} + \int_{\Gamma_h} \psi\,ds, \tag{14}$$

where Γ_h represents the free surface of the film, that is,

$$\Gamma_h := \partial\Omega_h \cap ((0,b) \times \mathbb{R}). \tag{15}$$

Since the functional $\mathcal{F}$ is not lower semicontinuous, and thus, in general, does not admit minimizers, we are led to study its relaxation. Let

$$X := \Big\{(\boldsymbol{u},h) : h : [0,b] \to [0,\infty) \text{ Lipschitz}, \quad \int_0^b h\,dx = d, \quad \boldsymbol{u} \in H^1_{\text{loc}}(\Omega_h;\mathbb{R}^2)\Big\}$$

and

$$X_0 = \Big\{(\boldsymbol{u},h) : h : [0,b] \to [0,\infty) \text{ lower semicontinuous}, \quad \text{var}_{[0,b]}\,h < \infty, \quad \int_0^b h\,dx = d, \quad \boldsymbol{u} \in H^1_{\text{loc}}(\Omega_h;\mathbb{R}^2)\Big\},$$

where $\text{var}_{[0,b]}\,h$ stands for the pointwise variation of the function h. Note that length Γ_h coincides with the pointwise variation of the function $x \in [0,b] \mapsto (x,h(x))$, and so

$$\text{var}_{[0,b]}\,h \leq \text{length}\,\Gamma_h \leq b + \text{var}_{[0,b]}\,h. \tag{16}$$

For $(\boldsymbol{u},h) \in X_0$ define

$$\mathcal{G}(\boldsymbol{u},h) := \int_{\Omega_h} W(\boldsymbol{E}(\boldsymbol{u})(\boldsymbol{x}) - \boldsymbol{E}_0(y))\,d\boldsymbol{x} + \gamma_{\text{film}}\,\text{length}\,\Gamma_h. \tag{17}$$

Theorem 1 (Existence). *The following equalities hold:*

$$\inf_{(\boldsymbol{u},h)\in X} \mathcal{F}(\boldsymbol{u},h) = \inf_{(\boldsymbol{u},h)\in X} \mathcal{G}(\boldsymbol{u},h) = \min_{(\boldsymbol{u},h)\in X_0} \mathcal{G}(\boldsymbol{u},h).$$

We refer to [20] for a proof.

Next we study regularity properties of minimizers of $\mathcal{G}$ in X_0. As customary in constrained variational problems, in order to have more flexibility in the choice of test functions, we prove that the volume constraint $\int_0^b h(x)\,dx = d$ can be replaced by a volume penalization.

Theorem 2 (Volume penalization). *Let $(\boldsymbol{u}_0,h_0) \in X_0$ be a minimizer of the functional $\mathcal{G}$ defined in* (17) *with $\int_0^b h_0(x)\,dx = d$. Then there exists $k_0 \in \mathbb{N}$ such that for every integer $k \geq k_0$, $(\boldsymbol{u}_0,h_0)$ is a minimizer of the penalized functional*

$$\mathcal{G}_k(\boldsymbol{u},h) := \int_{\Omega_h} W(\boldsymbol{E}(\boldsymbol{u}) - \boldsymbol{E}_0)\,d\boldsymbol{x} + \gamma_{\text{film}}\,\text{length}\,\Gamma_h + k\left|\int_0^b h\,dx - d\right| \tag{18}$$

over all $(\boldsymbol{u},h) \in X_0$.

Proof. An argument similar to that of the proof of Theorem 1 guarantees that for every $k \in \mathbb{N}$ there exists a minimizer $(\boldsymbol{v}_k,f_k)$ of $\mathcal{G}_k$. If $\int_0^b f_k\,dx = d$ for all k sufficiently large, then

$$\mathcal{G}(\boldsymbol{u}_0,h_0) \leq \mathcal{G}(\boldsymbol{v}_k,f_k) = \mathcal{G}_k(\boldsymbol{v}_k,f_k) \leq \mathcal{G}_k(\boldsymbol{u}_0,h_0) = \mathcal{G}(\boldsymbol{u}_0,h_0) < \infty,$$

and so $(\boldsymbol{u}_0,h_0)$ is a minimizer of $\mathcal{G}_k$.

Assume now that there is a subsequence, not relabeled, such that $\int_0^b f_k\,dx \neq d$ for all k. If

$$\int_0^b f_k\,dx > d \tag{19}$$

for countably many k, define

$$h_k := \min\{f_k, t_k\},$$

where $t_k > 0$ has been chosen so that $\int_0^b h_k\,dx = d$. Note that length $\Gamma_{h_k} \le$ length Γ_{f_k}. Indeed, for every partition $x_0 = 0 < \cdots < x_n = b$, we have that

$$\left(h_k(x_i) - h_k(x_{i-1})\right)^2 \le \left(f_k(x_i) - f_k(x_{i-1})\right)^2$$

for all $i = 1, \dots, n$. Hence,

$$\mathcal{G}(\boldsymbol{v}_k, h_k) = \mathcal{G}_k(\boldsymbol{v}_k, h_k) < \mathcal{G}_k(\boldsymbol{v}_k, f_k),$$

which is a contradiction. Therefore, for all k sufficiently large

$$\int_0^b f_k\,dx < d.$$

Since

$$\mathcal{G}_k(\boldsymbol{v}_k, f_k) \le \mathcal{G}_k(\boldsymbol{u}_0, h_0) = \mathcal{G}(\boldsymbol{u}_0, h_0) < \infty, \tag{20}$$

it follows from (18) and (20) that $\int_0^b f_k\,dx \to d$ as $k \to \infty$ and that $\sup_k$ length $\Gamma_{f_k} < \infty$. In turn, by (16), $\|f_k\|_\infty \le c$ for some constant c independent of k.

Let k_1 be so large that $\int_0^b f_k\,dx > \frac{d}{2}$ for all $k \ge k_1$. Then

$$t_k := \frac{d}{\int_0^b f_k\,dx} \in (0,2)$$

and the function $h_k(x) := t_k f_k(x)$, $x \in (0,b)$, satisfies

$$\int_0^b h_k\,dx = d.$$

Consider a partition $0 = x_0 < \cdots < x_\ell = b$. Then

$$\begin{aligned}
&\sum_{i=1}^{\ell} \sqrt{(x_i - x_{i-1})^2 + \left(h_k(x_i) - h_k(x_{i-1})\right)^2} \\
&\qquad = \sum_{i=1}^{\ell} \sqrt{(x_i - x_{i-1})^2 + t_k^2\left(f_k(x_i) - f_k(x_{i-1})\right)^2} \\
&\qquad \le t_k \sum_{i=1}^{\ell} \sqrt{(x_i - x_{i-1})^2 + \left(f_k(x_i) - f_k(x_{i-1})\right)^2} \\
&\qquad \le t_k \operatorname{length} \Gamma_{f_k},
\end{aligned}$$

where we used the fact that $t_k > 1$. Hence,

$$\operatorname{length} \Gamma_{h_k} \le t_k \operatorname{length} \Gamma_{f_k},$$

and so, by (20),

$$\begin{aligned}
&\gamma_{\text{film}} \operatorname{length} \Gamma_{h_k} - \gamma_{\text{film}} \operatorname{length} \Gamma_{f_k} \\
&\qquad \le (t_k - 1)\gamma_{\text{film}} \operatorname{length} \Gamma_{f_k} \le (t_k - 1)\mathcal{G}_k(\boldsymbol{v}_k, f_k) \\
&\qquad \le (t_k - 1)\mathcal{G}(\boldsymbol{u}_0, h_0).
\end{aligned}$$

We deduce that

$$\gamma_{\text{film}} \operatorname{length} \Gamma_{h_k} \le \gamma_{\text{film}} \operatorname{length} \Gamma_{f_k} + (t_k - 1)\mathcal{G}(\boldsymbol{u}_0, h_0). \tag{21}$$

For $(x, y') \in \Omega_{h_k}$ define

$$\boldsymbol{w}_k(x, y') := \left((\boldsymbol{v}_k)_1\left(x, \frac{y'}{t_k}\right), \frac{1}{t_k}(\boldsymbol{v}_k)_2\left(x, \frac{y'}{t_k}\right)\right).$$

By a change of variables and (10), we have

$$\begin{aligned}
&\int_{\Omega_{h_k}} W\left(\boldsymbol{E}(\boldsymbol{w}_k)(x, y') - \boldsymbol{E}_0(y')\right)\,dx dy' \\
&\qquad = \frac{1}{t_k} \int_{\Omega_{f_k}} W\left(\tilde{\boldsymbol{E}}(\boldsymbol{v}_k)(x) - \boldsymbol{E}_0(y)\right)\,dx,
\end{aligned}$$

where $\tilde{\boldsymbol{E}}(\boldsymbol{v}_k)(x)$ is the 2×2 matrix whose entries are

$$\begin{aligned}
\tilde{\boldsymbol{E}}_{11}(\boldsymbol{v}_k)(x) &= \boldsymbol{E}_{11}(\boldsymbol{v}_k)(x), \qquad \tilde{\boldsymbol{E}}_{12}(\boldsymbol{v}_k)(x) = \frac{1}{t_k}\boldsymbol{E}_{12}(\boldsymbol{v}_k)(x), \\
\tilde{\boldsymbol{E}}_{22}(\boldsymbol{v}_k)(x) &= \frac{1}{t_k^2}\boldsymbol{E}_{22}(\boldsymbol{v}_k)(x).
\end{aligned} \tag{22}$$

Observe that

$$\begin{aligned}
&\left(|\tilde{\mathbf{E}}(\boldsymbol{v}_k) - \boldsymbol{E}_0| + |\boldsymbol{E}(\boldsymbol{v}_k) - \boldsymbol{E}_0|\right)|\tilde{\mathbf{E}}(\boldsymbol{v}_k) - \boldsymbol{E}(\boldsymbol{v}_k)| \\
&\qquad \le c(t_k - 1)\left(|\tilde{\mathbf{E}}(\boldsymbol{v}_k) - \boldsymbol{E}_0| + |\boldsymbol{E}(\boldsymbol{v}_k) - \boldsymbol{E}_0|\right)|\boldsymbol{E}(\boldsymbol{v}_k)| \\
&\qquad \le c(t_k - 1)\left(|\boldsymbol{E}(\boldsymbol{v}_k)| + |\boldsymbol{E}_0|\right)\left(|\boldsymbol{E}(\boldsymbol{v}_k) - \boldsymbol{E}_0| + |\boldsymbol{E}_0|\right) \\
&\qquad \le c(t_k - 1)\left(|\boldsymbol{E}(\boldsymbol{v}_k) - \boldsymbol{E}_0| + |\boldsymbol{E}_0|\right)^2.
\end{aligned} \tag{23}$$

Since $W(\boldsymbol{E})$ is a positive definite quadratic form over the 2×2 symmetric matrices (see (11)), we have that

$$|W(\boldsymbol{E}) - W(\boldsymbol{E}_1)| \le c(|\boldsymbol{E}| + |\boldsymbol{E}_1|)|\boldsymbol{E} - \boldsymbol{E}_1|$$

for all 2×2 symmetric matrices $\boldsymbol{E}$ and $\boldsymbol{E}_1$. Hence by (1), (10) and (23)

$$\begin{aligned}
&\int_{\Omega_{h_k}} W\left(\boldsymbol{E}(\boldsymbol{w}_k)(x, y') - \boldsymbol{E}_0(y')\right)\,dx' - \int_{\Omega_{f_k}} W\left(\boldsymbol{E}(\boldsymbol{v}_k)(x) - \boldsymbol{E}_0(y)\right)\,dx \\
&\quad = \frac{1}{t_k} \int_{\Omega_{f_k}} \left[W\left(\tilde{\mathbf{E}}(\boldsymbol{v}_k)(x) - \boldsymbol{E}_0(y)\right) - W\left(\boldsymbol{E}(\boldsymbol{v}_k)(x) - \boldsymbol{E}_0(y)\right)\right]\,dx \\
&\quad \le c \int_{\Omega_{f_k}} \left(|\tilde{\mathbf{E}}(\boldsymbol{v}_k) - \boldsymbol{E}_0| + |\boldsymbol{E}(\boldsymbol{v}_k) - \boldsymbol{E}_0|\right)\left(|\tilde{\mathbf{E}}(\boldsymbol{v}_k) - \boldsymbol{E}(\boldsymbol{v}_k)|\right)\,dx \\
&\quad \le c(t_k - 1) \int_{\Omega_{f_k}} \left(|\boldsymbol{E}(\boldsymbol{v}_k) - \boldsymbol{E}_0| + |\boldsymbol{E}_0|\right)^2\,dx \\
&\quad \le c(t_k - 1)\left(\mathcal{G}_k(\boldsymbol{v}_k, f_k) + |\hat{\boldsymbol{E}}_0|^2\right) \le c(t_k - 1)\left(\mathcal{G}(\boldsymbol{u}_0, h_0) + |\hat{\boldsymbol{E}}_0|^2\right),
\end{aligned} \tag{24}$$

where c depends only on the ellipticity constants of W and $\sup_k \|f_k\|_\infty$. By (20), (21), and (24), we have that

$$\begin{aligned}
&\mathcal{G}(\boldsymbol{u}_0, h_0) \\
&\quad \le \mathcal{G}(\boldsymbol{w}_k, h_k) \le \mathcal{G}(\boldsymbol{v}_k, f_k) + (t_k - 1)\left[(c+1)\mathcal{G}(\boldsymbol{u}_0, h_0) + c|\hat{\boldsymbol{E}}_0|^2\right] \\
&\quad = \mathcal{G}_k(\boldsymbol{v}_k, f_k) + (t_k - 1)\left[(c+1)\mathcal{G}(\boldsymbol{u}_0, h_0) + c|\hat{\boldsymbol{E}}_0|^2\right] \\
&\qquad\qquad - k\left(d - \int_0^b f_k\,dx\right) \\
&\quad = \mathcal{G}_k(\boldsymbol{v}_k, f_k) + (t_k - 1)\left[(c+1)\mathcal{G}(\boldsymbol{u}_0, h_0) + c|\hat{\boldsymbol{E}}_0|^2\right] \\
&\qquad\qquad - (t_k - 1)k \int_0^b f_k\,dx \\
&\quad \le \mathcal{G}(\boldsymbol{u}_0, h_0) + (t_k - 1)\left[(c+1)\mathcal{G}(\boldsymbol{u}_0, h_0) + c|\hat{\boldsymbol{E}}_0|^2 - k\frac{d}{2}\right].
\end{aligned}$$

Thus, if

$$k \ge \frac{2}{d}\left[(c+1)\mathcal{G}(\boldsymbol{u}_0, h_0) + c|\hat{\boldsymbol{E}}_0|^2\right] + 1,$$

we get a contradiction, and this completes the proof. ∎

To prove the regularity of the free boundary we use the following internal sphere condition.

Theorem 3 (Internal Sphere's Condition). *Let* $(\boldsymbol{u}_0, h_0) \in X_0$ *be a minimizer of the functional* $\mathcal{G}$ *defined in* (17). *Then there exists* $r_0 > 0$ *with the property that for every* $\boldsymbol{z}_0 \in \overline{\Gamma}_{h_0}$ *there exists an open ball* $B(\boldsymbol{x}_0, r_0)$*, with* $B(\boldsymbol{x}_0, r_0) \cap ((0,b) \times \mathbb{R}) \subseteq \Omega_{h_0}$*, such that*

$$\partial B(\boldsymbol{x}_0, r_0) \cap \overline{\Gamma}_{h_0} = \{\boldsymbol{z}_0\}.$$

This result was first proved in a slightly different context by Chambolle and Larsen [11] (see also [9, 20]). The argument is entirely two-dimensional and its extension to three dimensions is open.

Remark 4. Note that if $\boldsymbol{\nu}_0 \in \partial B(\boldsymbol{0}, 1)$ is the outward unit normal to $B(\boldsymbol{x}_0, r_0)$ at $\boldsymbol{z}_0$, then $\boldsymbol{x}_0 = \boldsymbol{z}_0 - r_0 \boldsymbol{\nu}_0$. Thus, the set

$$N_{\boldsymbol{z}_0} := \left\{\boldsymbol{\nu} \in \partial B(\boldsymbol{0}, 1) : B(\boldsymbol{z}_0 - r_0 \boldsymbol{\nu}, r_0) \cap \left((0,b) \times \mathbb{R}\right) \subseteq \Omega_{h_0}\right\} \quad (25)$$

is nonempty.

In the next theorem we prove that h_0 admits a left and right derivative at all but countably many points.

Theorem 5 (Left and Right Derivatives of h). *Let* $(\boldsymbol{u}_0, h_0) \in X_0$ *be a minimizer of the functional* $\mathcal{G}$ *defined in* (17). *Then* $\overline{\Gamma}_{h_0}$ *admits a left and a right tangent at every point* $\boldsymbol{z}$ *not of the form* $\boldsymbol{z} = (x, h_0(x))$ *with* $x \in S$*, where*

$$S := \left\{x \in (0,b) : h_0(x) < \liminf_{t \to x} h_0(t)\right\}. \qquad (26)$$

Define

$$\Gamma_{\text{cusps}} := \left\{\boldsymbol{z} \in \overline{\Gamma}_{h_0} : \pm \boldsymbol{e}_1 \in N_{\boldsymbol{z}}\right\} \qquad (27)$$

and

$$\Gamma_{\text{cuts}} := \left\{(x,y) : x \in (0,b) \cap S,\ h_0(x) \le y \le \liminf_{t \to x} h_0(t)\right\}, \quad (28)$$

where $N_{\boldsymbol{z}}$ is the set defined in (25) and S is the set defined in (26).

Theorem 6 (Cusps and Cuts). *Let* $(\boldsymbol{u}_0, h_0) \in X_0$ *be a minimizer of the functional* $\mathcal{G}$ *defined in* (17). *Then the sets* Γ_{cusps} *and* Γ_{cuts} *contain at most finitely many vertical segments.*

Remark 7. If $-\boldsymbol{e}_1 \in N_{\boldsymbol{z}_0}$, then since $B((x_0 + r_0, y_0), r_0) \cap ((0,b) \times \mathbb{R}) \subseteq \Omega_{h_0}$ and h_0 is lower semicontinuous, for all $x > x_0$ sufficiently close to x_0, we have that

$$h_0(x) \ge y_0 + \sqrt{r_0^2 - \left(x - (x_0 + r_0)\right)^2},$$

and so

$$\frac{h_0(x) - y_0}{x - x_0} \ge \frac{\sqrt{2r_0 - (x - x_0)}}{\sqrt{x - x_0}} \to \infty$$

as $x \to x_0^+$. By Theorem 5 it follows that $\overline{\Gamma}_{h_0}$ admits a right vertical tangent at $\boldsymbol{z}_0$. Similarly, if $\boldsymbol{e}_1 \in N_{\boldsymbol{z}_0}$ then for $x < x_0$, $\overline{\Gamma}_{h_0}$ admits a left vertical tangent at $\boldsymbol{z}_0$. In particular, if $\pm \boldsymbol{e}_1 \in N_{\boldsymbol{z}_0}$ and h_0 is continuous at x_0, then

$$(h_0)'_-(x_0) = -\infty, \quad (h_0)'_+(x_0) = \infty. \qquad (29)$$

The next theorem shows that, except for cut and cusp points, $\overline{\Gamma}_{h_0}$ is locally Lipschitz.

Theorem 8. *Let* $(\boldsymbol{u}_0, h_0) \in X_0$ *be a minimizer of the functional* $\mathcal{G}$ *defined in* (17). *If* $\boldsymbol{z}_0 \in \overline{\Gamma}_{h_0} \setminus (\Gamma_{\text{cuts}} \cup \Gamma_{\text{cusps}})$*, then* $\overline{\Gamma}_{h_0}$ *is Lipschitz in a neighborhood of* $\boldsymbol{z}_0$.

In order to improve the regularity results for h, we restrict our attention to the linearly isotropic case in which

$$W(\boldsymbol{E}) = \frac{1}{2}\lambda\left[\operatorname{tr}(\boldsymbol{E})\right]^2 + \mu \operatorname{tr}(\boldsymbol{E}^2), \qquad (30)$$

where λ and μ are the (constant) Lamé moduli with

$$\mu > 0, \quad \mu + \lambda > 0. \qquad (31)$$

Note that in this range, the quadratic form W is coercive. We also assume that the matrix $\hat{\boldsymbol{E}}_0$ in (10) takes the form

$$\hat{\boldsymbol{E}}_0 = \begin{pmatrix} e_0 & 0 \\ 0 & 0 \end{pmatrix} \qquad (32)$$

for some $e_0 > 0$, which measures the mismatch between the lattices of the two materials.

Since h_0 is now Lipschitz with left and right derivatives at all but, at most, a finite number of points, we can now obtain classical decay estimates for the solution $\boldsymbol{u}_0$. In turn, these will exclude corners in the graph Γ_{h_0} of h_0.

Theorem 9 (Decay Estimate). *Assume* (30) *and* (32). *Let* $(\boldsymbol{u}_0, h_0) \in X_0$ *be a minimizer of the functional* $\mathcal{G}$ *defined in* (17). *Suppose that* $\overline{\Gamma}_{h_0}$ *has a corner at some point* $\boldsymbol{z}_0 \in \overline{\Gamma}_{h_0} \setminus (\Gamma_{\text{cusps}} \cup \Gamma_{\text{cuts}})$. *Then there exist a constant* $c > 0$*, a radius* r_0*, and an exponent* $\frac{1}{2} < \alpha < 1$ *such that*

$$\int_{B(\boldsymbol{z}_0, r) \cap \Omega_{h_0}} |\nabla \boldsymbol{u}_0|^2 \, d\boldsymbol{x} \le c r^{2\alpha} \qquad (33)$$

for all $0 < r < r_0$.

Using the previous decay estimate, it can be shown that for $(\boldsymbol{u}_0, \Omega) \in X$ the upper boundary $\overline{\Gamma}_{h_0}$ is of class C^1 away from $\Gamma_{\text{cusps}} \cup \Gamma_{\text{cuts}}$.

Theorem 10 (C^∞ Regularity of Γ). *Assume* (30) *and* (32). *Let* $(\boldsymbol{u}_0, h_0) \in X_0$ *be a minimizer of the functional* $\mathcal{G}$ *defined in* (17). *Then* $\overline{\Gamma}_{h_0} \setminus (\Gamma_{\text{cusps}} \cup \Gamma_{\text{cuts}})$ *is of class* C^1.

Theorem 10 can be significantly improved. Indeed, using another blow-up argument it is possible to show that $\overline{\Gamma}_{h_0} \setminus (\Gamma_{\text{cusps}} \cup \Gamma_{\text{cuts}})$ is of class $C^{1,\alpha}$ for all $0 < \alpha < \frac{1}{2}$. In turn, this implies that $\boldsymbol{u}_0$ is of class $C^{1,\beta}$ for some $\beta > 0$ away from the x-axis and from $\Gamma_{\text{cusps}} \cup \Gamma_{\text{cuts}}$. By a classical bootstrap argument, one can obtain C^∞ regularity and then use results of [34] by Koch, Morini and the second author to prove analyticity of $\overline{\Gamma}_{h_0} \setminus (\Gamma_{\text{cusps}} \cup \Gamma_{\text{cuts}})$ away from the x-axis. We refer to [20] for more details.

2 Evolution of epitaxially strained elastic films: The 2D case

The evolution of epitaxially strained elastic films depends strongly on the possible anisotropy of the surface energy density. For this reason, in (17) we replace the isotropic surface energy γ_{film} length Γ_h by

$$\int_{\Gamma_h} \psi(\boldsymbol{\nu})\, d\mathcal{H}^1,$$

where $\psi : \mathbb{R}^2 \to [0, \infty)$ is a positively one-homogeneous function of class C^2 away from the origin. Also, the mismatch between the substrate and film crystalline structures is represented by the Dirichlet condition (see (32))

$$\boldsymbol{u}(x,0) = (e_0 x, 0) \quad \text{for all } x \in (0,b).$$

As discussed in the introduction, strong anisotropy of ψ may lead to the ill-posedness of the evolution law, and thus we add a higher order regularizing term. To be precise, for $\varepsilon > 0$ small the energy under study becomes

$$\mathcal{J}(\boldsymbol{u},h) := \int_{\Omega_h} W\big(\boldsymbol{E}(\boldsymbol{u})\big)\, d\boldsymbol{x} + \int_{\Gamma_h} \Big(\psi(\boldsymbol{\nu}) + \frac{\varepsilon}{2}k^2\Big)\, d\mathcal{H}^1, \tag{34}$$

where k denotes the curvature of Γ_h and $\boldsymbol{\nu}$ is the outer unit normal to Ω_h.

We consider periodicity conditions. Hence, given a positive b-periodic function $h : \mathbb{R} \to [0, +\infty)$, with locally finite pointwise variation, we set

$$\Omega_h^\# := \{\boldsymbol{x} = (x,y) : x \in \mathbb{R}\,, 0 < y < h(x)\}\,,$$

and

$$\Gamma_h^\# := \{\boldsymbol{x} = (x,y) : x \in \mathbb{R}\,,\ y = h(x)\}\,.$$

Given $h \in W^{2,2}_\sharp((0,b);\mathbb{R}^2)$, where $W^{2,2}_\sharp((0,b);\mathbb{R}^2)$ is the space of b periodic functions in $W^{2,2}_{\text{loc}}(\mathbb{R};\mathbb{R}^2)$, we denote

$$X_\#(\Omega_h;\mathbb{R}^2) := \Big\{\boldsymbol{u} \in L^2_{\text{loc}}(\Omega_h^\#;\mathbb{R}^2) : \boldsymbol{u}(x,y) = \boldsymbol{u}(x+b,y) \text{ for } (x,y) \in \Omega_h^\#\,, \boldsymbol{E}(\boldsymbol{u})|_{\Omega_h} \in L^2(\Omega_h;\mathbb{R}^2)\Big\}$$

and

$$X_{e_0} := \Big\{(\boldsymbol{u},h) : h \in W^{2,2}_\sharp((0,b);\mathbb{R}^2),\ \boldsymbol{u} \in e_0(\cdot,0) + LD_\#(\Omega_h;\mathbb{R}^2)\,, \text{ and } \boldsymbol{u}(x,0) = (e_0 x, 0) \text{ for all } x \in \mathbb{R}\Big\}\,.$$

We next introduce the incremental minimum problems used to define the discrete time evolutions. This will lead to the existence of solutions for the evolution equation (40) below via minimizing movements. Let $(\boldsymbol{u}_0, h_0) \in X_{e_0}$ be such that

$$h_0 > 0, \tag{35}$$

and $\boldsymbol{u}_0$ minimizes the elastic energy in Ω_{h_0} among all $\boldsymbol{u}$ with $(\boldsymbol{u}, h_0) \in X_{e_0}$. Given $T > 0$, $N \in \mathbb{N}$, we set $\Delta T := \frac{T}{N}$. For $i = 1, \dots, N$, we define inductively $(\boldsymbol{u}_{i,N}, h_{i,N})$ as a solution of the minimum problem

$$\begin{aligned}\min\Big\{&\mathcal{J}(\boldsymbol{u},h) + \frac{1}{2\Delta T}\int_{\Gamma_{h_{i-1,N}}} \Big(\int_0^x \big(h(\zeta) - h_{i-1,N}(\zeta)\big)\, d\zeta\Big)^2 d\mathcal{H}^1(x,y) : \\ &(\boldsymbol{u},h) \in X_{e_0}, \quad \|h'\|_\infty \le \Lambda_0, \quad \int_0^b h\, dx = \int_0^b h_0\, dx\,, \\ &\int_{\Gamma_{h_{i-1,N}}} \int_0^x \big(h(\zeta) - h_{i-1,N}(\zeta)\big)\, d\zeta d\mathcal{H}^1(x,y) = 0\Big\},\end{aligned} \tag{36}$$

where $\|h_0'\|_\infty < \Lambda_0$.

Then for $x \in \mathbb{R}$ and $(i-1)\Delta T \le t \le i\Delta T$, $i = 1, \dots, N$, we set

$$h_N(x,t) := h_{i-1,N}(x) + \frac{1}{\Delta T}\big(t - (i-1)\Delta T\big)\big(h_{i,N}(x) - h_{i-1,N}(x)\big), \tag{37}$$

and we let $\boldsymbol{u}_N(\cdot,t)$ be the *elastic equilibrium corresponding to $h_N(\cdot,t)$, i.e., the minimizer of the elastic energy in $\Omega_{h_N(\cdot,t)}$ among all $\boldsymbol{u}$ such that $(h_N(\cdot,t), \boldsymbol{u}) \in X_{e_0}$.*

We remark the incremental minimum problem can be written as

$$\begin{aligned}\min\Big\{&\mathcal{J}(\boldsymbol{u},h) + \frac{1}{2\Delta T}\Big\|\frac{h - h_{i-1,N}}{\sqrt{1 + h'^2_{i-1,N}}}\Big\|^2_{H^{-1}(\Gamma_{i-1,N})} : (\boldsymbol{u},h) \in X_{e_0}, \\ &\|h'\|_\infty \le \Lambda_0, \quad \int_0^b h\, dx = \int_0^b h_0\, dx, \\ &\int_{\Gamma_{h_{i-1,N}}} \int_0^x \big(h(\zeta) - h_{i-1,N}(\zeta)\big)\, d\zeta d\mathcal{H}^1(x,y) = 0\Big\}.\end{aligned}$$

We now show that the incremental minimum problem (36) admits a solution.

Theorem 11. *For every $i = 1, \dots, N$, the minimum problem* (36) *admits a solution $(\boldsymbol{u}_{i,N}, h_{i,N}) \in X_{e_0}$.*

Proof. Let $\{(\boldsymbol{u}_n, h_n)\} \subset X_{e_0}$ be a minimizing sequence for (36). Since $\int_0^b h_n\, dx = \int_0^b h_0\, dx$,

$$\sup_n \int_0^b \frac{(h_n'')^2}{\sqrt{1 + (h_n')^2}}\, dx < \infty$$

and $\|h_n'\|_\infty \le \Lambda_0$, it follows that $\|h_n\|_{W^{2,2}} \le C$ for some constant $C > 0$ and for all n. Then, up to a subsequence (not relabelled), we may assume that $h_n \rightharpoonup h$ weakly in $W^{2,2}_\sharp((0,b);\mathbb{R}^2)$, and thus strongly in $C^1(\mathbb{R};\mathbb{R}^2)$. As a consequence,

$$\int_{\Gamma_h}\Big(\psi(\boldsymbol{\nu})+\frac{\varepsilon}{2}k^2\Big)\,d\mathcal{H}^1 \le \liminf_{n\to\infty}\int_{\Gamma_{h_n}}\Big(\psi(\boldsymbol{\nu})+\frac{\varepsilon}{2}k_n^2\Big)\,d\mathcal{H}^1, \quad (38)$$

and

$$\int_{\Gamma_{h_{i-1,N}}}\Big(\int_0^x\big(h(\zeta)-h_{i-1,N}(\zeta)\big)\,d\zeta\Big)^2\,d\mathcal{H}^1 = \lim_{n\to\infty}\int_{\Gamma_{h_{i-1,N}}}\Big(\int_0^x\big(h_n(\zeta)-h_{i-1,N}(\zeta)\big)\,d\zeta\Big)^2\,d\mathcal{H}^1. \quad (39)$$

Finally, since $\sup_n\int_{\Omega_{h_n}}|\boldsymbol{E}(\boldsymbol{u}_n)|^2\,dx<\infty$, reasoning as in [20, Proposition 2.2], from the C^1 convergence of $\{h_n\}$ to h and Korn's inequality we deduce that there exists $\boldsymbol{u}\in H^1_{\mathrm{loc}}(\Omega_h^\#;\mathbb{R}^2)$ such that $(\boldsymbol{u},h)\in X_{e_0}$ and, up to a subsequence, $\boldsymbol{u}_n\rightharpoonup\boldsymbol{u}$ weakly in $H^1_{\mathrm{loc}}(\Omega_h^\#;\mathbb{R}^2)$. Therefore, we have that

$$\int_{\Omega_h}W(\boldsymbol{E}(\boldsymbol{u}))\,dx \le \liminf_{n\to\infty}\int_{\Omega_{h_n}}W(\boldsymbol{E}(\boldsymbol{u}_n))\,dx,$$

which, together with (38) and (39), allows us to conclude that $(\boldsymbol{u},h)$ is a minimizer. ∎

Next, we show that solutions of the discrete time evolution problems converge to a function $h=h(x,t)$, which is a weak solution of the following geometric evolution equation,

$$\frac{\partial h}{\partial t}=\Bigg[\frac{1}{J}\bigg(\varepsilon\Big(\frac{h_{xx}}{J^5}\Big)_{xx}+\frac{5\varepsilon}{2}\Big(\frac{h_{xx}^2}{J^7}h_x\Big)_x + \big(\psi_x(-h_x,1)\big)_x+W(\boldsymbol{E}(\boldsymbol{u}))\bigg)_x\Bigg]_x \quad (40)$$

for a short time interval $[0,T_0]$, where $0<T_0\le T$, where T_0 depends on $(\boldsymbol{u}_0,h_0)$. Here $J:=\sqrt{1+(h_x)^2}$. Since $\|h_0'\|_\infty<\Lambda_0$, for all t sufficiently small we have that $\|\frac{\partial h}{\partial t}\|_\infty<\Lambda_0$, and so we are allowed to take admissible variations of h to obtain (40).

Theorem 12. *There exist $T_0\in(0,T]$ and $C>0$ depending only $(h_0,\boldsymbol{u}_0)$ such that:*

(i) *$h_N\to h$ in $C^{0,\beta}([0,T_0];C^{1,\alpha}([0,b]))$ for every $\alpha\in(0,\frac{1}{2})$, and $\beta\in(0,(1-2\alpha)/32)$,*
(ii) *$\boldsymbol{E}(\boldsymbol{u}_N(\cdot,h_N))\to\boldsymbol{E}(\boldsymbol{u}(\cdot,h))$ in $C^{0,\beta}([0,T_0];C^{1,\alpha}([0,b]))$ for every $\alpha\in(0,\frac{1}{2})$, and $0\le\beta<(1-2\alpha)/32$, where $\boldsymbol{u}(\cdot,t)$ is the elastic equilibrium in $\Omega_{h(\cdot,t)}$,*

and h is a weak solution to (40) with initial data h_0. Moreover, if $\psi\in C^3(\mathbb{R}^2\setminus\{0\})$ then $h(\cdot,t)\in H^5_\#(0,b)$ for almost every $t\in[0,T_0]$, and h is the unique solution.

For linearly isotropic energy densities of the form (30), where λ and μ satisfy (31), and for sufficiently regular surface energy densities, we can prove asymptotic stability of the flat configuration $h_{\mathrm{flat}}\equiv d/b$ when d is sufficiently small. Consider the *Grinfeld function K* defined by

$$K(y):=\max_{n\in\mathbb{N}}\frac{1}{n}J(ny),\quad y\ge 0, \quad (41)$$

where

$$J(y):=\frac{y+(3-4\nu_p)\sinh y\cosh y}{4(1-\nu_p)^2+y^2+(3-4\nu_p)\sinh^2 y},$$

and ν_p is the *Poisson modulus* of the elastic material, i.e.,

$$\nu_p:=\frac{\lambda}{2(\lambda+\mu)}. \quad (42)$$

It turns out that K is strictly increasing and continuous, $K(y)\le Cy$, and $\lim_{y\to+\infty}K(y)=1$, for some positive constant C.

Theorem 13. *Assume that W takes the form (30), where λ and μ satisfy (31), and that $\psi\in C^3(\mathbb{R}^2\setminus\{0\})$ satisfies $\partial^2_{11}\psi(0,1)>0$ and*

$$D^2\psi(\boldsymbol{\xi})[\boldsymbol{\tau},\boldsymbol{\tau}]>0\quad \text{for all } \boldsymbol{\tau}\perp\boldsymbol{\xi},\ \boldsymbol{\tau}\ne\mathbf{0}$$

for every $\boldsymbol{\xi}\in S^1$. Let $d_{\mathrm{loc}}:(0,\infty)\to(0,\infty]$ be defined as $d_{\mathrm{loc}}(b):=\infty$ if $0<b\le\frac{\pi}{4}\frac{(2\mu+\lambda)\partial^2_{11}\psi(0,1)}{e_0^2\mu(\mu+\lambda)}$, and as the solution to

$$K\Big(\frac{2\pi d_{\mathrm{loc}}(b)}{b}\Big)=\frac{\pi}{4}\frac{(2\mu+\lambda)\partial^2_{11}\psi(0,1)}{e_0^2\mu(\mu+\lambda)}\frac{1}{b} \quad (43)$$

otherwise. Then, for all $d\in(0,d_{\mathrm{loc}}(b))$ the flat configuration $h_{\mathrm{flat}}\equiv d/b$ is asymptotically stable, that is, there exists $\delta>0$ such that if $h_0\in W^{2,2}_\sharp((0,b);\mathbb{R}^2)$ with $\int_0^b h_0\,dx=d$ and $\|h_0-h_{\mathrm{flat}}\|_{W^{2,2}}\le\delta$, then the solution h to (40) with initial datum h_0 exists for all times and

$$\|h(\cdot,t)-h_{\mathrm{flat}}\|_{W^{2,2}}\to 0$$

as $t\to\infty$.

Acknowledgements. The research of I. Fonseca was partially funded by the National Science Foundation under Grants No. DMS-1411646 and DMS-1906238, and the one of G. Leoni by DMS-1714098.

References

[1] L. Ambrosio, Minimizing movements. *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5)* **19**, 191–246 (1995)
[2] P. Bella, M. Goldman and B. Zwicknagl, Study of island formation in epitaxially strained films on unbounded domains. *Arch. Ration. Mech. Anal.* **218**, 163–217 (2015)
[3] M. Bonacini, Epitaxially strained elastic films: The case of anisotropic surface energies. *ESAIM Control Optim. Calc. Var.* **19**, 167–189 (2013)
[4] M. Bonacini, Stability of equilibrium configurations for elastic films in two and three dimensions. *Adv. Calc. Var.* **8**, 117–153 (2015)

[5] E. Bonnetier and A. Chambolle, Computing the equilibrium configuration of epitaxially strained crystalline films. *SIAM J. Appl. Math.* **62**, 1093–1121 (2002)

[6] A. Braides, A. Chambolle and M. Solci, A relaxation result for energies defined on pairs set-function and applications. *ESAIM Control Optim. Calc. Var.* **13**, 717–734 (2007)

[7] M. Burger, F. Haußer, C. Stöcker and A. Voigt, A level set approach to anisotropic flows with curvature regularization. *J. Comput. Phys.* **225**, 183–205 (2007)

[8] J. W. Cahn and J. E. Taylor, Overview no. 113, surface motion by surface diffusion. *Acta Metall. Mater.* **42**, 1045–1063 (1994)

[9] G. M. Capriani, V. Julin and G. Pisante, A quantitative second order minimality criterion for cavities in elastic bodies. *SIAM J. Math. Anal.* **45**, 1952–1991 (2013)

[10] M. Caroccia, R. Cristoferi and L. Dietrich, Equilibria configurations for epitaxial crystal growth with adatoms. *Arch. Ration. Mech. Anal.* **230**, 785–838 (2018)

[11] A. Chambolle and C. J. Larsen, C^∞ regularity of the free boundary for a two-dimensional optimal compliance problem. *Calc. Var. Partial Differential Equations* **18**, 77–94 (2003)

[12] A. Chambolle and M. Solci, Interaction of a bulk and a surface energy with a geometrical constraint. *SIAM J. Math. Anal.* **39**, 77–102 (2007)

[13] V. Crismale and M. Friedrich, Equilibrium configurations for epitaxially strained films and material voids in three-dimensional linear elasticity. *Arch. Ration. Mech. Anal.* **237**, 1041–1098 (2020)

[14] G. Dal Maso, Generalised functions of bounded deformation. *J. Eur. Math. Soc. (JEMS)* **15**, 1943–1997 (2013)

[15] E. Davoli and P. Piovano, Analytical validation of the Young–Dupré law for epitaxially-strained thin films. *Math. Models Methods Appl. Sci.* **29**, 2183–2223 (2019)

[16] E. Davoli and P. Piovano, Derivation of a heteroepitaxial thin-film model. *Interfaces Free Bound.* **22**, 1–26 (2020)

[17] B. De Maria and N. Fusco, Regularity properties of equilibrium configurations of epitaxially strained elastic films. In *Topics in modern regularity theory*, CRM Series 13, Ed. Norm., Pisa, 169–204 (2012)

[18] A. Di Carlo, M. E. Gurtin and P. Podio-Guidugli, A regularized equation for anisotropic motion-by-curvature. *SIAM J. Appl. Math.* **52**, 1111–1119 (1992)

[19] I. Fonseca, N. Fusco, G. Leoni and V. Millot, Material voids in elastic solids with anisotropic surface energies. *J. Math. Pures Appl. (9)* **96**, 591–639 (2011)

[20] I. Fonseca, N. Fusco, G. Leoni and M. Morini, Equilibrium configurations of epitaxially strained crystalline films: existence and regularity results. *Arch. Ration. Mech. Anal.* **186**, 477–537 (2007)

[21] I. Fonseca, N. Fusco, G. Leoni and M. Morini, Motion of elastic thin films by anisotropic surface diffusion with curvature regularization. *Arch. Ration. Mech. Anal.* **205**, 425–466 (2012)

[22] I. Fonseca, N. Fusco, G. Leoni and M. Morini, Motion of three-dimensional elastic films by anisotropic surface diffusion with curvature regularization. *Anal. PDE* **8**, 373–423 (2015)

[23] I. Fonseca, N. Fusco, G. Leoni and M. Morini, A model for dislocations in epitaxially strained elastic films. *J. Math. Pures Appl. (9)* **111**, 126–160 (2018)

[24] I. Fonseca, G. Leoni and M. Morini, Equilibria and dislocations in epitaxial growth. *Nonlinear Anal.* **154**, 88–121 (2017)

[25] I. Fonseca, A. Pratelli and B. Zwicknagl, Shapes of epitaxially grown quantum dots. *Arch. Ration. Mech. Anal.* **214**, 359–401 (2014)

[26] N. Fusco, Equilibrium configurations of epitaxially strained thin films. *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* **21**, 341–348 (2010)

[27] N. Fusco, V. Julin and M. Morini, The surface diffusion flow with elasticity in the plane. *Comm. Math. Phys.* **362**, 571–607 (2018)

[28] N. Fusco, V. Julin and M. Morini, The surface diffusion flow with elasticity in three dimensions. *Arch. Ration. Mech. Anal.* **237**, 1325–1382 (2020)

[29] N. Fusco and M. Morini, Equilibrium configurations of epitaxially strained elastic films: second order minimality conditions and qualitative properties of solutions. *Arch. Ration. Mech. Anal.* **203**, 247–327 (2012)

[30] M. Goldman and B. Zwicknagl, Scaling law and reduced models for epitaxially strained crystalline films. *SIAM J. Math. Anal.* **46**, 1–24 (2014)

[31] M. E. Gurtin and M. E. Jabbour, Interface evolution in three dimensions with curvature-dependent energy and surface diffusion: interface-controlled evolution, phase transitions, epitaxial growth of elastic films. *Arch. Ration. Mech. Anal.* **163**, 171–208 (2002)

[32] C. Herring, Some theorems on the free energies of crystal surfaces. *Phys. Rev.* **82**, 87 (1951)

[33] S. Y. Kholmatov and P. Piovano, A unified model for stress-driven rearrangement instabilities. *Arch. Ration. Mech. Anal.* **238**, 415–488 (2020)

[34] H. Koch, G. Leoni and M. Morini, On optimal regularity of free boundary problems and a conjecture of De Giorgi. *Comm. Pure Appl. Math.* **58**, 1051–1076 (2005)

[35] L. C. Kreutz and P. Piovano, Microscopic validation of a variational model of epitaxially strained crystalline films. *SIAM J. Math. Anal.* **53**, 453–490 (2021)

[36] R. V. Kukta and L. B. Freund, Minimum energy configuration of epitaxial material clusters on a lattice-mismatched substrate. *J. Mech. Phys. Solids* **45**, 1835–1860 (1997)

[37] P. Piovano, Evolution of elastic thin films with curvature regularization via minimizing movements. *Calc. Var. Partial Differ. Equ.* **49**, 337–367 (2014)

[38] P. Piovano and I. Velčić, Microscopical justification of solid-state wetting and dewetting. arXiv:2010.08787 (2020)

[39] A. Rätz, A. Ribalta and A. Voigt, Surface evolution of elastically stressed films under deposition by a diffuse interface model. *J. Comput. Phys.* **214**, 187–208 (2006)

[40] M. Siegel, M. J. Miksis and P. W. Voorhees, Evolution of material voids for highly anisotropic surface energy. *J. Mech. Phys. Solids* **52**, 1319–1353 (2004)

[41] B. J. Spencer, Asymptotic derivation of the glued-wetting-layer model and contact-angle condition for Stranski–Krastanow islands. *Phys. Rev. B*, **59**, 2011 (1999)

[42] B. J. Spencer and J. Tersoff, Equilibrium shapes and properties of epitaxially strained islands. *Phys. Rev. Lett.* **79**, 4858 (1997)

Irene Fonseca's main contributions have been on the variational study of ferroelectric and magnetic materials, composites, thin structures, phase transitions, and on the mathematical analysis of image segmentation, denoising, detexturing, registration and recolorization in computer vision. She continues to explore modern methods in the calculus of variations motivated by problems issuing from materials science and imaging science. She is a Fellow of the American Mathematical Society, and of the Society for Industrial and Applied Mathematics. She was SIAM President in 2013 and 2014. She is a Grand Officer of the "Ordem Militar de Sant'Iago da Espada" (a Portuguese decoration).

fonseca@andrew.cmu.edu

Giovanni Leoni is a professor of mathematics at Carnegie Mellon University, His areas of expertise are the calculus of variations, geometric measure theory, and partial differential equations with applications to engineering, materials science, and mechanics. He is the recipient of the award "Premio Giuseppe Bartolozzi" of the Italian Mathematical Society for best Italian mathematician under 34, 2001, and of the Julius Ashkin Teaching Award and of the Richard Moore Award of the Mellon College of Sciences. He wrote two books *Modern methods in the calculus of variations: L^p spaces* (with I. Fonseca) and *A first course in Sobolev spaces*.

giovanni@andrew.cmu.edu

Young mathematicians' column

The pandemic situation again forces us to write about "online everything", but we do it not only to complain; we do believe that some of the new habits are actually fruitful, and may possibly remain even when the situation returns to normal. In this issue, we will address the question of conferences and other collective events. You yourself have surely "attended" some online seminars, or given an online talk, or even organized something (either because you really wished to do this, or because you were obliged to take online an event that had been planned much earlier). And of course, you have also noticed how much easier it is than it used to be to listen to "great people" all around the world.

But here, we will not merely talk about online conferences in general: rather, we will focus on those that are organized for/with/by young scientists. This is a very important question, because while for a senior researcher this last year may have been no more than an unfortunate less-productive period, for a young postdoc one year is practically half-life. The following two texts are just a couple of examples of a great number of events that were organized during this difficult year. Some of the ideas presented may seem quite obvious when you read them; nevertheless we believe they are worth spelling out in order to share our experiences and provide motivation. The first one, the "Junior Global Poisson Workshop", concerns an online conference that also included a part specifically devoted to social interaction. The second one, "Finding paths in a totally disconnected space: collaborating across universities" is written in an informal style, since its purpose is to recount activities that normally would happen precisely in informal circumstances, such as around the blackboard at the coffee machine. As always, please enjoy reading, and feel free to share your thoughts and experiences with the editorial board.

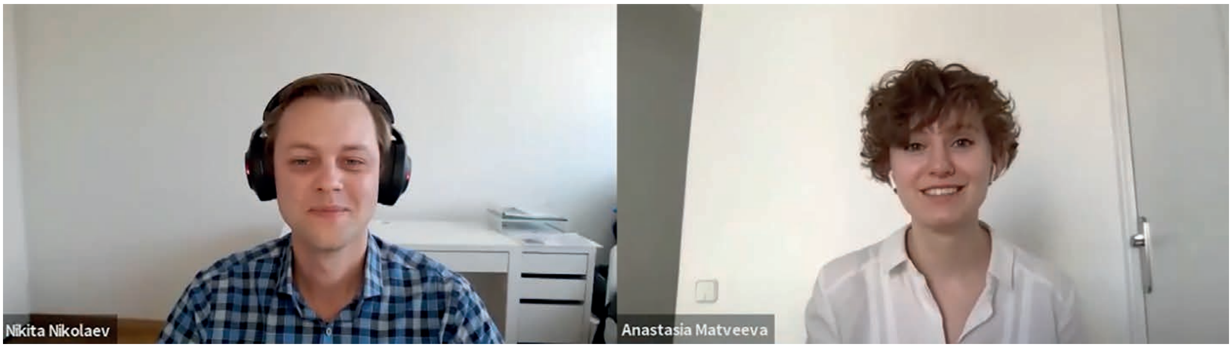
Vladimir Salnikov

Junior Global Poisson Workshops

Like all major global events, the coronavirus pandemic of 2020–2021 (let's be optimistic!) is going to have a lasting effect on our society, and in particular on the mathematical community, presenting us with new challenges but also new opportunities. Just one year ago, the mere suggestion of inviting a speaker to give a virtual seminar or organising a fully online conference might have sounded ludicrous to many of us. But, in the face of a new reality, our community stepped up and quickly adapted, and we demonstrated time and again that we can continue to interact, to collaborate and exchange ideas, and even to see each other and socialise (in many cases more often now than we ever did before!). Across multiple time zones, we helped each other unmute our microphones and connect our tablets, amused each other with silly virtual backgrounds, and embraced the healthy bring-your-children-and-pets-to-work attitude.

This new normal also presented us with an opportunity to address another old problem in our community: the struggle for many young mathematicians to get noticed. Last year, we (the authors of this letter) – encouraged by our colleague, friend, and mentor Eva Miranda, and with the financial support from the NCCR SwissMAP – created a recurring series of fully online workshops called *Junior Global Poisson Workshops* (JGPW). The aim was to have a global online venue for young mathematicians which can easily and instantly reach hundreds in the scientific community (in this case, in areas loosely connected to Poisson geometry). We thought it was extremely important to give our young colleagues the necessary space and platform to advertise their work, their ideas, and their ambition and vision. Especially, we felt such events would benefit those who are currently on the job market looking for their first or second postdoc position. This, we thought, would be their chance to attract attention to their work.

The fact that our mathematical community has so quickly embraced the new reality of virtual conferencing was the ideal condition for JGPW to be as successful as it was. Thanks to our wonderful contacts around the world who helped us advertise the workshop as widely as possible, our participants came in numbers from all corners of the globe: all in all, we welcomed almost 300 participants who tuned in from 45 separate countries, spanning a total



Nikita and Anastasia welcoming the participants on YouTube

of 14 different time zones, and representing every continent (except Antarctica: despite their piscivorous diet, sadly not a single penguin accepted our invitation). We were also delighted to see that our audience had a healthy mix of junior and senior researchers: nearly 1/3 of our participants were faculty members. Many senior researchers took a very active part in discussion sessions and even social events, causing the workshop to really bring together all generations of the Poisson geometry community whilst focusing on the achievements of some of its most junior members. Altogether, the scale and the scope that the event eventually attained was truly humbling for us, the organisers.

We pursued three main goals for each speaker. First, we wanted to give them an opportunity to give a short advertisement of their work, perhaps an overview of their problem or a brisk list of their main recent results, aimed more or less at a general member of the Poisson geometry community. Second, we also wanted to give each speaker an opportunity to explain aspects of their work in a lot more technical detail to the specialists in their respective areas. To achieve this, each speaker was asked to give one talk in two separate parts. The first part was a short uninterrupted formal presentation which was recorded and streamed live on YouTube (youtube.com/playlist?list=PLCzLB5TLzwFsineo36sCuxx1SOavHAqdn), which the speaker could later link to (for example, in a job application). The second part – which was neither recorded nor broadcast live – was an informal presentation in the style of a working group seminar, followed by a longer discussion; the idea was that this informal unrecorded setting would encourage more questions and more interaction from the participants.

The third goal – which from our point of view was perhaps the most important one – was to add to our workshop the social element which is often the root of new collaborations. We blocked out several time slots specifically dedicated to social activities, in order to let participants get acquainted and make new friends. The idea was not to simply recreate what usually happens at physical conferences, but rather to design activities adapted to the new virtual medium. We tried several different formats on a purely experimental basis, and obviously some were more successful than others. But amidst crashing into one another in Mario Kart, laughing together during Random Chats, and playing Codenames (which ultimately gained the title of the official JGPW board game!), we all had terrific fun getting to know each other.

Of course, not everything was perfect: there were emergencies and last minute changes, and altogether, as the conference went on, we realised pretty quickly that some things don’t work as well as we thought they would, whilst others work better than expected. We learnt many lessons, and we will keep adjusting, trying new ideas, and fine-tuning to make future JGP workshops even more exciting, engaging, and valuable to the mathematical community. And so there is no better way to conclude this letter than to cordially invite you, our dear reader, to participate in the next *Junior Global Poisson Workshop 2021* in early May: for details, go to our conference webpage www.unige.ch/math/folks/nikolaev/JGPW2021.html. Join us to make friends and hear interesting maths!

Anastasia Matveeva and Nikita Nikolaev

Anastasia Matveeva is a PhD student at Universitat Politécnica de Catalunya and Barcelona Graduate School of Mathematics. She is working in the field of *b*-symplectic and Poisson geometry under the supervision of Eva Miranda.

anastasiia.matveeva@upc.edu

Nikita Nikolaev has just started as a research associate at the University of Sheffield where his mentor is Tom Bridgeland. His expertise is in the theory of meromorphic connections and algebraic analysis of singular perturbation theory.

n.nikolaev@sheffield.ac.uk

Finding paths in a totally disconnected space: Collaborating across universities

Imaginary Interviewer: Alfonso and Leo, last fall you started organizing an online seminar. It is called "Good morning SFARS". What does the name stand for?

Alfonso: Well II, do you mind if we call you II? I think it sounds pretty good! We named it SFARS which is an acronym for Singular Foliations And Related Structures and the "Good morning" indicates that, well, the seminar is in the mornings.

Leo: And we thought the name fits well with the informal style of the seminar, and the fact that we start the day by seeing our mathematical friends. Btw I also like the II abbreviation.

II: But the whole thing did not start out as a seminar. First there was a conference. How did that come into existence?

Alfonso: Yes, you are right! At the start of this global singularity, I was not sure if I was working too little or too much, but I was definitely not having enough personal interaction to release my stress. So I started trying many online tools to recreate some kind of social life, things like online games and communication software.

Leo: On the mathematical side, there were very few opportunities to talk to fellow (young) mathematicians. A friend had organized an online workshop on multisymplectic geometry and I really enjoyed it, so I thought it might be nice to organize a similar event around singular foliations. Hence, I wrote a message to Alfonso.

Alfonso: And that was a really good news for me! At the same time, I already attended some online conferences using Zoom, and wondered if I could use my gaming experience to improve the running of conferences. So I was thrilled by this challenge. Of course, I said "yes!" and we started planning.

II: So back to the seminar: What is it about?

Alfonso: The seminar it is about nothing in particular. Talks are often related to singular foliations because most of the participants work on that. Nevertheless, speakers select their own theme; they can talk about something they want to learn, or about a question they have. Just like informal everyday math conversations.

Some participants of the SFARS seminar. From top left to bottom right: Leonid Ryvkin, Sara Azzali, Alfonso Garmendia, Jorret Bley, Malte Leimbach, Oscar Cosserat, Charlotte Kirchhoff-Lukat, Karandeep Singh, Vladimir Salnikov.

II: What are the main motivations for participating in the seminar?

Leo: Of course, we have our own ideas about that, but to be sure of ourselves, we distributed a questionnaire among the participants. For most of us, it seems that the seminar is a way to have the "coffee-break" math discussions that we have been missing lately. The short and regular format made it easy to stay engaged, even for people with a busy schedule. And the social time before the talks helped to maintain a friendly environment and supported the informal style of the meetings. However, we admit that for many people it is quite difficult to attend a social activity that early in the morning …

II: How does the technical side of running the seminar work?

Alfonso: Well, we use Zoom for the actual meeting, and Discord as a chat server for written discussions, sharing slides and so forth. Of course, we would have preferred an open source alternative to Zoom, but most people were already accustomed to Zoom and we didn't want to lose participants by unnecessarily changing the setup. In Discord, there is a separate channel for each talk, an "announcements" channel, and a space for general discussions and potential future talk topics.

Leo: And about the talks, we originally decided that they should be short, around 20 minutes. Standing in front of a computer paying close attention for a long period of time is not exactly an appealing and relaxing activity. However, 20 minutes turned out to be a little too short to elaborate on the topics, so we will try 30 minutes next term.

II: What were your experiences with this format?

Alfonso: Having a permanent chat for each topic was very helpful. Sometimes questions or interesting ideas arise a few days after the talk, and they can still be discussed on the corresponding chat. Moreover, it is nice to have a "Journal" of the past topics. Our original idea was that we would decide on topics spontaneously as we went along. But this turned out to be difficult, given the weekly rhythm of the meetings. Now we keep a set of topics in reserve, which we work through if nothing else comes up.

Leo: On the social side, one of the most important aspects is keeping a warm and personal climate. For instance, the participants obviously change, and new ones need to be welcomed into the group. In offline events this happens more naturally, but online it takes a bit more effort.

II: That is a very understandable difficulty, but certainly an important one to overcome. I hope to hear about more initiatives of this sort soon. One last thing, do you want to leave a message to our readers out there?

Alfonso and Leo: (in duet) Hello Mom! I am on EMS!

Alfonso Garmendia and Leonid Ryvkin

Alfonso Garmendia is a Postdoc at the University of Potsdam. His interests lie in symplectic geometry and mathematical physics: Poisson geometry, singular foliation theory, Lie theory (groups, algebras, groupoids and algebroids), deformation quantization and applications in physics. Besides he spends his free time making bad jokes.

garmendiagonzalez@uni-potsdam.de

Leonid Ryvkin is a Postdoc at the Unviersity of Göttingen. He likes to apply methods from homotopical and homological algebra to differential geometry. In addition to multisymplectic geometry and singular foliations, he is interested in mechanics and salsa dancing.

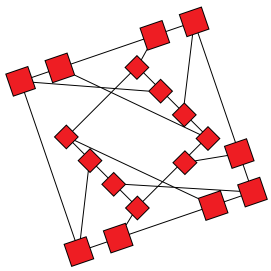
leonid.ryvkin@mathematik.uni-goettingen.de

Croatian Mathematical Society

Franka Miriam Brueckler

The Croatian Mathematical Society (CMS, www.matematika.hr) is the professional association of mathematicians in the Republic of Croatia, whose goal is to improve and promote mathematical science, teaching mathematics at all levels, application of mathematics in other disciplines and improving the social position of mathematicians in general. It was constituted in 1990, but its history is much longer. The formal creation of the CMS after Croatia became an independent state is the result of long-existing scientific societies existing in the former Yugoslavian Socialist Republic of Croatia.

The oldest of the societies which preceded and influenced the formation of the CMS is the Croatian Society of Natural Sciences (CSNS), created back in 1885. This society is a predecessor of all current scientific societies in Croatia, and it still exists, serving its original purpose of development and popularisation of science. Its publications from the period 1886–1944, among others, also contained mathematical articles.



After World War II, in 1945, the Mathematical-Physical Section of the CSNS was formed, and one year later the formation of the Educational Section for Mathematics and Physics followed. The latter organised both discussion meetings for teachers of mathematics and physics, and popular talks for the general public. The first-mentioned, in cooperation with the Astronomical-Geophysical Section, published the scientific magazine *Glasnik matematičko-fizički i astronomski* (1946–1965), in which Blanuša's snark, the graph on the CMS logo was published. It is a graph discovered and published in 1946 by Croatian mathematician Danilo Blanuša (1903–1987) related to his work on the, at that time still open, 4-colour theorem. It was known that in order to prove the conjecture it is sufficient to consider 3-regular bridgeless planar graphs. At that time only one snark (a simple, connected, bridgeless 3-regular graph with chromatic index 4) was known, the Petersen graph, and in his paper Blanuša gave his snark as an example of a 3-regular bridgeless graph which is not 3-edge-colourable, the first of two snarks named after him. In 1996 the CMS chose this graph to be a part of its logo.

Another society, the Society of Croatian Mathematicians and Physicists (SCMP), was also soon created (1949), and it existed until 1990 as part of the Association of Societies of Mathematicians, Physicists and Astronomers of Yugoslavia. Its educational section was very active right from the start, and it began organising several seminars, meetings and events which have survived, more or less continuously, to the present day. This society also began to publish the journal *Matematičko-fizički list*, a professional journal on maths and physics for high school students, which is still being published, jointly by the CMS and the Croatian Physical Society. The still-existing research journal *Glasnik matematički* also began its existence 1966 within the SCMP.

Today, the CMS is a member of the EMS and the IMU (both since 1993). It has seven member societies (Croatian Actuarial Society, Croatian Biometric Society, Croatian Society for Geometry and Graphics, Mathematical Society Istria, Mathematical Society Zadar, Split Mathematical Society and Association of Mathematicians Osijek), and more than 1000 individual members.

The CMS activities are organised within five sections: the Educational Section (CMSEdS), the Scientific Section (CMSSS), the Engineering Section (SMSEnS), the Professional Section (CMSPS) and the Student Section. The CMSEdS also founded the CMS Juniors for students of primary and secondary schools. The CMSSS is involved in the advancement and unification of mathematical research in the Republic of Croatia. Among other activities, it organises regular scientific colloquia and a quadrennial international maths conference, the Croatian mathematical congress. The CMSEdS is active in the improvement of mathematics teaching at all levels. It organises regular professional lectures and conferences (quadrennial Congress of mathematics teachers of the Republic of Croatia). Also, the CMSEdS participates in the organization and implementation of all competitions in mathematics in the Republic of Croatia, including the qualifications for the International Mathematical Olympiad, Middle European Mathematical Olympiad, European Girls Mathematical Olympiad, Cyberspace Mathematical Competition, Junior Balkan Mathematical Olympiad and Mediterranean Mathematical Olympiad. The CMSEdS also organises the

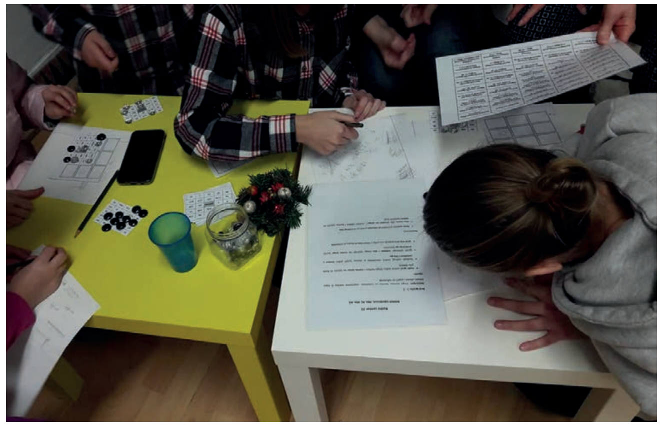
Mathematical Kangaroo Competition in Croatia with more than 50,000 participants every year. The CMSEnS connects mathematicians from various technical and business fields and promotes applications of mathematics. It organises regular popular, professional and scientific lectures. The CMSPS connects mathematics teachers employed at polytechnics and independent colleges in the Republic of Croatia. Its main activity is the organisation of professional colloquia, public lectures and round tables on various topics related to the teaching of mathematics subjects in professional studies.

The CMS also organises a series of very attractive, high-level international conferences in the Inter-University Centre (IUC, www.iuc.hr) in Dubrovnik. They began as quadrennial, however since 2001 they are biannual. The first nine, from 1981 to 2005, were named "Functional Analysis (I–IX)" (including Operator Theory, Banach Algebras, Probability Theory, Differential Operators, Wavelets, Representation Theory), then the name and the subject specialised to "Representation Theory (X–XVI)". Since 2019, Number Theory has also been included. Many well-known mathematicians have participated in these conferences: P. Halmos, G. Weiss, A. V. Skorokhod, V. Kac, D. Vogan (who is one of the co-directors since 2001), M. Duflo, G. Pedersen, G. Pisier, R. Howe, T. Arakawa, W. Soergel, K. Vilonen, J. Lepowski, T. Kobayashi, D. Gourevitch, M. Frank, ….

In 1995, the general assembly of the CMS constituted the CMS Prize for Scientific Contribution to Mathematics. It is awarded quadrennially to a young mathematician up to 35 years of age, living in Croatia. So far the recipients of the prize have been I. Slapničar (1996), A. Dujella (2000), G. Muić (2004), O. Perše (2008), I. Velčić (2012) and A. Mimica (2016). The last recipient, A. Mimica, unfortunately passed away before receiving the prize, following which the CMS decided to name the future CMS prizes after him. In 2020, the recipient of the CMS Ante Mimica Prize was S. Kožić.

The CMS publishes several research and professional mathematics journals: *Matka* (since 1992, for children up to about 15 years of age), the online-journal *math.e* (aimed at high school students, since 2004), *Poučak* (journal for didactics and teaching of mathematics, published since 1998), *Glasnik matematički* (publishes original research papers from all fields of pure and applied mathematics, has been published since 1966, since 2008 it is cited in WOS and SCIE) and *Vjesnik HMD-a* (yearly bulletin on current activities of the society). The CMS also, since 1998, publishes popular mathematics books in a series called *Matkina biblioteka*. So far 20 books of this series were published, the two most recent in 2020.

We end this short presentation of the CMS with the mention of an important popular mathematical activity of the CMS, the yearly Evening of Mathematics, held every year since 2013 on first Thursday in December. It is a collection of mathematical talks, workshops and other activities aimed at children of all ages which has proven to be a great success, with increasing number of participants every year up to sometimes more than 100,000.



Pictures from the Evening of Mathematics

Franka Miriam Brueckler was born 1971 in Essen, Germany and works as assistant professor at the Department of Mathematics in Zagreb, where she received her PhD in operator theory (2002). Her current main research interest is history of mathematics. She teaches Mathematics 1 and 2 for chemists, History of Mathematics, and Crystallography for geologists, is active in popularisation of mathematics, was a member of the EMS Raising Public Awareness Committee (2009–2017, vice-chair 2012–2017) and received the 2009 Croatian State Award for Popularisation of Science.

bruckler@math.hr

ICMI column

Susanne Prediger

Mathematics education in times of the pandemic

All over the world, schooling is currently hindered by the second lockdowns enforced by the pandemic. Teachers and students have to learn with limited time tables or even completely remotely. This works sufficiently well in regions where digital tools have already been established as a normal medium of learning, but very badly where digital tools are not accessible for all teachers and particularly for all students. Some of these students have submerged completely, others are reached only by teachers who cycle around the towns and bring worksheets to everybody personally. Can you imagine communicating about mathematics only on WhatsApp? Can you imagine students sitting at home and filling out worksheets without any oral communication? We can perhaps imagine exercising some procedural skills this way, but surely not developing new conceptual understanding.

In these times, parents and students in many countries become aware of the important advantages of reliable formal education in schools, even for mathematics education. In these times, we also become aware that access to formal education is a privilege that is available to some students much more than to others. Even in privileged countries such as most countries in Europe, the school system is currently not able to ensure fully equal access to mathematics for all students. What a failure of the school system! Some countries have decided to avoid closures in the second waves, while others closed down again.

Once the lockdowns are over (while writing this column in January, I still hope this will be the case by the end of March), we will be able to return to normal schooling. Then we will be able to see how the lockdown has impacted students' mathematical abilities and mathematical understanding. We will need to develop creative ideas for programs compensating for the lost time. We will need to develop programs to support teachers and school administrators in implementing these compensation programs, in order to avoid creating a lost generations, in particular from among the most vulnerable students. Can we, the university mathematicians and university mathematics education researchers, contribute to this process? Let us start discussing this!

ICME 14 in Shanghai planned as hybrid conference in July 2021

Usually, mathematics education researchers from all over the world meet every four years for their *International Congress on Mathematics Education* (ICME). The congress in Shanghai that was originally planned for July 2020 was postponed to July 11 to 18, 2021. It is now planned as a hybrid conference, since many researchers expect difficulties with international travel in Summer 2021.

A hybrid conference poses many challenges. How does one deal with the time shifts in different time zones of the world? How does one include people in the conference site as well as people on the screens?

The main strategy for dealing with the time zones is to fix a core-time that is feasible for nearly all countries in the world. This has been fixed as 19:30–21:30 Shanghai time. During this core time-period, interactive formats are scheduled, such as Topic Study groups. Plenary and invited lectures, on the other hand, will be scheduled outside the core time, and potentially video-recorded. Only those formats which require the most intense interaction (e.g., the Early Career Researchers Day and the poster presentation) will be reserved for the "physical participants", i.e., those who can physically be in Shanghai. In this way, we hope to include as many people as possible in the conference, in spite of the potential travel limitations.

In order to find space to talk about the consequences of the pandemic on mathematics education, an additional panel is planned with this specific focus.

The *registration will be open till June 20th, 2021.* The registration fee for the online participation will only be RMB 2000 (and RMB 3500 for the physical participants) if early payment is made before March, 31, 2021.

More information on ICME 13 in Shanghai can be found at www.icme14.org.

Susanne Prediger is a full professor in mathematics education research at TU Dortmund University, and director of the DZLM network at the IPN Leibniz Institute for Science and Mathematics Education. Her research interests include developing and investigating learning opportunities for secondary mathematics education as well as mathematics teacher education.

prediger@math.uni-dortmund.de

ERME column

Elisabeth Rathgeb-Schnierer, Renata Carvalho, Beatriz Vargas Dorneles and Judy Sayers

ERME Thematic Working Groups. *The European Society for Research in Mathematics Education (ERME) holds a biennial conference (CERME), in which research is presented and discussed in Thematic Working Groups (TWG). We continue the initiative of introducing the working groups, which we began in the September 2017 issue, focusing on ways in which European research in the field of mathematics education could be interesting or relevant for research mathematicians. Our aim is to extend the ERME community to new participants, who may benefit from hearing about research methods and findings and contribute to future CERMEs.*

Introducing CERME Thematic Working Group 2 – Arithmetic and Number Systems

Learning arithmetic and number systems is a relevant field in mathematics education. Kindergarten students make their first steps in developing a number concept of natural numbers with counting and subitizing. Primary school aims to develop a deep conceptual understanding of numbers and basic arithmetic operations, and, a main focus is on building number- and structure-sense as foundation for flexibility in mental arithmetic. A deep understanding of natural numbers is an important prerequisite for developing conceptual understanding and procedural skills in rational numbers in secondary school. Learning arithmetic and number systems is not only a matter of content. Beyond the content, arithmetic education research focuses on models for teaching and learning arithmetic, approaches for heterogeneous and inclusive classrooms, analogue and digital tools to support understanding, and, let us not forget, cultural practices.

Teaching and learning arithmetic and number systems through activities in kindergarten and school is a broad field. For that reason, the thematic working group TWG2 as a forum for presenting and discussing research and development in this subject area is characterized by a great variety of topics. The scope of the TWG comprises kindergarten to 12th grade, and emphasizes for instance research-based specifications of domain-specific goals, analysis of learning processes and learning outcomes in domain-specific learning environments and classroom cultures, as well as new approaches to the design of meaningful and rich learning environments and assessments.

Brief History of TWG2. TWG2 is a relatively young working group which started at CERME7 (2011) in Rzeszów (Poland) with a small group with nine papers, which nevertheless covered various topics including place value, number concepts, adaptive and flexible use of strategies, number sense, fractions in primary education, qualitative and quantitative reasoning, representations and collaborative learning. Two years later, at CERME8 in Antalya (Turkey), the group discussed thirteen papers and one poster. The papers dealt with topics related to natural numbers, fractions and decimals, proportion and negative numbers, and mainly investigated learners' thinking and understanding related to particular mathematics content. Developing conceptual understanding and flexibility in primary and secondary students was shared common ground. The TWG grew steadily, and at CERME9 in Prague (Czech Republic), twenty-one papers and one poster were submitted. The various topics can be categorized in three thematic groups:

(1) number sense, conceptual understanding and flexibility;
(2) the role of models;
(3) subject matter analysis.

Even if the number of discussed papers and posters slightly decreased, CERME10 was characterized by an even greater variety of themes. Fifteen papers and one poster addressing research for different ages and different approaches were intensively discussed. The key themes were number sense and structure sense, estimation and estimation tasks, flexibility in mental calculation, derived fact-based strategies for multiplication in low-achieving students, understanding of rational numbers and ratios, didactic models as scaffolds for the evolution of mathematical knowledge, as well as teachers' knowledge about rational numbers, ratio and place value. The most recent meeting of TWG2 was CERME11. While CERME9 and CERME10 were dominated by design-based research, CERME11 was characterized by a balance of qualitative and quantitative approaches (more details below).

Working spirit of TWG2. TWG2 is a diverse group which gathers researchers from different domains (mathematics education, psychology, mathematics), and from different levels of experience. Approximately one third of the researchers have been with the group from the beginning, and have built a stable basis for developing common ideas and concepts. The fluctuation of group members between the conferences is challenging in terms of developing a common understanding of concepts, but it is also a great opportunity for enhancing our perspectives and notions.

The TWG work always starts with a small working-group with two aims: firstly, to give all participants – especially researchers who are new to the group – an opportunity to engage in personal interaction, and secondly, to engage everybody in an intensive discussion from the start, especially those people who are not as fluent and secure in the conference language as others. According to our experience of recent years, this small group discussion on thematically related papers provides a fruitful basis for joint work during the conference week. This is especially efficient because we assign each group member to two related papers four weeks in advance, and provide specific questions for individual preparation.

CERME has the spirit of a working conference, and participants are supposed to read all the papers of the TWG in advance. This makes it possible to place more emphasis on discussion and scientific exchange during the joint work sessions. Depending on the amount of submitted papers, it is our goal to discuss each paper at least for 30 minutes after a short introduction in the form of a five-minute-reminder. Finally, each group member gives written feedback after each discussion.

Current and further work of TWG2. Due to the specific situation caused by the Covid-19-pandemic, the last time TWG2 was able to meet was in spring 2019 at CERME 11 in Utrecht (Netherlands). Here, twenty papers and four posters that reflected the richness of the field were introduced and discussed. The group had the chance to broaden the discussion by regarding the same topics from the perspectives of both cognitive psychologists and mathematics education researchers. In comparison to the TWG work of the previous years, there was also a change regarding the approach to research.

The various papers presented and discussed were grouped into four overarching themes: The role of manipulatives in learning processes in early arithmetic, learning and teaching numbers and operations in kindergarten to first grade, learning and teaching arithmetic in second to sixth grade, and various approaches to learning rational numbers. Additionally, there was one paper on metacognition. Like all the previous years, the group work is characterized by a great variety of topics and approaches. This is always a challenge for developing a common understanding, but it is also the strength of TWG2. The great variety of papers under discussion gave rise to interesting and often animated discussions that went far beyond the specific topics of single papers. We worked out differences and commonalities of used terms, concepts, theoretical frameworks and methodological approaches. Once again, we identified different conventions of naming numbers and operations, different understandings of terms and concepts such as number sense, flexibility and mastery, and partly conflicting views of quantitative and qualitative approaches. Altogether, we broadened our own culturally-influenced perspectives and drew new inspiration for further research.

The aspects we discussed were manifold and reflected the whole spectrum of learning arithmetic in primary and secondary level, dealing with such topics as:

- The role and relevance of counting in different cultures and curricula, and a critical reflection on counting for developing strategies in mental calculation.
- The importance of designing learning environments based on our knowledge of students' strategies, conceptions and misconceptions.
- The emphasis on fostering students' conceptual knowledge regarding numbers and operations in different number systems.
- The clarification of methodology regarding what is measured, why measures are used, what is meant by a result, and how it contributes to a better understanding of teaching and learning.
- The challenges of transition between number systems, between representations, and from intervention research to students' long-term understanding.

Our research is characterized by a huge variety of questions, yet it is driven by a common goal: supporting understanding and meaningful learning in all students so as to develop their conceptual knowledge, number sense and structure sense in different number systems, as well as flexible and adaptive expertise in arithmetic and number systems. Empirical results suggest the early development of a conceptual understanding of numbers and basic arithmetic operations as a precursor for mathematical achievement in school (Krajewski & Schneider 2009). Learning processes, teaching approaches and learning trajectories from early arithmetic to university mathematics is an important field of research which opens up the possibility of joint work between mathematicians and math educators.

In Utrecht, we decided to meet in May 2020 for an ERME Topic Conference at the University of Leeds, to specify terms and work on common understanding of number sense, flexibility and mastery. All planning was successfully completed in February 2020, and 32 abstracts were submitted and accepted, but we were then obliged to postpone the conference due to Covid-19. The situation still has not changed, and so the ETC will take place as a virtual conference on May, 12, 2021, with a young researcher day on May, 11, 2021 to enable joint work to take place.

Reference

[1] K. Krajewski and W. Schneider, Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learn. Instr.* **19**, 513–526 (2009)

Elisabeth Rathgeb-Schnierer is professor for mathematics education at the University of Kassel. She has been engaged in TWG2 since CERME8, and is a member of the leading team since CERME9. With regard to learning arithmetic, she pursues various research interests such as learning arithmetic in heterogeneous groups and fostering students with learning disabilities. One of her main foci is developing cognitive flexibility in mental arithmetic, a field in which she has carried out various national and international projects.

rathgeb-schnierer@mathematik.uni-kassel.de

Renata Carvalho holds a PhD in mathematics education. She collaborates with the UIDEF of the Institute of Education of the University of Lisbon as a researcher and with the School of Education of Lisbon in preservice education. She also works with in-service teachers in professional development courses. Her first participation in the TWG2 was in CERME8 as a young researcher, and has acted as co-leader since CERME10. Her research focus is mental computation and rational numbers at the elementary levels.

renatacarvalho@campus.ul.pt

Beatriz Vargas Dorneles is full professor for Cognitive Psychology at the Federal University of Rio Grande do Sul (UFRGS), Brazil. She has been engaged in TWG2 since CERME10, and is a member of the leading team since CERME11. She pursues research interests such as learning arithmetic in deaf and hearing students as well as in different groups of students with learning disabilities. One of her foci is cognitive process supporting arithmetic and estimation.

beatriz.dorneles@ufrgs.br

Judy Sayers is a lecturer and researcher at the University of Leeds in England. She first joined CERME5 in Cyprus, but joined the newly formed TWG2 Arithmetic group in CERME7 in Rzeszow where she continues to enjoy great conversation, debates and empowering new ideas. Her research interest in early years' mathematics led her to the development of a large funded project Foundational Number Sense (FoNS) with colleagues at Stockholm University, where she continues to collaborate.

j.m.sayers@leeds.ac.uk

Short note: zbMATH Open

Klaus Hulek

A long term project has finally become reality: zbMATH has become an open access database as of 1st January 2021, allowing free usage worldwide. This is the result of a process which has lasted several years and our thanks go to all the individuals and institutions who have made this possible.

The mathematical community is invited to participate in the future development of the database. We will now be open to share data and links with other non-commercial databases. This opens up the way to new cooperations and, we hope, novel and potentially unexpected new developments in the future. We are currently working on an API which will allow much of our data for research to be downloaded for research and non-commercial purposes. Please share you ideas on the future of the database with us via editor@zbmath.org. And of course, we are always looking for new reviewers: you can register via our website: zbmath.org/become-a-reviewer/

Last but not least: we thank Springer Verlag for many years of good cooperation. zbMATH would not exist without Springer Verlag. At the same time the landscape of publishing is undergoing fundamental changes and we believe that the new model is the right direction to follow in the future.

Klaus Hulek is professor of mathematics at Leibniz University Hannover and Editor-in-Chief of zbMATH Open. His field of research is algebraic geometry.

hulek@math.uni-hannover.de

DOI 10.4171/MAG-11

zbMATH Open: Towards standardized machine interfaces to expose bibliographic metadata

Moritz Schubotz and Olaf Teschke

In this article, we give motivation for the need for standardized machine interfaces to zbMATH open data, outline the target audience, describe our preliminary strategy to develop API interfaces, and report on the details of the first interface we implemented.

1 Target audience

As announced in the previous note, zbMATH is becoming open access from the 1st of January 2021.[1] For most working mathematicians, this means that they can access zbMATH from anywhere in the world without a subscription or authentication. Additionally, we envision benefits to the community through our efforts to connect zbMATH data with information systems of research data, collaborative platforms, funding agencies, and intradisciplinary efforts, as outlined in [2]. We expect that our efforts to disseminate the results of mathematical research will provide this research with increased visibility. However, to target domain-independent information systems, we need to comply with standardized information exchange protocols and interfaces.

In what follows, we describe potential partners that might interact with zbMATH. We will offer the data via so-called Application Programming Interfaces (APIs). Moreover, in this report, we focus on how others can make use of zbMATH open data, rather than how zbMATH can use other data sources. As depicted in Figure 1, the potential consumers can be clustered into at least five groups, which we will describe below.

[1] The open web platform is now available under the name zbMATH Open, while we will address the zbMATH content and services under the traditional umbrella name zbMATH for convenience.

Bibliographic consumer	MathOverflow Wikimedia arXiv Zotero	a_1 Selection of individual items a_2 High throughput a_3 End-user-friendly formats a_4 Various representations a_5 Fuzzy search
Aggregators	OpenAIRE/ERC NFDI/DFG	b_1 Standard compliance b_2 Incremental updates b_3 Projection on properties
Archives	Software Heritage Internet archive	c_1 Fetch everything c_2 Reduce traffic c_3 Traceability of versions
Search engines	Firefox search plugin	d_1 Selection of individual items d_2 High throughput d_3 End-user-friendly formats d_4 Various representations d_5 Fuzzy search d_6 Formula search
Individuals	Blog on specific topic Personal reference list	e_1 Easy to setup e_2 Long term stability

Figure 1. Envisioned consumer (left) and desiderata (right) [5]

Bibliographic consumers are information systems that display references to scientific publications. They often deal with user-generated content that references individual research articles. For websites like Wikipedia or MathOverflow, users interactively search for references to support their statements. The remote information system, e.g., MathOverflow, sends the user's search-string to a designated zbMATH API endpoint, which then returns a ranked list of possible references. The remote information system takes care of the formatting. While MathOverflow, for instance, might use zbMATH exclusively, others, such as Wikipedia, might want to fuse results from zbMATH with results from other providers of bibliographic metadata. Standardized protocols drastically reduce the implementation effort for intra-domain information systems. Even before the transformation to zbMATH open, we provided a simple API for MathOverflow [4], which was limited to the top three search results. This legal restriction has now vanished. In contrast, to interactive bibliographic customers described before, arXiv and other publishers might use zbMATH's bibliographic metadata to disambiguate references, which is an essential prerequisite for many information retrieval tasks such as recommendations, semantic searches, or plagiarism detection [3].

Aggregators such as the OpenAIRE research explorer, SemanticScholar, DataCite, or Altmetric extract information from different sources, transform them to standardized representations and load them into their specific data models. Additionally, in some countries, researchers are also required to report their publications to government platforms. At the end of the process, funding agencies or other decision-makers can use these data sources for so-called data-driven decision making. Here standardized interfaces and formats evolved to simplify the aggregators' job, as crawling through web-pages optimized for human consumption is error-prone and involves complicated heuristics that are fragile and vulnerable to layout changes.

Archives such as the Internet Archive and Software Heritage capture the digital history in the forms of websites or software code. Since their mission is digital preservation, an API that enables replaying the entire history of the website would be ideal. Moreover, they strive to reduce traffic consumption and avoid redundancy. Their infrastructure is optimized to preserve HTML websites in the form presented to a user at a particular point in time.

Search Engines might use our API to present search results in a different format. For instance, Mozilla Firefox has a built-in feature to include custom search engines that implement the OpenSearch standard. One interesting feature to consider is if and how mathematical formulæ are represented in OpenSearch.

Individuals or small groups of people with particular needs are of particular importance to us. We aim to provide as much support as possible to research groups, either from mathematics or from the field of bibliometric research. Highly motivated individuals who aim to use our data creatively are also on our schedule. Here, we are open to requests, and need to investigate potential uses case-by-case. A typical, not too exceptional use case we envision would be to set up a personal publication list or to enrich a personal website with the latest news of specific Mathematics Subject Classification (MSC) classes. While many of these functionalities are already possible with zbMATH's news-feed functionality, we expect the API functionality to be more flexible.

2 First steps towards APIs

Given the diverse expectations and needs described above, we do not see a one-size-fits-all solution that would fulfill the diverse requirements. Therefore, we decided to pursue an iterative approach to building API solutions. As a first step, we aim to start with a first API version that is well-established, easy to implement, and has a substantial positive impact on working mathematicians. According to our analysis, aggregators, archives, and bibliometric researchers commonly use the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). OAI-PMH seems to be well-established, sufficiently documented, and relatively easy to implement. This protocol is also well-suited to downloading, i.e. to harvesting the entire open collection of zbMATH document data. These data come along with a CC-BY-SA license, which facilitates both reusability and allowing derived work to remain in the open ecosystem, although this comes at the cost that some third-party content (such as abstracts) is not included due to legal constraints. Additionally, one may harvest well-defined subsets and consume updates since the last download without requiring to redownload a dump. We expect that this format will also be well-suited to individuals working with zbMATH data. Especially consumers that work with other datasets besides zbMATH will appreciate the standardized functionality of the protocol. However, the format is not well-suited for bibliographic consumers, given the overhead caused by the standard and the lack of flexibility. Because of this, we decided to create at least two APIs, with the OAI-PMH API as a starting point, and other more flexible APIs to be determined.

In Figure 2, we display a possible scenario for zbMATH's future API development efforts. The blue boxes (Reviewer Interface, Internal Interfaces, zbMATH database, and zbMATH Website) show the well-established components of zbMATH. The dark gray box (OAI-PMH API) shows the newly released API described in this paper. With this setup, all write operations to the database will be performed from the reviewer interface and other internal interfaces. In contrast, the Website and the OAI-PMH interfaces are read-only interfaces that present the zbMATH database's contents without modifying it.

As a next step, we are working on an API to create links from zbMATH to external sources and vice versa. One commonly accepted format to describe these connections is the Scholix format. Therefore we labeled this project "Scholix Link API" in Figure 2. Independently of this task, we are also working on a general-purpose API (Gray box Custom API in the figure) that replicates the current website's functionality but produces the results in a better machine-processible form, such as JSON instead of HTML. In theory, one might use this API for a far-future version of the zbMATH website, given that efficient caching layers are implemented. While we are pursuing the linking and custom API efforts in parallel, our goal is to limit as far as possible the number of distinct API endpoints. Another vital link will be a bidirectional link to research data in the context of the German Mathematical Research Data Initiative (MaRDI), a consortium formed for applications within the National Research Data Infrastructure (see [1]). In the MaRDI project (light gray box at the top), we plan to repurpose the generic WikiBase knowledge graph software that supports many well-established structured graph data exchange protocols, such as RDF and SPARQL among others.

3 Implementation

Given the above motivations, we have implemented a first version of the OAI-PMH interface. The current demo is available from purl.org/zb/10. As required by the protocol, our OAI-PMH api offers six endpoints, namely (1 Identify, 2 ListMetadataFormats, 3 ListSets, 4 ListIdentifiers, 5 ListRecords, 6 GetRecord):

- Endpoint 1 helps aggregators and archives to discover the new API fully unsupervised, identify which version of the OAI-PMH standards we are using, and other technicalities.
- Endpoint 2 lists the formats that we use to expose zbMATH data. We implemented two flavors. The first is the required standard Dublin Core Metadata Record format, which contains standardized fields like abstract, publisher, creator, or title. For the second, we implemented a format that is closer to zbMATH's internal data model. Many domain-specific classifications can be expressed in terms of Dublin core vocabulary. However, the MSC is not predefined in the Dublin core standard, even though it seems to us that it could be modeled. However, expressing all the details of zbMATH's data in Dublin core terms would require an immense effort of coordination with librarians to ensure that our encodings are modeled according to common best practices for modelling specifics in Dublin core. In other words, we are addressing the issue from two ends. With the DC standard, we encode the low hanging fruits in a standard way. With our additional zbMATH custom format, we ensure that we expose all the data we are legally allowed to by the API.
- Endpoint 3 lists the subsets of the zbMATH dataset that we think could be relevant. In the first version, we provide the following sets: document type, year, document author, classification, keyword, document language, author variation, author reference, biographic reference, software, review type, review language, reviewer, serial publisher. For example, the set `document_author:Noether, Emmy` is the subset of all zbMATH entries authored by Emmy Noether. As an extension to OAI-PMH's built-in set logic, we implemented magic characters `|&~` that indicate the standard set operations 'or', 'and', and 'not' respectively, allowing users to combine sets at will. Obviously, in endpoint 3, we only enumerate the 1 125 144 base sets.
- Endpoints 4 and 5 list the currently 4 206 870 list zbMATH identifiers and records, respectively. This endpoint is predestined to

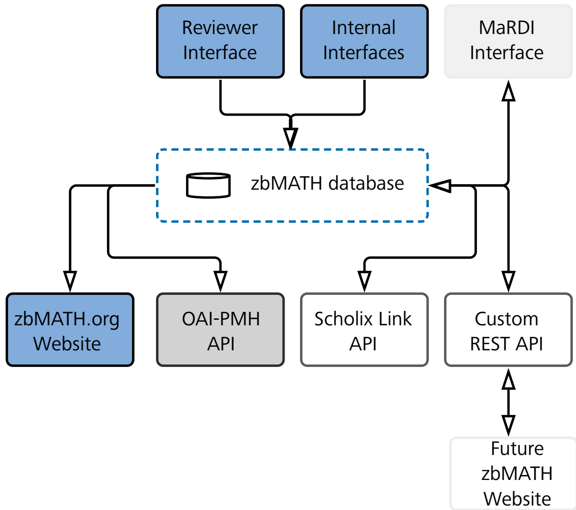


Figure 2. A conceptual overview of the zbMATH database and the data flow from an to the database

obtain a dump of all public zbMATH open data. Here OAI-PMH's built-in cursor and resumption mechanism ensure an efficient and seamless retrieval of the data. For the convenience of end-users, one can use one of the many available OAI-PMH metadata harvesters from www.openarchives.org/pmh/tools/ to retrieve all the data.
- Endpoint 6 gets individual zbMATH entries.

To conclude with a real example, assume that we want to retrieve the OAI-PMH metadata for the entry with zblnumber 1200.35057. We would first need to retrieve the corresponding internal identifier (DE number), which can be done by clicking on the BibTeX button below the article. In this example, a BibTeX entry with key zbMATH05797851 will be downloaded open. The last digits following the word zbMATH, i.e., 5797851, are the DE number. One can then use this number to retrieve the metadata from the API by appending the query

```
verb=GetRecord&identifier=oai:zbmath.org:
5797851&metadataPrefix=oai_zb_preview
```

to the root of the API endpoint in the browser. Here "verb" identifies the endpoint (6=GetRecord), and the DE number is prefixed with identifier prefix and postfixed with the desired metadata format. The browser will then display a large XML file that contains the review text and other public information available on the zbMATH website. See purl.org/zb/11 for comparison to the website view. One can use this method to obtain any document from the zbMATH open database without downloading large sets of articles.

4 Conclusion

We have introduced the target audience of our API, discussed our strategy of rolling out APIs to cover a wide range of potential users, and described details of our API infrastructure's first pillar. While our plans for future endpoints are subject to change and the current OAI-PMH endpoint is subject to continual improvement, we have taken the first step towards standardized machine interfaces to make the data of zbMATH available to a broader audience.

References

[1] K. Hulek, F. Müller, M. Schubotz and O. Teschke, Mathematical research data – an analysis through zbMATH references. *Eur. Math.Soc. Newsl.* **113**, 54–57 (2019)

[2] K. Hulek and O. Teschke, Die Transformation von zbMATH zu einer offenen Plattform für die Mathematik. *Mitt. Dtsch. Math.-Ver.* **28**, 108–111 (2020)

[3] N. Meuschke, V. Stange, M. Schubotz, M. Kramer and B. Gipp, Improving academic plagiarism detection for STEM documents by analyzing mathematical content and citations. *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, Champaign, IL, USA, 120–129 (2019)

[4] F. Müller, M. Schubotz and O. Teschke, References to research literature in QA forums – a case study of zbMATH links from MathOverflow. *Eur. Math. Soc. Newsl.* **114**, 50–52 (2019)

[5] M. Schubotz, D. Trautwein and O. Teschke, zbMATH is open: A practical guide to open an informationservice. *Proceedings of The Open Science Conference 2021 (OSC '21), February 17–19, 2020*, online

Moritz Schubotz is a senior researcher for mathematical information retrieval and open science. He maintains the support for mathematical formulæ in Wikipedia and is off-site collaborator at NIST.

moritz.schubotz@fiz-karlsruhe.de

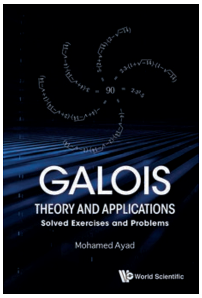
Olaf Teschke is Managing editor of zbMATH and Vice-chair of the EMS Committee on publications and electronic dissemination.

olaf.teschke@fiz-karlsruhe.de

Book reviews

About Galois Theory and Applications – Solved Exercises and Problems by Mohamed Ayad

Reviewed by Jean-Paul Allouche



A book entirely devoted to exercises and problems with solutions about Galois theory and its applications: is this something new? is this something interesting? The answer to the first question is probably yes. Of course there are already several books on the subject that contain exercises; there is even a book entitled *Galois Theory Through Exercises* by J. Brzeziński. But the book under review contains nothing but a very large number of exercises (285) that one *irresistibly* wants to attack – without even looking at the solutions provided by the author. That is what I did, browsing through the eleven chapters of the book (Polynomials, Fields, Generalities; Algebraic extension, Algebraic closure; Separability, Inseparability; Normal extensions, Galois extensions, Galois groups; Finite fields; Permutation polynomials; Transcendental extensions, Linearly disjoint extensions, Luroth's theorem; Multivariate polynomials; Integral elements, Algebraic number theory; Derivations), picking exercises, and trying to solve them. Here are a few examples: Exercise 2.31 begins with solving $y^k - (1 + x) = 0$ in an algebraic closure of $K(x)$, and ends by asking for a proof that $3x^4 + 6x^3 + 5x^2 + 2x$ can be expressed as the composition of two polynomials of degree 2. Exercise 5.26 plays with the cyclic extensions of degree 4 over the rationals; Exercise 8.1 looks innocent but is related to the Christol theorem on algebraicity of "automatic" series; Exercise 8.3 requires proving (with a hint …) that if L is an algebraically closed field such that $L \subset E$, where E is an extension of finite type of an algebraically closed field K, then one must have $L \subset K$. Exercise 10.24 leads to the determination of all the distinct factorizations of 90 into a product of irreducibles in $\mathbb{Z}[\sqrt{-14}]$, a problem which yields the nice image of the cover page. Solving (or trying to solve …) all the exercises gives a clear answer to the second question at the beginning of this survey: yes, this book is definitely interesting, and I warmly recommend it. It can be used by beginners who want to learn about Galois theory in a more playful manner, by colleagues who want to teach really everything about Galois theory, and even by researchers who might discover useful results and ideas there.

Mohamed Ayad, *About Galois Theory and Applications. Solved Exercises and Problems*. World Scientific, 2018, 452 pages.
Hardcover ISBN 978-981-3238-30-5. eBook ISBN 978-981-3238-32-9.

Meilensteine der Rechentechnik (Milestones in Analog and Digital Computing) by Herbert Bruderer

Reviewed by Jean-Paul Allouche

After the impressive and remarkable first edition (one volume of more than 800 pages, see the *Newsletter of the EMS*, December 2016, Issue 102, p. 154), this second edition consists of two volumes totaling over 1500 pages. It has the same good qualities as the previous edition but contains twice as much material, which makes the set of the two volumes of the second edition an extremely useful contribution to the history of computing machines. I will concentrate on the second volume. The first part is devoted to general questions and answers about computers, from "Who invented the first computer" to "What is a Turing machine?" through questions about theoretical computer science, algorithms and universal machines; from "What is a von Neumann computer" (and is it a series or parallel computer?) to theoretical questions about storage; from political and historical issues to technical developments. This first chapter is already extraordinarily rich. It is followed by chapters that provide a detailed analysis of events in three different countries, namely Germany, Great Britain, and Switzerland. In these chapters, we learn an incredible number of things that most of us probably never suspected, e.g., the difference between computing machines and logical devices according to Konrad Zuse (and questions about "computers" playing chess), the whole history of *Enigma* and of the "Turing–Welchman bomb", the question of whether Churchill really ordered all "colossal" computers to be destroyed, the history of the Swiss computer *Ermeth*, acronym for "Elektronische Rechenmaschine der ETH" (followed in particular by *Lilith* and *Ceres*: given that these last two names are related to religious figures, one might ask whether the acronym Ermeth had something to do with the Hebrew word for truth, namely *emeth*, see, e.g., biblicalstudies.org.uk/pdf/ifes/5-4_blocher.pdf, page 50). A further chapter is devoted to the first computing devices from almost twenty other countries: let me just cite Spain with Leonardo Torres y Quevedo, his analog computer, his chess-playing computer and his analytical machine. The book also contains an extremely useful dictionary for all technical terms, giving the English equivalent for all German words or expressions and vice versa, not to mention a bibliography over 300 pages long and an amazing set of images! After having read this volume and the first volume, what strikes me most is the incredibly rich history of computer science, and the incredibly deep ignorance of this history by essentially everybody who uses computers for whatever purpose. This is just one of the reasons for which these two beautiful and well-documented volumes should definitely be necessary reading.

Added note: I was going to add that it would be good to translate these volumes into English (and French), when I learned that an English version is due to appear very soon – the electronic English version was made available on January 6 (www.springer.com/de/book/9783030409739).

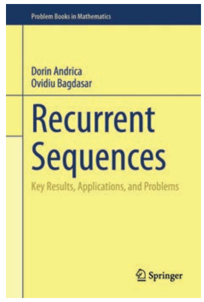
Herbert Bruderer, *Meilensteine der Rechentechnik. Erfindung des Computers, Elektronenrechner, Entwicklungen in Deutschland, England und der Schweiz*, 2. Auflage, Band 2 (*Milestones in Analog and Digital Computing*, 2nd edition, volume 2). De Gruyter Oldenburg, 2018, 829 pages. ISBN 978-3-11-060088-9. e-ISBN 978-3-11-060261-6.

Jean-Paul Allouche is "Directeur de recherche" emeritus, at CNRS. He is working at IMJ-PRG, Sorbonne, Paris (France) on subjects relating number theory and theoretical computer science, including the so-called "automatic sequences". He was editor of the *EMS Newsletter* in 2013–2020.

jean-paul.allouche@imj-prg.fr

Recurrent Sequences. Key Results, Applications and Problems
by Dorin Andrica and Ovidiu Bagdasar

Reviewed by Michael Th. Rassias



This book contains an ample presentation of recurrent sequences from multiple perspectives, initiating the readers with classical results and gradually leading them to the very frontier of what is known in the subject. The expository style is engaging and the succinct presentation of theoretical results is accompanied by short but tricky examples which invite the reader to investigate the topic further.

The first six chapters of the book present classical and recent results on the topic. Numerous results have been obtained by the authors, and highlight connections between recurrences and combinatorics, number theory, integer sequences, and random number generation. The diagrams of orbits of second and third-order recurrent sequences in the complex plane presented in the book add significantly to its artistic quality. About a third of the book is devoted to an inspired selection of 123 (the 10th Lucas number) Olympiad training problems, accompanied by detailed solutions.

Chapter 1 offers a succinct presentation of the fundamentals of recurrence relations, along with examples of recurrent sequences naturally arising in algebra, combinatorics, geometry, analysis, and mathematical modelling.

Chapter 2 is devoted to first and second-order linear recursions, as well as homographic recurrences. Examples include the Fibonacci sequence and its close companions: the Lucas, Pell or Pell–Lucas sequences, for which the authors present a palette of interesting identities with elegant proofs. The discussion extends to special families of polynomials, which are then related back to the Fibonacci, Lucas, Pell and Pell–Lucas sequences, and used to establish novel number theoretic results. This chapter also presents homographic sequences with constant and variable coefficients.

Chapter 3 presents arithmetic properties of the Fibonacci, Lucas, Pell and Pell–Lucas sequences, with links to pseudoprimality. The authors prove new theoretical results, present recent entries to the Online Encyclopedia of Integer Sequences, and formulate a few interesting conjectures. Some of these results have already been extended to generalized Pell and Pell-Lucas sequences in the recent paper [Andrica, D. and Bagdasar, O.: On some new arithmetic properties of the generalized Lucas sequences. *Mediterr. J. Math.*, to appear (2021)]. The complex factorization of the polynomials

$$U_n = \frac{x^n - y^n}{x - y} \qquad \text{and} \qquad V_n = x^n + y^n, \; n = 0, 1, \ldots,$$

is used to derive some elegant trigonometric formulae for these classical sequences.

Chapter 4 is devoted to ordinary and exponential generating functions, which are used to evaluate the general term formulae for many classical polynomials and sequences in Sections 4.1 and 4.2. The interesting version of Cauchy's integral formula given in Section 4.3 is used to derive integral formulae for the Fibonacci, Lucas, Pell and Pell–Lucas sequences. In some recent papers the authors have used this approach to establish novel integral formulae for the coefficients of cyclotomic, Gaussian, multinomial, or polygonal polynomials.

Chapter 5 explores second order linear recurrences depending on a family of four complex coefficients (often called Horadam sequences). The results include necessary and sufficient conditions for periodicity (Section 5.2), the geometric structure (Section 5.3), and the enumeration of Horadam orbits with a fixed length (Section 5.4). An atlas presenting numerous beautiful diagrams of periodic and non-periodic Horadam patterns is given in Section 5.6, while Section 5.7 presents a Horadam-based pseudo-random number generator. Some examples of periodic non-homogeneous Horadam sequences are given in Section 5.8. The chapter is based on many recent articles.

Chapter 6 further develops the ideas presented in Chapter 2, featuring a collection of useful methods related to generating functions, matrices and interpolating geometric inequalities. Some results for systems of linear recurrence sequences are also given, with applications to Diophantine equations. Extending the results from Chapter 5, the authors present complex linear recurrent sequences of higher order, periodicity conditions, geometric structure, and enumeration of periodic orbits with a fixed length. An atlas of exciting geometric patterns produced by third-order linear recursions in the complex plane is also showcased. The chapter concludes with a presentation of connections between the theory of linear recurrences and finite differences.

Chapter 7 contains 123 Olympiad training problems involving recurrent sequences, which are solved in detail in Chapter 8, sometimes with multiple solutions. The problems concern linear recurrence sequences of first, second and higher orders, some classical sequences, homographic sequences, systems of sequences, complex recurrence sequences, and recursions in combinatorics. Many of the problems were actually proposed by the authors, while the others were selected from international competitions or classical journals.

The book ends with an appendix and a rich bibliography including 177 references, many of which represent contributions by the authors. An index is also provided.

This book teaches numerous fundamental facts and techniques which are central in mathematics. It is both a research monograph and a delightful problem book, which I feel will spark the interest of a wide audience, from mathematics Olympiad competitors and their coaches to undergraduate or postgraduate students, or professional mathematicians with an interest in recurrences and their multiple applications.

Dorin Andrica and Ovidiu Bagdasar, *Recurrent Sequences. Key Results, Applications and Problems*, Springer International Publishing, 2020, 402 + xiv pages. Hardcover ISBN 978-3-030-51501-0.
eBook ISBN 978-3-030-51502-7.

Michael Th. Rassias is a member of the Editorial Board of the Newsletter/Magazine of the EMS. He is a Research Fellow at the I-Math of the University of Zürich and a visiting researcher at PIDS of the IAS, Princeton. He holds a Diploma from the NTUA, Greece, a Master of Advanced Study from the University of Cambridge, and a PhD from ETH Zürich. He has been awarded with two Gold medals in Mathematical Olympiads in Greece, a Silver medal in the IMO, and with the Notara Prize of the Academy of Athens. He has published 16 books and volumes by Springer, including *Open Problems in Mathematics* with J. F. Nash, Jr. He has published several research papers in Mathematical Analysis and Analytic Number Theory. His homepage is www.mthrassias.com.

michail.rassias@math.uzh.ch

Solved and unsolved problems

Michael Th. Rassias

Probability theory is nothing but
common sense reduced to calculation.

Pierre-Simon Laplace (1749–1827)

The present column is devoted to Probability Theory and related topics.

I Six new problems – solutions solicited

Solutions will appear in a subsequent issue.

237

We take for our probability space (X, m): the unit interval $X = [0,1]$ equipped with the Lebesgue measure m defined on $\mathcal{B}(X)$, the Borel subsets of X and let (X, m, T) be an invertible measure preserving transformation, that is $T : X_0 \to X_0$ is a bimeasurable bijection of some Borel set $X_0 \in \mathcal{B}(X)$ of full measure so that and $m(TA) = m(T^{-1}A) = m(A)$ for every $A \in \mathcal{B}(X)$.

Suppose also that T is ergodic in the sense that the only T-invariant Borel sets have either zero- or full measure ($A \in \mathcal{B}(X),\ TA = A \implies m(A) = 0, 1$).

Birkhoff's ergodic theorem says that for every integrable function $f : X \to \mathbb{R}$,

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \xrightarrow[n\to\infty]{} \mathbb{E}(f) := \int_X f dm \text{ a.s.}$$

The present exercise is concerned with the possibility of generalizing this. Throughout, (X, m, T) is an arbitrary ergodic, measure-preserving transformation as above.

Warm-up 1
Show that if $f : X \to \mathbb{R}$ is measurable, and

$$m\left(\left[\overline{\lim_{n\to\infty}}\left|\sum_{k=0}^{n-1} f \circ T^k\right| < \infty\right]\right) > 0,$$

then $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ converges in $\mathbb{R}$ a.s.

Warm-up 1 is [1, Lemma 1]. For a multidimensional version, see [1, Conjecture 3].

Warm-up 2
Show that if $f : X \to \mathbb{R}$ is as in Warm-up 1, then $\exists\ g,\ h : X \to \mathbb{R}$ measurable with h bounded so that $f = h + g - g \circ T^n$.

Warm-up 2 is established by adapting the proof of [3, Theorem A].

Problem
Show that there is a measurable function $f : X \to \mathbb{R}$ satisfying $\mathbb{E}(|f|) = \infty$ so that

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$$

converges in $\mathbb{R}$ a.s.

The existence of such f for a specially constructed ergodic measure-preserving transformation is shown in [2, example b]. The point here is to prove it for an arbitrary ergodic measure preserving transformation of (X, m).

References
[1] P. Hagelstein, D. Herden and A. Stokolos, A theorem of Besicovitch and a generalization of the Birkhoff Ergodic Theorem. *Proc. Amer. Math. Soc. Ser. B* **8**, 52–59 (2021)
[2] D. Tanny, A zero-one law for stationary sequences. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **30**, 139–148 (1974)
[3] D. Volný and B. Weiss, Coboundaries in L_0^∞. *Ann. Inst. H. Poincaré Probab. Statist.* **40**, 771–778 (2004)

Jon Aaronson (Tel Aviv University, Israel)

238

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{X_n : n \geq 1\}$ a sequence of independent and identically distributed (i.i.d.) random variables on Ω. Assume that there exists a sequence of positive numbers $\{b_n : n \geq 1\}$ such that $\frac{b_n}{n} \leq \frac{b_{n+1}}{n+1}$ for every $n \geq 1$, $\lim_{n\to\infty} \frac{b_n}{n} = \infty$, and $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geq b_n) < \infty$. Prove that if $S_n := \sum_{j=1}^{n} X_j$ for each $n \geq 1$, then

$$\lim_{n\to\infty} \frac{S_n}{b_n} = 0 \text{ almost surely.}$$

Comment. The desired statement says that if such a sequence $\{b_n : n \geq 1\}$ exists, then $\{X_n : n \geq 1\}$ satisfies the (generalized) Strong Law of Large Numbers (SLLN) when averaged by $\{b_n : n \geq 1\}$. If $X_n \in L^1(\mathbb{P})$ for every $n \geq 1$, then the desired statement follows trivially from Kolmogorov's SLLN, since in that case, with probability one, we have

$$\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}[X_1]$$

and hence

$$\frac{S_n}{b_n} = \frac{S_n}{n} \cdot \frac{n}{b_n}$$

must converge to 0 under the assumptions on $\{b_n : n \geq 1\}$. Therefore, the desired statement can be viewed as an alternative to Kolmogorov's SLLN for i.i.d. random variables that are not integrable.

Linan Chen (McGill University, Montreal, Quebec, Canada)

239

In Beetown, the bees have a strict rule: all clubs must have exactly k membees. Clubs are not necessarily disjoint. Let $b(k)$ be the smallest number of clubs that it is possible for the $n \geq k^2$ bees to form, such that the set of clubs has the property that no matter how the bees divide themselves into two teams to play beeball, there will always be a club all of whose membees are on the same team. Prove that

$$2^{k-1} \leq b(k) \leq Ck^2 \cdot 2^k$$

for some constant $C > 0$.

Rob Morris (IMPA, Rio de Janeiro, Brasil)

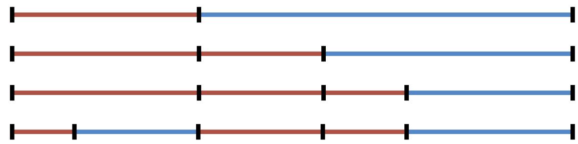
240

N agents are in a room with a server, and each agent is looking to get served, at which point the agent leaves the room. At any discrete time-step, each agent may choose to either shout or stay quiet, and an agent gets served in that round if (and only if) that agent is the only one to have shouted. The agents are indistinguishable to each other at the start, but at each subsequent step, every agent gets to see who has shouted and who has not. If all the agents are required to use the same randomised strategy, show that the minimum time expectation to clear the room is $N + \big(2 + o(1)\big)\log_2 N$.

Bhargav Narayanan (Rutgers University, Piscataway, USA)

241

Consider the following sequence of partitions of the unit interval I: First, define π_1 to be the partition of I into two intervals, a red interval of length 1/3 and a blue one of length 2/3. Next, for any $m > 1$, define π_{m+1} to be the partition derived from π_m by splitting all intervals of maximal length in π_m, each into two intervals, a red one of ratio 1/3 and a blue one of ratio 2/3, just as in the first step. For example π_2 consists of three intervals of lengths 1/3 (red), 2/9 (red) and 4/9 (blue), the last two are the result of splitting the blue interval in π_1. The figure below illustrates $\pi_1, \ldots, \pi_4$, from top to bottom.



Let $m \in \mathbb{N}$ and consider the m-th partition π_m.

1. Choose an interval in π_m uniformly at random. Let R_m be the probability you chose a red interval, does the sequence $(R_m)_{m\in\mathbb{N}}$ converge? If so, what is the limit?
2. Choose a point in I uniformly at random. Let A_m be the probability that the point is colored red, does the sequence $(A_m)_{m\in\mathbb{N}}$ converge? If so, what is the limit?

Yotam Smilansky (Rutgers University, NJ, USA)

242

Prove that there exist $c < 1$ and $\epsilon > 0$ such that if $A_1, \ldots, A_k$ are increasing events of independent boolean random variables with $\Pr(A_i) < \epsilon$ for all i, then

$$\Pr(\text{exactly one of } A_1, \ldots, A_k \text{ occurs}) \leq c.$$

(What is the smallest c that you can prove?)

Here $A \subset \{0,1\}^n$ is an "increasing event" if whenever $x \in A$, then the vector obtained by changing any coordinates of x from 0 to 1 still lies in A.

A useful fact is Harris inequality, which says that for increasing events A and B of boolean random variables, $\Pr(A \cap B) \geq \Pr(A)\Pr(B)$.

I learned of this problem from Jeff Kahn.

Yufei Zhao (MIT, Cambridge, USA)

II Open problems

Equidistributed orbits in 2-adic integers

by Hillel Furstenberg (Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Israel)

The Collatz problem, also known as the $3x+1$ problem is very well known. We will formulate a general conjecture motivated by the Collatz problem, as well as another related conjecture. Both have to do with ordinary integer sequences which should be regarded as subsequences of the 2-adics. First some notation. Every positive integer can be written as $x = g(x)h(x)$, with $g(x)$ a power of 2, and $h(x)$ odd. We define a transformation on $\mathbb{N} = \{1, 2, 3, \ldots\}$ by

$$T(x) = h(3x+1).$$

The mysterious phenomenon is that every orbit seems to reach the fixed point 1. The Collatz problem is to prove this. Probability enters as a heuristic explanation of why the orbit is finite, always reaching some cycle. A more "robust" phenomenon is that for any odd b, if we were to set

$$T(x) = h(3x+b),$$

the orbits are finite.

We point out that it is not hard to show that b can be chosen so that the corresponding transformation has as many distinct cycles as we like. Nevertheless it seems that all orbits are finite. We propose a still more general conjecture which implies this. For this we need the notion of "equidistribution in the odd 2-adics".

If we fix a natural number k, every odd integer lies in one of the $2^{(k-1)}$ arithmetic progressions

$$P_{j,k} = j + 2^k Z, \quad j = 1, 3, 5, \ldots, 2^k - 1.$$

Definition. A sequence S of odd integers is equidistributed in the odd 2-adics if for every j, k, the proportion of S in $P_{j,k}$ is $(1/2)^{(k-1)}$. "Proportion" means the relative density.

Now define

$$T_{a,b}(x) = h(ax+b),$$

where a and b are odd integers, $a > 2$, $b > 0$.

*243**

Conjecture A. For any natural odd numbers a, b, x, the orbit of x under $T_{a,b}$ is either finite or equidistributed in the odd 2-adics.

Consequence: All the orbits of $T_{3,b}$ are finite.

To prove this one shows that the assumption of an infinite equidistributed orbit leads to a contradiction. This follows from the fact that on account of equidistribution, the expectation of $\log g(x)$ is $\log 4$. The idea now is to observe that for large x, the effect of applying $T_{3,b}$ is – on the average – multiplying by 3 and dividing by 4. This can be made precise to show that an equidistributed, infinite orbit is an impossibility.

The next conjecture involves equidistribution in the full compact group of 2-adics. Define $R(x) = 3[x/2]$. Here as usual $[z]$ denotes the largest integer less than z. We have

$$R(0) = R(1) = 0,\ R(2) = 3,\ R(3) = 3.$$

But for $x > 3$,

$$R(x) > x,$$

so that for $x > 3$, the orbit of x will be infinite.

*244**

Conjecture B. For any $x > 3$, the orbit under R is equidistributed in the 2-adics.

In particular, this would imply that in every such orbit, even and odd integers appear with the same frequency. However, we are unable to verify this for any orbit.

Here too we can consider R_a defined by

$$R_a(x) = a[x/2], \quad \text{for odd } a.$$

We expect the same phenomenon: orbits that are either finite or equidistributed. Little numerical work has been done on this conjecture.

III Solutions

226

Let $\mathbb{C}^n$ stand for the space of complex *column* n-vectors, and let $\mathbb{M}_n$ stand for the space of complex $n \times n$ matrices.

The inner product $\langle x|y\rangle$ of $x, y \in \mathbb{C}^n$ is defined as

$$\langle x|y\rangle = x^* y \quad \text{(matrix product)}.$$

Therefore $\langle x|y\rangle$ is linear in y and conjugate linear in x.

Let A, B be $n \times m$ complex matrices. Write them as

$$A = [a_1, \ldots, a_m] \quad \text{and} \quad B = [b_1, \ldots, b_m]$$

with $a_j, b_j \in \mathbb{C}^n$ $(j = 1, 2, \ldots, m)$.

Then it is immediate from the definition of matrix multiplication that

$$A^*B = \left[\langle a_j|b_k\rangle\right]_{j,k=1}^{m} \in \mathbb{M}_m.$$

Show the following relation:

$$AB^* = \sum_{j=1}^{m} a_j b_j^* \in \mathbb{M}_n$$

where each product $a_j b_j^*$ $(j = 1, \ldots, n)$ is a rank-one matrix in $\mathbb{M}_n$.

T. Ando (Hokkaido University, Sapporo, Japan)

Solution by the proposer
Consider the canonical orthonormal basis in $\mathbb{C}^n$;

$$e_j := \left[\delta_{j,k}\right]_{k=1}^{n}, \quad (j = 1, 2, \dots, n).$$

Then it suffices to prove the identity in the assertion for the case

$$A = \left[0, \dots, 0, \overset{(p)}{e_j}, 0, \dots, 0\right], \quad \exists\ 1 \le j \le n,\ \exists\ 1 \le p \le m$$

and

$$B = \left[0, \dots, 0, \overset{(q)}{e_k}, 0, \dots, 0\right], \quad \exists\ 1 \le k \le n,\ \exists\ 1 \le q \le m$$

In this case

$$AB^* = \delta_{p,q} e_j e_k^*$$

while

$$\sum_{j=1}^{m} a_j b_j^* = \delta_{p,q} e_j e_k^*$$

as expected.

Also solved by John N. Daras (Athens, Greece), Muhammad Thoriq (Yogyakarta, Indonesia), and Socratis Varelogiannis (France)

227

Let p and q be two distinct primes with $q > p$ and G a group of exponent q for which the map $f_p : G \to G$ defined by $f_p(x) = x^p$, for all $x \in G$, is an endomorphism. Show that G is an abelian group.

Dorin Andrica and George Cătălin Țurcaş (Babeş-Bolyai University, Cluj-Napoca, Romania)

Solution by the proposer
We first prove the following auxiliary result.

Claim. The map $f_k : G \to G$ defined by $f_k(x) = x^k$ is an endomorphism for all $k \in \mathbb{Z}$ with $k \equiv 1 \pmod{p(p-1)}$.

First, it is not hard to see that $(xy)^p = x^p y^p$ implies that $(yx)^{p-1} = x^{p-1}y^{p-1}$, for all $x, y \in G$. Then we observe that

$$(xy)^{(p-1)^2} = (y^{p-1}x^{p-1})^{p-1} = x^{(p-1)^2} y^{(p-1)^2}, \quad \forall x, y \in G.$$

It is also easy to show that

$$x^p y^{p-1} = y^{p-1} x^p, \quad \forall x, y \in G. \qquad (\bigstar)$$

Using the latter, we see that

$$(xy)^{p(p-1)} = (x^p y^p)^{p-1} = y^{p(p-1)} x^{p(p-1)} = x^{p(p-1)} y^{p(p-1)}$$

for all $x, y \in G$. We just showed that $f_{p(p-1)}$ is an endomorphism of G. We proceed by showing that f_{p^2-p+1} is also an endomorphism.

For every $x, y \in G$, we have

$$(xy)^{p^2-p+1} = (xy)^p \cdot (xy)^{(p-1)^2} = x^p \cdot y^p \cdot x^{(p-1)^2} \cdot y^{(p-1)^2}.$$

In the above chain of equalities we just used that f_p and $f_{p(p-1)}$ are endomorphisms. We previously mentioned ($\bigstar$) that the middle terms commute, which shows that f_{p^2-p+1} is an endomorphism of G.

Observe that for all $x, y \in G$ we have

$$x^{p^2-p+1} y^{p^2-p+1} = (xy)^{p^2-p+1} = (xy) x^{p^2-p} y^{p^2-p},$$

which implies that $x^{p^2-p} y = y x^{p^2-p}$. It follows that x^{p^2-p} belongs to the center of G, for any $x \in G$.

We are now ready to prove our claim. Let $k = 1 + mp(p-1)$ for some $m \in \mathbb{Z}$. Then

$$(xy)^k = (xy)^{mp(p-1)} \cdot (xy) = x^{p(p-1)m} y^{p(p-1)m} \cdot (xy) = x^k y^k.$$

The claim is proved.

Since $\gcd(q, p(p-1)) = 1$, by the Chinese Remainder Theorem we know that there is an integer n such that

$$\begin{cases} n \equiv 1 \pmod{p(p-1)}, \\ n \equiv 2 \pmod{q}. \end{cases}$$

We therefore have that the map $f_n : G \to G$ defined by $f_n(x) = x^n$ is an endormorphism. However, $x^n = x^2$ for all $x \in G$, since q is the exponent of G. Now, since $f_2(x) = x^2$ is an endomorphism of G, it follows that for all $x, y \in G$ we have $(yx)^2 = y^2x^2 \Leftrightarrow yxyx = y^2x^2 \Leftrightarrow xy = yx$.

Remark. The conclusion also holds if p and q are not prime. One just needs G to have a finite exponent q such that q and $p^2 - p$ are coprime.

Also solved by Mihaly Bencze (Romania), Tomek Jedrzejak (University of Szczecin, Poland), and Efstathios S. Louridas (Athens, Greece)

228

Let $(G, \cdot)$ be a group with the property that there is an integer $n \ge 1$ such that the map $f_n : G \to G, f_n(x) = x^n$ is injective and the map

$$f_{n+1} : G \to G, f_{n+1}(x) = x^{n+1}$$

is a surjective endomorphism. Prove that G is an abelian group.

Dorin Andrica and George Cătălin Țurcaş (Babeş-Bolyai University, Cluj-Napoca, Romania)

Solution by the proposer

From the second hypothesis we have

$$(xy)^{n+1} = x^{n+1}y^{n+1},$$

which implies that

$$(yx)^n = x^n y^n \quad \text{for all } x, y \in G.$$

Using the above, we see that $x(yx)^n = x^{n+1}y^n$ for all $x, y \in G$. On the other hand,

$$x(yx)^n = (xy)^n x = y^n x^{n+1} \quad \text{for all } x, y \in G.$$

We just showed that

$$x^{n+1}y^n = y^n x^{n+1} \quad \text{for all } x, y \in G. \tag{$\star$}$$

Now, using the surjectivity of f_{n+1} we obtain that for every $z \in G$ there is $x \in G$ such that $f_{n+1}(x) = z$, that is $x^{n+1} = z$. The relation ($\star$) can be written as $zy^n = y^n z$, for all $y, z \in G$. From this relation we get

$$y(xy)^n = (xy)^n y \quad \text{for all } x, y \in G,$$

that is

$$y(xy)(xy)\cdots(xy) = (xy)^n y.$$

This is equivalent to

$$(yx)(yx)\cdots(yx)y = (xy)^n y,$$

hence $(yx)^n = (xy)^n$ and the conclusion follows from the injectivity of the map f_n.

Also solved by John N. Daras (Athens, Greece), Tomek Jedrzejak (University of Szczecin, Poland), Muhammad Thoriq (Yogyakarta, Indonesia), and Socratis Varelogiannis (France)

229

Let A and $B \in \mathrm{Mat}_k(K)$ be two matrices over a field K. We say that A and B are *similar* if there exists an invertible matrix $C \in \mathrm{GL}_k(K)$ such that $B = C^{-1}AC$.

Let A and $B \in \mathrm{GL}_k(\mathbb{Q})$ be two similar invertible matrices over the field of rational numbers $\mathbb{Q}$. Assume that for some integer l, we have $A^{l+1}B = BA^l$. Then A and B are the identity matrices.

Andrei Jaikin-Zapirain (Departamento de Matemáticas, Universidad Autónoma de Madrid and Instituto de Ciencias Matemáticas, CSIC-UAM-UC3M-UCM, Spain) and Dmitri Piontkovski (Faculty of Economic Sciences, Moscow Higher School of Economics, Russia)

Solution by the proposer

Consider first a finite group G having three elements x, y, z satisfying

$$y = z^{-1}xz \quad \text{and} \quad x^{l+1}y = yx^l.$$

Let us show that $x = y = 1$.

By way of contradiction we assume that the order of x is $n > 1$. Since x and y are conjugate, the order of y is also n. The cases $l = 0$ and $l = -1$ are trivial, so we assume that $l \neq 0, -1$.

Let $\mathrm{GCD}(a, b)$ denote the greatest common divisor of two integers a and b. The order of x^l is $n/\mathrm{GCD}(n, l)$ and the order of x^{l+1} is $n/\mathrm{GCD}(n, l+1)$. Since x^l and x^{l+1} are conjugate, their orders coincide. Therefore, $\mathrm{GCD}(n, l) = \mathrm{GCD}(n, l+1)$. Thus, since l and $l+1$ are coprime, we obtain that

$$\mathrm{GCD}(n, l) = \mathrm{GCD}(n, l+1) = 1.$$

This implies that there exists a natural number $q > 1$, which is coprime with n and such that $ql \equiv l + 1 \pmod{n}$. Note that $yx^l y^{-1} = x^{lq}$. Therefore,

$$x^l = y^{-n}x^l y^n = y^{-n+1}x^{lq}y^{n-1} = \cdots = x^{lq^n}.$$

Since the order of x^l is n, n divides $q^n - 1$. Let p be the smallest prime divisor of n. Then the integers n and $p-1$ are coprime. Hence there exist $a, b \in \mathbb{Z}$ such that $an + b(p-1) = 1$. Observe also that

$$q^n \equiv q^{p-1} \equiv 1 \pmod{p}.$$

Therefore, $q^{an+b(p-1)} \equiv 1 \pmod{p}$, and so

$$l \equiv l \cdot q^{an+b(p-1)} \equiv l \cdot q \equiv l + 1 \pmod{p}.$$

We have arrived at a contradiction, which proves that $x = y = 1$.

Now let us come back to the original problem. Since A and B are similar, there exists an invertible matrix $C \in \mathrm{GL}_k(\mathbb{Q})$ such that $B = C^{-1}AC$. Let F be the subgroup of $\mathrm{GL}_k(\mathbb{Q})$ generated by A and C. Let m be such that A, $A^{-1}, C, C^{-1} \in \mathrm{GL}_k(\mathbb{Z}[\frac{1}{m}])$. Then $F \leq \mathrm{GL}_k(\mathbb{Z}[\frac{1}{m}])$.

Recall that a group H is called *residually finite* if for every non-trivial element $h \in H$ there exists a finite quotient G of H such that the image of h in G is non-trivial. Observe that the group $\mathrm{GL}_k(\mathbb{Z}[\frac{1}{m}]$ is residually finite (consider the natural maps from $\mathrm{GL}_k(\mathbb{Z}[\frac{1}{m}])$ to $\mathrm{GL}_k(\mathbb{F}_p)$, where p are prime numbers coprime with m). Therefore we conclude that the group F is also residually finite.

If A is not the identity matrix, then there exists a finite quotient G of F such that the image of A in G is not trivial. But this contradicts what we proved at the beginning. Thus A, so also B, are the identity matrices.

Remark. The problem is inspired by a result of Baumslag [1] where he constructed a two-generator one-relator group having only cyclic finite quotients. Instead of $\mathbb{Q}$ we could assume that the matrices A and B in the problem are considered over an arbitrary field K. In this case the problem can be solved using a theorem of Malcev [2] where he proved that any finitely generated group linear over a field is residually finite.

References

[1] G. Baumslag, A non-cyclic one-relator group all of whose finite quotients are cyclic. *J. Austral. Math. Soc.* **10**, 497–498 (1969)

[2] A. Malcev, On isomorphic matrix representations of infinite groups. *Rec. Math. [Mat. Sbornik] N.S.* **8 (50)**, 405–422 (1940)

Also solved by John N. Daras (Athens, Greece), George Miliakos (Sparta, Greece), and Moubinool Omarjee (Paris, France)

230

We are trying to hang a picture on a wall. The picture has a piece of string attached to it forming a loop, and there are 3 nails in the wall that we can wrap the string around. We want to hang the picture so that it does not fall down, but it will upon the removal of any of the 3 nails.

Dawid Kielak (Mathematical Institute, University of Oxford, UK)

Solution by the proposer

Let us start with 2 nails. Wrapping a loop around two nails in some way is equivalent to choosing an element of $\pi_1(\mathbb{C} \setminus \{0,1\})$. Of course, this fundamental group is the free group $F_2 = F(a,b)$ of rank 2, with generators corresponding to loops going around one of the nails.

If the picture is not to fall down, we need a non-trivial element x of F_2. The condition that the picture is supposed to fall when any of the nails is removed means that the image of x in $F_2/\langle\langle a\rangle\rangle \cong \mathbb{Z}$ and in $F_2/\langle\langle b\rangle\rangle \cong \mathbb{Z}$ has to be trivial. We immediately recognise that $x = [a,b]$ does the trick.

For three nails, we can take

$$[[a,b],c] \in F(a,b,c) \cong \pi_1(\mathbb{C} \setminus \{0,1,2\}).$$

The solution easily generalises to n nails.

Also solved by Mihaly Bencze (Romania), and George Miliakos (Sparta, Greece)

231

Given a natural number n and a field k, let $M_n(k)$ be the full $n \times n$ matrix algebra over k. A matrix $(a_{ij}) \in M_n(k)$ is said to be centrosymmetric if

$$a_{ij} = a_{n+1-i,n+1-j}$$

for $1 \le i,j \le n$. Let $C_n(k)$ denote the set of all centrosymmetric matrices in $M_n(k)$. Then $C_n(k)$ is a subalgebra of $M_n(k)$, called centrosymmetric matrix algebra over k of degree n. Centrosymmetric matrices have a long history (see [1, 5]) and applications in many areas, such as in Markov processes, engineering problems and quantum physics (see [2, 3, 4, 6]). In the representation theory of algebras, a fundamental problem for a finite-dimensional algebra is to know if it has finitely many nonisomorphic indecomposable modules (or in other terminology, representations). In our case, the concrete problem on $C_n(k)$ reads as follows.

Does $C_n(k)$ have finitely many nonisomorphic indecomposable modules? If yes, what is the number?

Changchang Xi (School of Mathematical Sciences, Capital Normal University, Beijing, and College of Mathematics and Information Science, Henan Normal University, Xinxiang, China)

Solution by the proposer

Strategy. To solve the problem, we use the fact that two algebras have the same number of nonisomorphic indecomposable modules if their module categories are equivalent. In our case of $C_n(k)$, a practical way is to look for a decomposition of $C_n(k)$ as a direct sum of indecomposable left ideals, and then take the direct sum of representives for each isomorphism classes of indecomposable left ideals. The endomorphism algebra of this direct sum of representives is called the basic algebra of $C_n(k)$, denoted $B_0(n,k)$. Then $C_n(k)$ and $B_0(n,k)$ have equivalent module categories, and therefore they have the same number of nonisomorphic indecomposable modules.

Technique. Let I_n be the identity matrix in $M_n(k)$, e_{ij} be the $n \times n$ matrix with (i,j)-entry 1 and other entries 0, and

$$f_i = e_{ii} + e_{n+1-i,n+1-i} \text{ for } 1 \le i \le n.$$

Then $f_i \in C_n(k)$. Calculations show that, for $n = 2m$,

1. $I_n = f_1 + \cdots + f_m$, $f_i f_j = \delta_{ij} f_i$, where δ_{ij} is the Kronecker symbol,

2. $C_n(k)f_1 \simeq C_n(k)f_2 \simeq \cdots \simeq C_n(k)f_m$ as left ideals, and
3. $\dim_k(C_{2m}(k) = 2m^2$;

and, for $n = 2m + 1$,

1. $I_n = f_1 + \cdots + f_m + e_{m+1,m+1}, f_i f_j = \delta_{ij} f_i$, where δ_{ij} is the Kronecker symbol, and
2. $C_n(k)f_1 \simeq \cdots \simeq C_n(k)f_m$ as left ideals, and
3. $\dim_k(C_{2m+1}(k) = 2m^2 + 2m + 1$.

Thus

$$B_0(2m,k) = f_1 C_{2m}(k) f_1$$

and

$$B_0(2m+1,k) = (f_1 + e_{m+1,m+1}) C_{2m+1}(k) (f_1 + e_{m+1,m+1}).$$

Further calculations lead to

$$B_0(2m,k) \simeq C_2(k),\ B_0(2m+1,k) \simeq C_3(k) \quad \text{for all } m \geq 1.$$

Answer. We give quiver presentations of $B_0(n,k)$ (see [7]).

Clearly, $B_0(1,k) = k\ \bullet$. If $\mathrm{char}(k) \neq 2$, then $B_0(2,k) = k(\bullet \quad \bullet)$ and $B_0(3,k) \simeq k(\bullet \quad \bullet)$. If $\mathrm{char}(k) = 2$, then

$$C_2(k) = k(\ \bullet \circlearrowleft \alpha\ )/(\alpha^2), \qquad C_3(k) = k(\ \bullet \underset{\beta}{\overset{\alpha}{\rightleftarrows}} \bullet\ )/(\alpha\beta).$$

From these quiver presentations of the basic algebras, we can draw their Auslander–Reiten quivers and gain a complete answer to the above problem.

(a) $C_1(k)$ has exactly 1 nonisomorphic indecomposable module.
(b) If $\mathrm{ch}(k) \neq 2$, then $C_n(k)$ has exactly 2 nonisomorphic indecomposable modules for all $n \geq 2$.
(c) If $\mathrm{ch}(k) = 2$, then $C_{2m}(k)$ has exactly 2 nonisomorphic indecomposable modules for all $m \geq 1$, and $C_{2m+1}(k)$ has exactly 5 nonisomorphic indecomposable modules for all $m \geq 1$.

References
[1] A. Aitken, *Determinants and matrices*. Oliver and Boyd, Edinburgh (1939)
[2] A. R. Collar, On centrosymmetric and centroskew matrices. *Quart. J. Mech. Appl. Math.* **15**, 265–281 (1962)
[3] L. Datta and S. D. Morgera, On the reducibility of centrosymmetric matrices – applications in engineering problems. *Circuits Systems Signal Process.* **8**, 71–96 (1989)
[4] I. J. Good, The inverse of a centrosymmetric matrix. *Technometrics* **12**, 925–928 (1970)
[5] T. Muir, *A treatise on the theory of determinants*. Revised and enlarged by William H. Metzler, Dover Publications, Inc., New York (1960)
[6] J. R. Weaver, Centrosymmetric (cross-symmetric) matrices, their basic properties, eigenvalues, and eigenvectors. *Amer. Math. Monthly* **92**, 711–717 (1985)
[7] C. Xi and S. Yin, Cellularity of centrosymmetric matrix algebras and Frobenius extensions. *Linear Algebra Appl.* **590**, 317–329 (2020)

An additional interesting problem (not intimately connected to Algebra)

Intervals of monotonic changes in a polynomial are located between the roots of its derivative. A derivative of a polynomial is also a polynomial, although of a lesser degree. Using these considerations, construct an algorithm for calculating the real roots of the quadratic equation. Improve it to calculate the real roots of a polynomial of the third, fourth and generally arbitrary degree.

Igor Kostin (Moscow, Russian Federation)

Solution by the proposer

We assume that the coefficient at x^2 of the square polynomial is greater than zero. We need to find the derivative of this polynomial.

This is a linear function and we can easily find its root. Let us denote it by xmin.

It is clear that if the value of the initial square polynomial at xmin is greater than zero, then such a polynomial has no real roots. Otherwise, we will look for an argument xref such that this polynomial is greater than zero. An easy way to find such an argument is to step back from xmin by some step and check the value of the polynomial in such a step.

If the calculated value of the polynomial is greater than zero, then the search is completed, otherwise we will continue to retreat, each time doubling the value of the step of the retreat. Having obtained xref, we find the root by the standard dichotomy method, dividing in half the segment between xmin and xref.

It is convenient to end the dichotomy process when, in the machine representation, the point dividing the segment in half coincides with one of the ends of the original segment. Due to the finite precision of real numbers, this will happen sooner or later on any computer.

We repeat the above method for calculating the root twice, departing from xmin in different directions.

It is now clear that in order to calculate the roots of a polynomial of third degree, one must first calculate its derivative polynomial and find the roots of this square polynomial.

These roots will determine the boundaries of the intervals of monotonic changes in the initial polynomial of third degree. The roots of this initial polynomial will be found by the dichotomy method on the segments of monotonic variation calculated in this way.

It is clear that the ladder of the described possibilities rises, in principle, to a polynomial of an arbitrarily large degree. Of course, the complete solution to this problem should be a computer program that implements the verbal instructions listed here. In practice, the capabilities of such an algorithm are limited in that the counting time increases with the degree of the original polynomial.

Finally, we note that the xref points replacing the infinity value for the boundaries of the segments of the monotonicity of the polynomial can be calculated in a less primitive way than the step-by-step search with doubling of the step considered above. We normalize the polynomial so that the coefficient of the highest power of the argument is equal to one.

Let M be the largest modulo value among all its coefficients. If the argument x of the polynomial is greater than $M+2$, then the value of the polynomial is greater than 1. To prove this, consider the calculation of the polynomial

$$p(x) = x^n + k[n-1]x^{(n-1)} + \cdots + k[1]x + k[0]$$

by the Horner scheme.

At the first step we calculate

$$p[1] = k[n-1] + x,$$

and it is obvious that $p[1] > 1$. In fact even if $k[n-1] < 0$, it does not exceed M in absolute value.

At the second step we calculate

$$p[2] = k[n-2] + xp[1],$$

and again it is obvious that $p[2] > 1$.

The same holds in the following steps.

At the last step we compute

$$p(x) = k[0] + xp[n-1]$$

and finally obtain $p(x) > 1$. Thus, if one needs to set a representative value of the polynomial with an infinite value of the argument, one should take the argument equal to $M+2$.

Solution to part (a) of Open problem 137 proposed by Ovidiu Furdui (Romania), September 2014 Issue of the EMS Newsletter*

Solution by Seán M. Stewart (Bomaderry, NSW, Australia)

Denote the integral to be found in (a) by I. In terms of known constants its value is:

$$I = 8 - \pi - \frac{\pi^2}{8} + \frac{\pi}{2}\log 2 + 2(\log 2)^2 - 6\log 2 - 2\mathbf{G}.$$

Here $\mathbf{G}$ is *Catalan's constant* defined by $\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2}$.

We first establish a number of preliminary results before evaluating the integral. The nth *harmonic number* is defined by

$$H_n = \sum_{k=1}^{n} \frac{1}{k}, \quad n \in \mathbb{N}.$$

By convention, $H_0 \equiv 0$. The harmonic numbers satisfy the following recurrence relation

$$H_{n+1} = H_n + \frac{1}{n+1}.$$

In terms of the harmonic numbers the following finite sum can be written as

$$\begin{aligned}\sum_{k=1}^{n} \frac{1}{2k-1} &= 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-1} \\ &= \left(1 + \frac{1}{2} + \cdots + \frac{1}{2n}\right) - \frac{1}{2}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \\ &= H_{2n} - \frac{1}{2}H_n. \qquad (1)\end{aligned}$$

The harmonic numbers are related to the *digamma function* by (see entry 5.4.14 in [1])

$$\psi(x+1) = -\gamma + H_x. \qquad (2)$$

Here γ is the Euler–Mascheroni constant. This allows the harmonic numbers to be analytically continued to all $x \in \mathbb{R}, x \neq -1, -2, -3, \ldots$. The functional relation for the digamma function is

$$\psi(x+1) = \psi(x) + \frac{1}{x}. \qquad (3)$$

For half-integer arguments the digamma function takes the values (see entry 5.4.15 in [1])

$$\psi\left(n + \frac{1}{2}\right) = -\gamma - 2\log 2 + 2\sum_{k=1}^{n} \frac{1}{2k-1}. \qquad (4)$$

Replacing n with $n + \frac{1}{2}$ in (2) we see that

$$H_{n+\frac{1}{2}} = \psi\left(n + \frac{3}{2}\right) + \gamma = \psi\left(n + \frac{1}{2}\right) + \frac{2}{2n+1} + \gamma,$$

where we have made use of the functional for the digamma function. In view of (4) we may rewrite this as

$$H_{n+\frac{1}{2}} = -2\log 2 + \frac{2}{2n+1} + 2\sum_{k=1}^{n} \frac{1}{2k-1} = -2\log 2 + 2\sum_{k=1}^{n+1} \frac{1}{2k-1}.$$

By applying (1), in terms of harmonic numbers this can be expressed as

$$H_{n+\frac{1}{2}} = 2H_{2n+2} - H_{n+1} - 2\log 2,$$

which reduces to

$$H_{n+\frac{1}{2}} = 2H_{2n+1} - H_n - 2\log 2, \qquad (5)$$

after the recurrence relation for the harmonic numbers is applied to each of the harmonic number terms that appear.

Two results from Cauchy products for power series will prove useful. The first is

$$\arctan^2(x) = -\sum_{n=1}^{\infty} \frac{(-1)^n}{n}\left(H_{2n} - \frac{1}{2}H_n\right)x^{2n}, \qquad |x| \leqslant 1. \qquad (6)$$

Showing this we have

$$\begin{aligned}\arctan^2(x) &= \left(\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1}\right) \cdot \left(\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1}\right) \\ &= x^2 \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{(-1)^n}{(2k+1)(2n-2k+1)} x^{2n} \\ &= \frac{x^2}{2}\sum_{n=0}^{\infty} \frac{(-1)^n}{n+1}\left[\sum_{k=0}^{n} \frac{1}{2k+1} - \sum_{k=0}^{n} \frac{1}{2k-2n-1}\right]x^{2n},\end{aligned}$$

where we have made use of the partial fraction decomposition of

$$\frac{1}{(2k+1)(2n-2k+1)} = \frac{1}{2(n+1)(2k+1)} - \frac{1}{2(n+1)(2k-2n-1)}.$$

Reindexing the second sum $k \mapsto n-k$ we have

$$\arctan^2(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} \left(\sum_{k=0}^{n} \frac{1}{2k+1} \right) x^{2n+2}.$$

Reindexing the infinite sum $n \mapsto n-1$ and the finite sum $k \mapsto k-1$ yields

$$\arctan^2(x) = -\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \left(\sum_{k=1}^{n} \frac{1}{2k-1} \right) x^{2n},$$

with the desired result then following on application of (1).

The second Cauchy product for power series is

$$\frac{x \arctan x}{1+x^2} = \sum_{n=1}^{\infty} (-1)^{n+1} \left(H_{2n} - \frac{1}{2} H_n \right) x^{2n}, \quad |x| < 1. \tag{7}$$

Showing this we have

$$\begin{aligned}
\frac{x \arctan x}{1+x^2} &= x^2 \left(\sum_{n=0}^{\infty} (-1)^n x^{2n} \right) \cdot \left(\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{2n+1} \right) \\
&= \sum_{n=0}^{\infty} (-1)^n \left(\sum_{k=0}^{n} \frac{1}{2k+1} \right) x^{2n+2} \\
&= \sum_{n=0}^{\infty} (-1)^n \left(\sum_{k=1}^{n+1} \frac{1}{2k-1} \right) x^{2n+2} \\
&= \sum_{n=0}^{\infty} (-1)^n \left(H_{2n+2} - \frac{1}{2} H_{n+1} \right) x^{2n+2},
\end{aligned}$$

where (1) has been used. The desired result then follows after a reindexing of $n \mapsto n-1$.

Turning our attention to the integral I we have

$$\begin{aligned}
I &= -\int_0^1 \log(1-x^2) \sum_{n=1}^{\infty} \frac{(-1)^n x^{2n}}{n} \, dx \\
&= -\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \int_0^1 x^{2n} \log(1-x^2) \, dx.
\end{aligned}$$

Here the interchange made between the summation and the integration is permissible due to Fubini's theorem. Using the variable change $x \mapsto \sqrt{x}$ leads to

$$I = -\frac{1}{2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \int_0^1 x^{n-\frac{1}{2}} \log(1-x) \, dx. \tag{8}$$

From the well-known result of (see, for example, [2, p. 2, Eq. (1.4)])

$$\int_0^1 x^{n-1} \log(1-x) \, dx = -\frac{H_n}{n},$$

replacing n with $n + \frac{1}{2}$ we see that

$$\int_0^1 x^{n-\frac{1}{2}} \log(1-x) \, dx = -\frac{2H_{n+\frac{1}{2}}}{2n+1},$$

allowing us to rewrite the integral appearing in (8) as

$$I = \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \frac{H_{n+\frac{1}{2}}}{2n+1}. \tag{9}$$

In view of (5) this can be rewritten as

$$I = \sum_{n=1}^{\infty} \frac{(-1)^n}{n(2n+1)} \big(2H_{2n+1} - H_n - 2\log(2) \big). \tag{10}$$

From the partial fraction decomposition of

$$\frac{1}{n(2n+1)} = \frac{1}{n} - \frac{2}{2n+1},$$

and the recurrence relation for the harmonic numbers, (10) may be expressed as

$$\begin{aligned}
I &= 2 \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \left(H_{2n} - \frac{1}{2} H_n \right) - 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{2n+1} \left(H_{2n} - \frac{1}{2} H_n \right) \\
&\quad + 2(1-\log 2) \sum_{n=1}^{\infty} \frac{(-1)^n}{n} + 4(\log 2 - 1) \sum_{n=1}^{\infty} \frac{(-1)^n}{2n+1} \\
&\quad - 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n+1)^2} \\
&= 2S_1 - 4S_2 + 2(1-\log 2) S_3 + 4(\log 2 - 1) S_4 - 4S_5.
\end{aligned} \tag{11}$$

We now find each of these five sums. For the first of these, setting $x = 1$ in (6) we see that

$$S_1 = -\arctan^2(1) = -\frac{\pi^2}{16}.$$

For the second of the sums, S_2, integrating (7) with respect to x from 0 to 1 we see that

$$S_2 = -\int_0^1 \frac{x \arctan x}{1+x^2} \, dx.$$

Using the variable change $x \mapsto \tan x$ before integrating by parts produces

$$S_2 = -\frac{\pi}{8} \log 2 - \int_0^{\frac{\pi}{4}} \log(\cos x) \, dx.$$

The integral that has now appeared can be evaluated as follows. Recalling Euler's famous log-sine integral

$$\int_0^{\frac{\pi}{2}} \log(\sin x) \, dx = -\frac{\pi}{2} \log 2,$$

and let

$$J_S = \int_0^{\frac{\pi}{4}} \log(\sin x) \, dx \quad \text{and} \quad J_C = \int_0^{\frac{\pi}{4}} \log(\cos x) \, dx.$$

Using the variable change $x \mapsto \frac{\pi}{2} - x$ in J_C gives

$$J_C = \int_{\pi/4}^{\pi/2} \log(\sin x) \, dx.$$

Thus

$$J_S + J_C = \int_0^{\frac{\pi}{2}} \log(\sin x) \, dx = -\frac{\pi}{2} \log 2.$$

Also

$$J_S - J_C = \int_0^{\frac{\pi}{4}} \log(\tan x)\, dx.$$

Substituting $y = \tan x$ we obtain

$$J_S - J_C = \int_0^1 \frac{\log(y)}{1+y^2}\, dy = \sum_{n=0}^{\infty} (-1)^n \int_0^1 y^{2n} \log(y)\, dy.$$

Integrating by parts gives

$$J_S - J_C = -\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \int_0^1 y^{2n}\, dy = -\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} = -\mathbf{G},$$

where we used the definition of Catalan's constant. Solving for J_C yields

$$J_C = \frac{\mathbf{G}}{2} - \frac{\pi}{4} \log 2.$$

Thus

$$S_2 = -\frac{\mathbf{G}}{2} + \frac{\pi}{8} \log 2.$$

The third sum comes directly from the Maclaurin series expansion for $\log(1-x)$ evaluated at $x = -1$. Here

$$S_3 = -\log 2.$$

The four sum is related to the Maclaurin series expansion for $\arctan x$. Here

$$S_4 = -1 + \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = -1 + \arctan(1) = -1 + \frac{\pi}{4}.$$

For the fifth and final sum

$$S_5 = -1 + \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} = -1 + \mathbf{G},$$

where again the definition for Catalan's constant has been used. Combining the values found for all five sums into (11) we find

$$I = 8 - \pi - \frac{\pi^2}{8} + \frac{\pi}{2} \log 2 + 2(\log 2)^2 - 6 \log 2 - 2\mathbf{G},$$

as announced.

Some comments on the general case (part (b) of the question)
Denoting the integral in (b) by I_n where $n \geqslant 2$ is an integer, following the identical procedure that led to (9) we find

$$I_n = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \frac{H_{k+\frac{1}{n}}}{nk+1}.$$

Using such an approach, the general problem therefore boils down to expressing $H_{k+\frac{1}{n}}$ in terms of harmonic numbers whose indices are of integer rather than fractional order before evaluating each of the resulting series. Unfortunately, how this can be achieved for cases when $n > 2$ is currently not clear to me, suggesting that to find I_n, we may need a different approach from the one we used to find I_2.

References

[1] F. W. J. Olver, D. W. Lozier, R. F. Boisvert and C. W. Clark (eds.), *NIST Handbook of mathematical functions*. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge (2010)

[2] C. I. Vălean, *(Almost) impossible integrals, sums, and series*. Problem Books in Mathematics, Springer, Cham (2019)

We would like to invite you to submit solutions to the proposed problems and/or ideas on the open problems. Send your solutions by email to Michael Th. Rassias, Institute of Mathematics, University of Zürich, Switzerland, michail.rassias@math.uzh.ch.

We also solicit your new problems, together with their solutions, for the next "Solved and Unsolved Problems" column, which will be devoted to the topic of Game Theory.

New editors appointed

Karin Baur has been a full professor in the School of Mathematics of the University of Leeds, UK, since 2018. She obtained her PhD in 2002 in Mathematics from the University of Basel, Switzerland. She held an SNF Professorship at the ETH from 2007 to 2011, and since 2011 she has been a full professor at the Department of Mathematics of the University of Graz, Austria (from which she is currently on leave). In 2018 she obtained a Wolfson Fellowship from the Royal Society.

Her research interests lie in algebraic, geometric and combinatorial methods in representation theory, including cluster algebras, cluster categories, module categories, frieze patterns, surface algebras, triangulations and tilings.

She is a member of the scientific committee for the Centre International de Rencontres en Mathématiques, France, and the standing committee of European Women in Mathematics, as well as an editorial advisor for the LMS journals and an editor for the *Springer Lecture Notes in Mathematics*. Her webpage is www1.maths.leeds.ac.uk/~pmtkb/.

Kathryn Hess is a professor of mathematics and life sciences at the Ecole Polytechnique Fédérale de Lausanne (EPFL). She received her PhD from MIT in 1989 and held postdoctoral positions at the universities of Stockholm, Nice, and Toronto before moving to the EPFL.

Her research focuses on algebraic topology and its applications, primarily in the life sciences, but also in materials science. She has published extensively on topics in pure algebraic topology including homotopy theory, operad theory, and algebraic K-theory. On the applied side, she has elaborated methods based on topological data analysis for high-throughput screening of nanoporous crystalline materials, classification and synthesis of neuron morphologies, and classification of neuronal network dynamics. She has also developed and applied innovative topological approaches to network theory, elaborating a mathematical framework relating the activity of a neural network to its underlying structure, both locally and globally.

She is one of the managing editors of *Algebraic and Geometric Topology* and on the editorial board of *Communications of the American Mathematical Society*, the *Journal of Applied and Computational Topology*, *Network Neuroscience*, *Publicacions Matemàtiques*, and *Theory and Applications of Categories*.

In 2016 she was elected to the Swiss Academy of Engineering Sciences and was named a fellow of the American Mathematical Society and a distinguished speaker of the European Mathematical Society in 2017.

Susanne Prediger has been a full professor for mathematics education research at TU Dortmund University in Germany since 2006, and part-time at the German Center for Mathematics Teacher Education (DZLM) at the Leibniz-Institute for Science and Mathematics Education since 2021. She received her PhD in mathematical logic in Darmstadt (in 1998, under the supervision of Rudolf Wille), and a habilitation in mathematics education research in Klagenfurt in 2004.

In Germany, she led the joint commission on teacher education between the mathematics association (DMV) and the mathematics education association (GDM), so she is experienced in maintaining contact between the two disciplines.

She was the president of the European Society for Research in Mathematics Education (ERME) from 2017 to 2021. She is an editor of Educational Studies in Mathematics (ESM) and a member of the executive committee of International Commission on Mathematical Instruction (ICMI-EC).

As a new editor of the EMS magazine, she will focus on the section about mathematics education. Her homepage is www.mathematik.uni-dortmund.de/~prediger/english.html.

EMS executive committee

President

Volker Mehrmann (2019–2022)
Technische Universität Berlin, Germany
mehrmann@math.tu-berlin.de

Vice presidents

Betül Tanbay (2019–2022)
Bogazici University, Istanbul, Turkey
tanbay@boun.edu.tr

Jorge Buescu (2021–2024)
University of Lisbon, Portugal
jbuescu@gmail.com

Treasurer

Mats Gyllenberg (2015–2022)
University of Helsinki, Finland
mats.gyllenberg@helsinki.fi

Secretary

Jiří Rákosník (2021–2024)
Czech Academy of Sciences, Praha, Czech Republic
rakosnik@math.cas.cz

Members

Frédéric Hélein (2021–2024)
Université de Paris, France
helein@math.univ-paris-diderot.fr

Barbara Kaltenbacher (2021–2024)
Universität Klagenfurt, Austria
barbara.kaltenbacher@aau.at

Luis Narváez Macarro (2021–2024)
Universidad de Sevilla, Spain
narvaez@us.es

Beatrice Pelloni (2017–2024)
Heriot-Watt University, Edinburgh, UK
b.pelloni@hw.ac.uk

Susanna Terracini (2021–2024)
Universita di Torino, Italy
susanna.terracini@unito.it

EMS publicity officer

Richard H. Elwes
University of Leeds, UK
r.h.elwes@leeds.ac.uk

EMS secretariat

Elvira Hyvönen
Department of Mathematics and Statistics
P.O. Box 68
00014 University of Helsinki, Finland
ems-office@helsinki.fi

Join the EMS

You can join the EMS or renew your membership online at euromathsoc.org/individual-members.

Individual membership benefits

- Printed version of the EMS Magazine, published four times a year for no extra charge
- Free access to the online version of the *Journal of the European Mathematical Society* published by EMS Press
- Reduced registration fees for the European Congresses
- Reduced registration fee for some EMS co-sponsored meetings
- 20 % discount on books published by EMS Press (via orders@ems.press)*
- Discount on subscriptions to journals published by EMS Press (via subscriptions@ems.press)*
- Reciprocity memberships available at the American, Australian, Canadian and Japanese Mathematical Societies

* These discounts extend to members of national societies that are members of the EMS or with whom the EMS has a reciprocity agreement.

Membership options

- 25 € for persons belonging to a corporate EMS member society (full members and associate members)
- 37 € for persons belonging to a society, which has a reciprocity agreement with the EMS (American, Australian, Canadian and Japanese Mathematical societies)
- 50 € for persons not belonging to any EMS corporate member
- A particular reduced fee of 5 € can be applied for by mathematicians who reside in a developing country (the list is specified by the EMS CDC).
- Anyone who is a student at the time of becoming an individual EMS member, whether PhD or in a more junior category, shall enjoy a three-year introductory period with membership fees waived.
- Lifetime membership for the members over 60 years old.
- Option to join the EMS as reviewer of zbMATH Open.