CIM

CENTRO INTERNACIONAL DE MATEMÁTICA

BULLETIN

ISSN 2183-8062 (print version)
ISSN 2183-8070 (electronic version)

2023 December .45

CIM

Contents

Editorial

In this issue of the bulletin, we pay a heartfelt tribute to Isabel Figueiredo, who passed away last September. Isabel was the President of CIM and a distinguished professor from the department of Mathematics of the University of Coimbra, but, more importantly, she was a dear friend, and we will miss her and her constant good mood. We have invited João Filipe Queiró to write a brief testimony honouring Isabel and her career.

We publish four scientific papers in different areas including Number Theory, Analysis, Geometry and Statistics. Namely, we present an article about Diophantine equations, one about control theory applied to fluid dynamics, one about geometry on surfaces and Higgs bundles and another one linking Bayesian analysis and Hilbert spaces.

Inserted in the cycle of historical articles, we feature an article dedicated to the work of Francisco de Melo and its applications to optics.

We publish two interviews. We interviewed Richard Davis who was the distinguished mathematician invited to deliver this year's Pedro Nunes' lecture, which is an emblematic initiative of CIM with the support of SPM. We also include an interview to Bruno Loff, who was awarded with a European Research Council starting grant in Computer Science.

We also count with several summaries and reports of some of the activities partially supported by CIM.

We recall that the bulletin continues to welcome the submission of review, feature, outreach and research articles in Mathematics and its applications.

Ana Cristina Moreira Freitas

Faculdade de Economia and
Centro de Matemática da Universidade do Porto

https://www.fep.up.pt/docentes/amoreira/

IN MEMORIAM

Isabel Maria Narra de Figueiredo
1959–2023

by **João Filipe Queiró***

This is a brief testimony about Isabel Narra de Figueiredo.

I met her for the first time 46 years ago, when she started her university studies. Hers was a very strong class, with several future university professors and researchers. And already she stood out. Later she obtained her doctorate in Paris, returning then to Coimbra and to a career in which she built a remarkable body of scientific work.

Her subject was always Applied Analysis, moving gradually from problems in Mechanics to medical applications, in imaging and diagnostics, with substantial results and national and international recognition.

I remember the occasion, around 15 years ago, when Isabel came to see me, to talk about a letter she had just received from an Israeli company that wanted to invest in her ideas. I wonder how many Portuguese scientists, especially in a fundamental area, may have experienced this type of situation. Isabel wished to discuss how to deal with the letter. I suggested some people she might contact, both inside and outside the University. Shortly afterwards, she began her long and fruitful collaboration with the firm Critical Software.

She always had a youthful attitude and the corresponding energy in her scientific work. All those who knew her can testify to that.

In the second part of her career, and particularly in the last decade, Isabel added to her research activities — which she never interrupted — a dimension of service to the academic community. Apart from positions in scientific societies, she became President of CIM — the International Centre for Mathematics — an important organization gathering research units from all over Portugal. She was also the director of the Coimbra Mathematics Library, which can seem a minor position until we recall that this is the best mathematical library in the country, and probably in the iberian peninsula.

By her continued efforts, dynamism and scientific activity, Isabel reached a high level of recognition, both in Portugal and abroad. That is the testimony I wish to leave here.

Based on a short improvised speech made at Isabel Narra de Figueiredo's funeral on 8 September 2023.

* Universidade de Coimbra

43rd Conference on Stochastic Processes and their Applications

by **Ana Bela Cruzeiro*** and **Patrícia Gonçalves****

The SPA2023 took place at the Faculty of Sciences of the University of Lisbon, from 24 to 28 July 2023. SPA Conferences are organised under patronage of the Bernoulli Society and can justifiably be regarded as the most important international scientific meeting on the theory and applications of Stochastic processes. They are held annually except for the years when the World Congress in Probability and Statistics takes place.

The SPA2023 had a total of 257 talks (13 plenary sessions, 30 invited sessions, 1 public lecture and 51 contributed sessions with 3 speakers each). The 496 participants represented 44 countries.

It was supported by the University of Lisbon, the Nova School of Science and Technology, the University Aberta, as well as by FCT, IST, ISEG, Faculty of Sciences, SPM, SPE, CIM and various research centres in the Lisbon area.

Local organizers: Fernanda Cipriano, Ana Bela Cruzeiro, Patrícia Gonçalves, Manuel Guerra, Maria João Oliveira, António Pacheco, Beatriz Costa Salvador, Jean-Claude Zambrini.

More information can be found at https://www.spa2023.org/

* Grupo de Física-Matemática e Dep. de Matemática do IST, Univ. de Lisboa. Email: ana.cruzeiro@tecnico.ulisboa.pt

** CAMGSD e Dep. de Matemática do IST, Univ. de Lisboa. Email: pgoncalves@tecnico.ulisboa.pt

AN INTERVIEW WITH

Bruno Loff

by **Carlos Florentino***

Bruno Loff is a Mathematician and Computer Scientist working at the Faculty of Sciences at the University of Lisbon (FCUL). Last year, he was awarded with one of the only three European Research Council grants for research in Computer Science hosted by a Portuguese institution.

After completing the Undergraduate and Master degrees in Computer Science and Software Engineering at the Instituto Superior Técnico (IST), Bruno earned an FCT fellowship (the Portuguese State's Science Institution), and went on to do his PhD at the University of Amsterdam, under the supervision of Prof. Harry Buhrman, on the topic of computational complexity, finishing in 2014.

During the years 2015 to 2020, he held one postdoc position in Prague and another one in Porto. He then obtained a tenure track position at the University of Porto. Returning to Lisbon, in the context of his ERC grant, he is now developing an active group on computational complexity within the Research Center LASIGE, based at FCUL.

I first met Bruno, back in the year 2007, when he was finishing his Master studies at IST, and it is with great satisfaction that I see him reaching high international recognition. Moreover, it is a great pleasure to have him as our newest colleague at the Mathematics department of FCUL. This became the perfect opportunity for an informal interview. In these lines, we delve into the world of Bruno Loff, talk about his research journey, and collect his opinions on Mathematics and Computing.

* CMAFcIO & FCUL

There is certainly no single path to a successful career in research, but everything has a beginning. Can you tell us how and when did you discover your interest in Mathematics and Computer Science?

Ever since I was 7 or 8 years old, I have had this fascination with computers . . . As a child and a teenager, I must have spent more hours in front of a computer than doing any other thing. Gaming, of course, but also coding, 3D modelling, audio editing, and so on. Even today, I'm still in awe of them. Time in front of a computer was replaced with time thinking about computers. Which is my job, nowadays. It just boggles my mind that reality includes these things.

So, during primary and secondary school, did you see Mathematics as a profession, as a challenge, or just as a fun thing? Did you ever think you would dedicate yourself to it?

Neither. The way I was taught Mathematics in high school is the reason I decided against pursuing Mathematics in university, and opted for computer science instead. In my days, and as far as I understand this is still largely the case today, high-school mathematics is taught as if it was some kind of game, where one learns how to do certain calculations.

I was always reasonably good at this game, but I also thought it was a waste of time, because of course: I can just program a computer to do the calculations for me. From this mistaken perspective, that Mathematics is just calculations, one saves a lot of time by learning Computer Science instead.

It was only in my third year of university that I finally understood what mathematical proofs are all about (of course, I had decent grades in all math courses, but understanding proofs is sadly not a requirement for non-math majors). Ironically, nowadays I have to spend a lot of time trying to catch up on all the Mathematics that I didn't learn in my Computer Science education.

I imagine your Master's degree in Computer Science and Software Engineering was very significant for your academic training. Some milestones certainly emerged along the way, as well as people who played an important, or even essential, role in terms of training, or as a source of inspiration. Which mathematicians or professors deserve such a mention and why?

One name before all others: José Félix Costa, my MSc advisor. (José Félix Costa is a professor in the Department of Mathematics at Técnico). It was from him that I first learned the theory of Computability, and then later Computational Complexity. His classes were marvelous: clear, compelling, exciting. I remember being aesthetically moved to tears by Rice's theorem. He somehow managed to teach the solution to Hilbert's 10th problem in a first course on Computability! (Hilbert's 10th problem asked for an algorithm to solve Diophantine equations, and a long line of work by Martin Davis, Hilary Putnam and Julia Robinson, culminating with a result of Yuri Matiyasevich, showed that such an algorithm does not exist). He is also a formidable role model as a researcher, hard working, with a vast knowledge of many fields, extremely ethical when it comes to collaboration and attribution, and driven by a love of the thing itself. Without him, I would be doing something else. He is also a really fun and wild person, and a humble guy, if you can believe it! We remain good friends to this day.

That explains why you chose him to supervise your Master's degree in Informatics. At this moment, were you already thinking about continuing for a PhD?

My mother tells this story . . . When I was 7 years old, my parents would hang out with this couple who had a daughter my age, Eva. Eva's mother once turned to me and asked the typical question: "So, Bruno, what do you want to do when you grow up?" Little 7-year-old Bruno looked up to her and said: "First I want to do a PhD in Mathematics, and then I want to do a PhD in Physics."

My mom says that she hesitated, and said: "Okay . . . so in the meantime would you like to go and play with Eva?" I have since gained some sense and have no plans to do a PhD in physics.

Was it obvious that you should obtain the doctorate abroad? Was it easy to choose the advisor (Harry Buhrman) and the thesis problem? How was your adaptation to another country, culture, way of teaching and studying?

I was excited to go abroad! New things! An adventure! I looked for people doing computational complexity in Europe, and Harry's name came up. I sent him a letter, said I would be coming with my own funding, and he took me in. I moved to Amsterdam on September 2008.

And within 4 months, I had entered a deep state of depression, including debilitating anxiety, panic attacks, mild paranoia and hallucination, and a complete lack of concentration, which made it very difficult to work. I also had brief periods when I was on top of the world, master of my game, had figured it all out, etc. A psychiatrist diagnosed me as bipolar, which presumably I had inherited from my mother, and which maybe, just maybe, was triggered by an LSD trip where I realized, in some deep, undeniable, immediate, visceral way, that every moment in time is dying all the time, and I myself am going to die some day.

The psychiatrist prescribed me antipsychotic medication. After reading about the side effects of the drug I had been prescribed, I decided against taking it, and started doing meditation instead. I did a vipassana retreat in August 2009, where it became clear that meditation really affects the condition I was in.

So by November 2009, I had all but decided to quit my PhD to become a Buddhist monk, when Harry gently suggested that I should take a temporary break instead, "do the meditation thing for a while", and see if I wanted to come back after that. So I asked for a temporary interruption of my grant, which FCT allowed. I did a solitary retreat in December 2009, and on the 30th I experienced a shift in my perception, and was never depressed again in the same way. With continued practice, eventually my emotions balanced out, and I have not experienced euphoria or depression since about 2014.

But sorry, I got a little sidetracked. I got a lot out of working at CWI (Center for Mathematics and Computer Science, in Amsterdam). They have a very strong scientific culture over there.

A doctoral thesis is the beginning of a research career. Right after finishing the thesis, were you prepared for this challenge? Do you think that, to gain experience, one or two postdoctoral positions are fundamental?

After my PhD, I had a lot of self-doubt and seriously considered giving up science. My PhD thesis is entitled *A Medley for Computational Complexity*. I.e., I had a bunch of disparate results and I stapled them together. It had always felt that every single result I discovered was a stroke of luck. I didn't really think I could turn such random events into a career. The word career suggests a straight line, of sorts, a natural uphill progression.

Again Harry offered me good advice: he said that giving up was completely fine, but he thought I was doing OK, and maybe I should give it a chance? So that was the second time a conversation with Harry pulled my career from the brink.

So I decided I should give it a fair shot, and if it flopped, then I had done the best I could. I contacted Michal Koucký, who had visited Amsterdam a couple of times, and proposed that we work on a particular problem (dynamic data structures for directed connectivity). I moved to Prague, and lived there for two years, with a lifestyle of a mathematical monk. I got up early, I went to my Tai Chi practice, and went to the office, where I would work until late. Next day, repeat. My bipolar disorder was gone at this point, so my concentration was back. I learned a huge amount. (But we never solved the above problem. It turns out to be a formidable problem.) I would never be able to do the research I do without those highly focused five years of postdoctoral research.

Oddly, each and every result I discover still feels like a stroke of luck. I'm just more used to it, I guess.

What do you consider to be your most relevant scientific contribution up to now, and why?

I think my favorite own paper, thus far, is *Computing with a full memory: catalytic space*. Quoting directly from the intro: "Imagine the following scenario. You want to perform a computation that requires more memory than you currently have available on your computer. One way of dealing with this problem is by installing a new hard drive. As it turns out you have a hard drive but it is full with data, pictures, movies, files, etc. You don't need to access that data at the moment but you also don't want to erase it. Can you use the hard drive for your computation, possibly altering its contents temporarily, guaranteeing that when the computation is completed, the hard drive is back in its original state with all the data intact? One natural approach is to compress the data on the hard disk as much as possible, use the freed-up space for your computation and finally uncompress the data, restoring it to its original setting. But suppose that the data is not compressible. In other words, your scheme has to always work no matter the contents of the hard drive. Can you still make good use of this additional space?"

Surprisingly, the answer is yes! It is possible to use full memory in a non-trivial way!

After the post-doctoral positions in Prague and Porto, you obtained an Assistant Professor position at the Faculty of Sciences of the University of Porto. How was the experience in Porto, in particular the need to balance teaching and research?

I was very lucky, because I got a CEEC grant (Concurso de Estímulo ao Emprego Científico). This grant disallowed the university of assigning me more than 6 hours of teaching duties per week. Even then, teaching made research significantly harder than it was during my postdoc years. It is very sad that there isn't really a research career in Portugal, and that so many researchers are working under precarious employment contracts. Some kind of solution really needs to be found.

In a world of research that tends to be very competitive, how did the idea of applying for an ERC grant come about? What aspects of the application were decisive for the positive evaluation?

Actually I did not plan and did not want to apply to an ERC grant. But Michal Koucký insisted that I should. Once I started working on it, I had a vision of what I wanted to do, and I wrote it down.

Well, I started writing a grant as usual, and then at some point I realized that this is a much bigger grant than FCT grants, so I threw away the few pages I had and started over. I also decided I would try to solve a difficult problem that people in my area care about, because why would anyone care otherwise? I also decided that people in the committee were probably really smart, so I would be brutally honest. I remember during the interview they asked me: "so, what applications do you think might come out of this project?" To which I promptly answered "probably none", and felt really dumb afterwards. But actually, I suspect that the committee knew as much, and they were testing whether I would reply honestly.

Curious fact: The project was awarded the ERC, which means the project was in the top 10% of the Computer Science projects submitted to the ERC that year. Well, between submitting to the ERC and getting the acceptance letter, I took the same project, trimmed it down to work with 1/5th of the budget, and submitted it to FCT. It was classified in the bottom 10% of the Computer Science projects submitted to FCT in that same year. This probably happened because that year, like, sadly, in most years, the FCT evaluation committee did not include people from Theoretical Computer Science (the ERC committee had several).

There also, I feel that something needs to change. Computer Scientists think of me as a Mathematician, and Mathematicians think of me as a Computer Scientist.

Amazing! . . . Ok, tell us a little about the objectives of your work plan for these five years. Do you think it is possible to achieve most or all the goals?

I will try, but the project is very ambitious. In Computational Compleixty, lower bounds are impossibility results showing that certain computational problems cannot be solved efficiently in a certain computational model. Some computational problems are harder than others, and some computational models are stronger than others. We know how to prove lower-bounds either for very hard problems no one cares about, or in very weak models no one cares about. The goal of the project is to prove lower- bounds in new ways. If we fail to do this, we would like to understand why we failed at it.

Since one of the goals is to advance on the resolution of the famous millennium problem: the "P versus NP" problem, we cannot resist asking: how close do you think we are to finding a solution? Will the answer be positive or negative, or will it be one of the undecidable questions, as in Gödel's incompleteness theorem?

This is a good question. Something which is not well understood outside Computational Complexity is that we know of a very good reason why lower bounds are hard to prove. For a long time it has been believed that there are functions which are easy to compute but hard to invert. So there exists an efficient algorithm for computing $f(x)$ when given x, but, simultaneously, there is no efficient algorithm to find x when given $f(x)$. Such one-way functions are known to exist in any sufficiently powerful computational model. For example, multiplication of natural numbers can be computed efficiently, but we do not know of any efficient (non-quantum) algorithm for factoring natural numbers.

Alexander Razborov and Stephen Rudich observed in the 90s that essentially every lower-bound proof technique that was known up to that point had a certain kind of structure. They called proofs with this kind of structure Natural Proofs. So all lower- bound proofs known at the time are natural, and this is still very much true today, with few, in my opinion not very relevant, exceptions. They then showed that natural proofs cannot be used to show lower-bounds against any computational model strong enough to compute one-way functions. It is a kind of independence result. We call it the natural proofs barrier.

So look at the difficult situation we were left with: we cannot prove lower-bounds by natural proofs on any model powerful enough to compute, say, multiplication. And yet every lower-bound proof we know is a natural proof. The big question is how to overcome this barrier.

Let's talk more about the research experience. How do you discover interesting and good problems to work on? And for solving them, are there methods or strategies that may be more effective?

A math problem is like a chronic disease, I don't go looking for them, they find me and won't let go, unless by chance I find the cure by solving the problem. I wish I knew of some effective general approach that works. I feel completely stuck 99% of the time. It's a very frustrating profession, at least for me.

But I should add, of course, under the conjecture that P is not equal to NP, there does not exist any method or strategy that will be effective 100% of the time at solving math problems. Under a slightly stronger complexity-theoretic conjecture, e.g. that k-SAT or CLIQUE are hard on any sufficiently random efficiently samplable distribution, there does not exist any method or strategy that is effective even 1% of the time.

Of course, I'm totally stuck at proving these conjectures, hehe.

Even so, doing research is certainly satisfying and rewarding. What do you think is fascinating in the field of Mathematics?

Speaking for myself, I learned Mathematics because I wanted to understand computers. Understanding is the highest form of love, and I love computers. Computers were invented by Alan Turing, a mathematician, not a physicist, or an inventor, or anything more applied. And there is good reason why this was the case: the computer is the most mathematical of all human-made objects. So Mathematics drew me in. It took me some time to realize that: to understand computers is to understand lower-bounds. We understand algorithms quite well, i.e., we understand what computers can do very well. We are good at coming up with new algorithms. But we really don't understand computation, because we cannot prove lower-bounds, i.e., we cannot understand the things that computation cannot do. That understanding can only come from Mathematics.

Of course, Mathematics is beautiful and fascinating. And I have known of people who do Mathematics as a kind of leisurely stroll, just looking out over beautiful vistas, smelling each nice flower they come across. But I have a very goal-oriented approach to Mathematics, I want to get something out of it, and this adds a certain degree of stress. Maybe one day Mathematics will finally help me understand computation, and then I will be able to relax more into it.

Research collaborations are obviously important, both in terms of work and visibility. On the other hand, management of a grant adds responsibility towards the colleagues that work with us. What is the most important characteristic that a collaborator of yours needs to have? How important it is to have PhD students and develop a research group on your own main topics?

Skill is important, of course, but the most important characteristic of a collaborator of mine is, without a doubt, a love for what they do. You would think that this is easy to come by, but so many researchers have their egos wrapped up in their work, with their love of Mathematics soiled by a stronger desire to be a great Mathematician, or something along those lines. I am sad to say that I also have a big ego, but I do sincerely strive my best to keep an above-unit quotient of love-for-mathematics over ego (He said, in the magazine interview he accepted to participate in. It's work-in-progress).

How do you see the relationships, differences and similarities between research in Mathematics, Computer Science and Informatics?

There is one fundamental thing that good Mathematicians and good Computer Scientists have in common: an understanding of what it means to be precise. A mathematical statement is precise in very much the same way that an instruction in a computer program is precise. A good programmer can easily be taught what a proof is, and a good mathematician can easily be taught how to code. I work in Theoretical Computer Science, which studies computers with the methodology of mathematics. So I don't really know how research happens in applied computer science. But I can say this: applied Computer Science is a discipline that entails very many non-mathematical skills. There is little mathematics going into requirements analysis, software architecture, interface design, testing, deployment, load balancing, etc. Not to mention management skills and people skills, all of which are necessary to produce a usable, reliable software product. Naturally, research in applied Computer Science includes all of these things, it's a whole other world.

What advice would you give to young mathematicians just starting their research careers? What specific skills or competences are essential for success in Mathematics research?

Wow, some advice for a young mathematician, let's see... You're probably pretty smart, try not to be a dick about it. Create and nurture a circle of mathematical friends. Work with researchers with all levels of skill. Get used to the feeling of being stuck, it will be with you for your entire mathematical life.

How do you balance your work as a mathematician with other aspects of your life? Are there hobbies or activities that are particularly enjoyable or essential to maintaining a healthy work/life balance?

I wish I knew. There is really no balance. Sometimes I'm so engrossed in a math problem that I get insomnia, spend hours awake at night thinking about it, and get out of bed exhausted the next day. I have forfeited entire holidays, quality time with loved ones, because I had some idea and couldn't let it go. Exercise helps a lot. I still meditate. But the job takes its toll. I have been noticing, lately, that I find it extra annoying when people use imprecise language in our day-to-day lives. It's such a silly thing to be annoyed about, but one's job shapes one's mind. And it is so isolating, to work in a field so abstract that you can't explain what you do to your friends, your partner, your family. It really is a labor of love, as nothing else would justify the sacrifice.

Mathematics is generally considered a difficult and hermetic subject. How could we make Mathematics more accessible and engaging to a wider audience, including students and the general public?

To be engaged with Mathematics, people have to experience the pleasure of understanding it. Maybe not for geniuses or whatever, but for the rest of us Mathematics is an acquired taste, and in that way it is not like hot chilies or sour pickles: it is particularly difficult to acquire because it takes a lot of time. There is no magical substitution for time spent together with people who already love mathematics, in a place that is suitable for it. A teacher, a friend, a desk, a classroom, a club. Yes, I would say, we need more math clubs.

Do you think Mathematics can play an important role in solving real-world problems, and contributing to facing global challenges?

No, of course not.

Seriously, though? The universe is made of the stuff. To forget mathematics is to forget a fundamental ingredient that everything is made with. I suppose a fish does not need to know what water is, if all he wants to do is be a fish, as God intended. But if the fish wants to have any semblance of control over his surroundings, he will need to understand water. There is no civilized world where Mathematics doesn't play a very important, fundamental role.

Having said that, none of the most important problems facing humanity today are mathematical problems. Just a few days ago ended COP28, which was held in one of the worlds largest fossil-fuel exporting nation, and, one might say unsurprisingly, resulted in a multinational agreement far less ambitious than what our climate scientist colleagues say is necessary to maintain global warming below 1,5 C°. I would happily give up my job as a mathematician if, in magical exchange, all humans everywhere would be 10% more reasonable.

CIM Bulletin

Editor-in-Chief

Ana Cristina Moreira Freitas [amoreira@fep.up.pt]

Editorial Board

António Fernandes* [antonio.marques.fernandes@tecnico.ulisboa.pt]
Carlos Florentino [caflorentino@fc.ul.pt]
Samuel Lopes [slopes@fc.up.pt]

*Responsible for the production and graphic design of the bulletin

Address

Centro de Matemática da Universidade do Porto
Rua do Campo Alegre, 687
4169-007 Porto
Portugal

The CIM Bulletin is published once a year. Material intended for publication shoud be sent to one of the editors. The Bulletin is available at: www.cim.pt. The CIM acknowledges the financial support of FCT- Fundação para a Ciência e Tecnologia.

The influence of greek authors on Francisco de Melo's theory of vision

by **Daniel Pinto***

In the history of science, of any branch of science, there are no periods of complete stagnation. One can always find at least a few small advances or a couple of false steps that turn out to be crucial. In any case, even taking into account the growing number of studies on perspective in painting, which revived interest in various topics of geometric optics, it would not be out of place to classify the 14th and 15th centuries as two reasonably quiet centuries in the history of optics. So when we look at the theory of vision presented by the Portuguese mathematician Francisco de Melo (born about 1490), it is unsurprising that his approach is still deeply linked to the dominant theories in the Middle Ages. The influence of Ibn-al Haytham, also known by the Latinized name of Alhacen, who wrote, in the 11th century, one of the most groundbreaking treatises on optics, and the impact of Witelo, a 13th-century author, on Melo's thinking would be easily detectable even if he had not referred to them. Still, the truth is that Melo mentions the names of both in the *Corollary to Euclid's Perspective*, a text dedicated to the nature and fundamental principles of vision (and written to complement Melo's version of Euclid's *Optics*). But besides his interest in the medieval period, Francisco de Melo, who for some years attended the University of Paris, back then one of the most important academic centres in the world, was nonetheless a man of his time, strongly influenced by the Renaissance movement that sought to revisit ancient Greek texts. Here we will try to display a significant number of concepts that Melo retrieved from Greek authors to build his theory of vision.

The intersection of ideas, the succinct way in which Melo condenses and cross-references, in a few pages, many of the main currents in optics known until then, is one of the most peculiar facets of the *Corollary to Euclid's Perspective*. That whole process of synthesis is backed up by Melo's tendency towards abstract thinking. He also makes use of experimental results and some concrete examples, but it is in the manipulation of mathematical tools that he shows greater ability. To write his version of Euclid's *Optics*, Melo completely reworked the very unclear demonstrations that were available in the Latin edition (translated from Greek) by Bartolomeo Zamberti, printed a few years earlier (1505). He did not just make small adjustments; almost all the demonstrations were rebuilt from scratch with remarkable detail and precision. Even though, to complete that task, Melo may have followed some of the comments of Pierre Brissot, with whom he worked in France, such an ambitious programme would have proved impossible had Melo not been a talented mathematician himself. In the *Corollary to Euclid's Perspective*, maybe because of the hybrid nature of the text (a combination of geometry, anatomy and natural philosophy), he is not so original or audacious in his demonstrations, but it is still possible to find substantial differences when we compare them with proofs of similar results in previous works on the subject.

Euclid

In the last paragraph of the *Corollary to Euclid's Perspective*, Melo warns the reader that the theory of vision he had elaborated had been quickly put together and that it may not be perfectly articulated. This self-inflicted depreciation of his own text is intended not only to point out that Melo did not invest much time in dealing with

* Centro de Matemática da Universidade de Coimbra
Email: dpinto@mat.uc.pt

Figure 1.—Johann Zahn, *Oculus Artificialis Teledioptricus Sive Telescopium*, 1685

the details of a complex field, that he wants to summarise in an original way, but also to draw attention to the fact that his version of Euclid's *Optics* is a much more ambitious project. Although the *Corollary to Euclid's Perspective* could be read as a completely autonomous text, it would be disconcerting if the main ideas defended by Euclid in the *Optics* had not been incorporated into Melo's theory of vision. But that is not the case. Euclid (4th–3rd century BC) is possibly the strongest presence in Melo's theory of vision. This is evident in the structure of the text, which is very Euclidean, with two postulates followed by propositions and lemmas, as well as in the more substantive content.

Euclid is considered an extramissionist since, in one of the definitions of the *Optics*, he admits the existence of rays travelling from the eye to the objects, despite not using this orientation in his proofs (Figure 1). For centuries, even after Kepler, the extramission (or emission) hypothesis had many supporters who rejected the intromission theory, in which rays emitted by the eye were not necessary to explain vision, only rays travelling in the opposite direction (a position advocated, for example, by Alhacen). Melo agreed that something had to reach the eyes, but his theory of vision does not dispense with the emission of visual rays. Furthermore, Melo is also aligned with Euclid in terms of both the rectilinear and

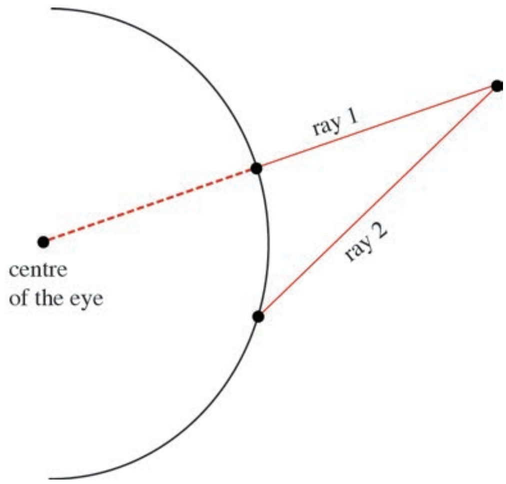


Figure 2.—Ray 1 (the perpendicular ray) is shorter than ray 2.

the discrete nature of those rays. For Melo, visual rays are not continuous, there are intervals between them, which would explain why, in an instant, we might not see a small needle on the ground (fallen between two consecutive rays) but, by moving our eyes a little, the needle might appear to us (after being hit by some visual ray). In Euclid's *Optics*, the eye is represented by a point, and binocular vision is not considered (with a few exceptions). Melo does not devote much time to this subject either. He only directly deals with binocular vision in three of the twenty propositions.

Even when he addresses anatomical issues, which are absent from the Euclidean text, Melo transfers the analysis of the components of the eye to a geometric setting. It is undoubtedly Euclid's *Optics*, the earliest extant treatise on geometric optics, that Melo has as his primary reference. And if the style of Euclid's *Optics* is present in the geometrisation of the anatomical features, the *Elements* are called upon in the course of various demonstrations—particularly the results from Book III—involving circles and circumferences.

Aristotle

Although Aristotle is explicitly mentioned several times, his most relevant contribution to Melo's work is tacitly included in the first postulate of the *Corollary to Euclid's Perspective*, formulated with an undeniable Aristotelian slant:

> Firstly, it must be accepted that every natural agent acts more quickly and vigorously towards what is near than towards what is far.

It is this postulate that will allow Melo, later on, to establish that faithful and distinct vision is realised by means of rays that fall perpendicularly on the eye, the ones with the shortest length (Figure 2). If all the rays, and not just the perpendicular ones, were of equal importance, the image in the eye would appear confused, which does not occur in a person without ocular disorders.

There are passages in Aristotle´s body of work where he seems to be moving closer to the extramission theory, namely in his studies of the rainbow. But, unlike Euclid, he is generally connected, by commentators, with the intromission theory. Melo is not an exception and also associates Aristotle with the idea that something from a visible object must reach the eye. However, in the *Corollary to Euclid's Perspective*, Francisco de Melo often mentions Aristotle in paragraphs in which he wants to reaffirm that his theory of vision does not dispense with rays emitted by the eye. For instance, Melo uses an observation that he partly attributes to Aristotle (why do we see further and more clearly through a tube or with half-closed eyes?) to reinforce his appetite for extramission theories. According to Melo, this is due to the greater number of rays that hit the object since, under normal conditions, some of them would disperse. Curiously, the oddest reference to Aristotle is also related to extramission. For Melo, something must be travelling from the eyes to the objects since, following an observation from Aristotle's *On Dreams*, menstruating women allegedly infect mirrors.

Despite the fact that Melo does not always stress the relevance of colour in his theory of vision, he seems to agree with Aristotle for whom colour was not only a characteristic of objects (not dependent on the observer or other factors) but precisely what makes them potentially visible. As for the importance of the eye, Melo is again in tune with Aristotle, considering vision to be dominant over the other senses.

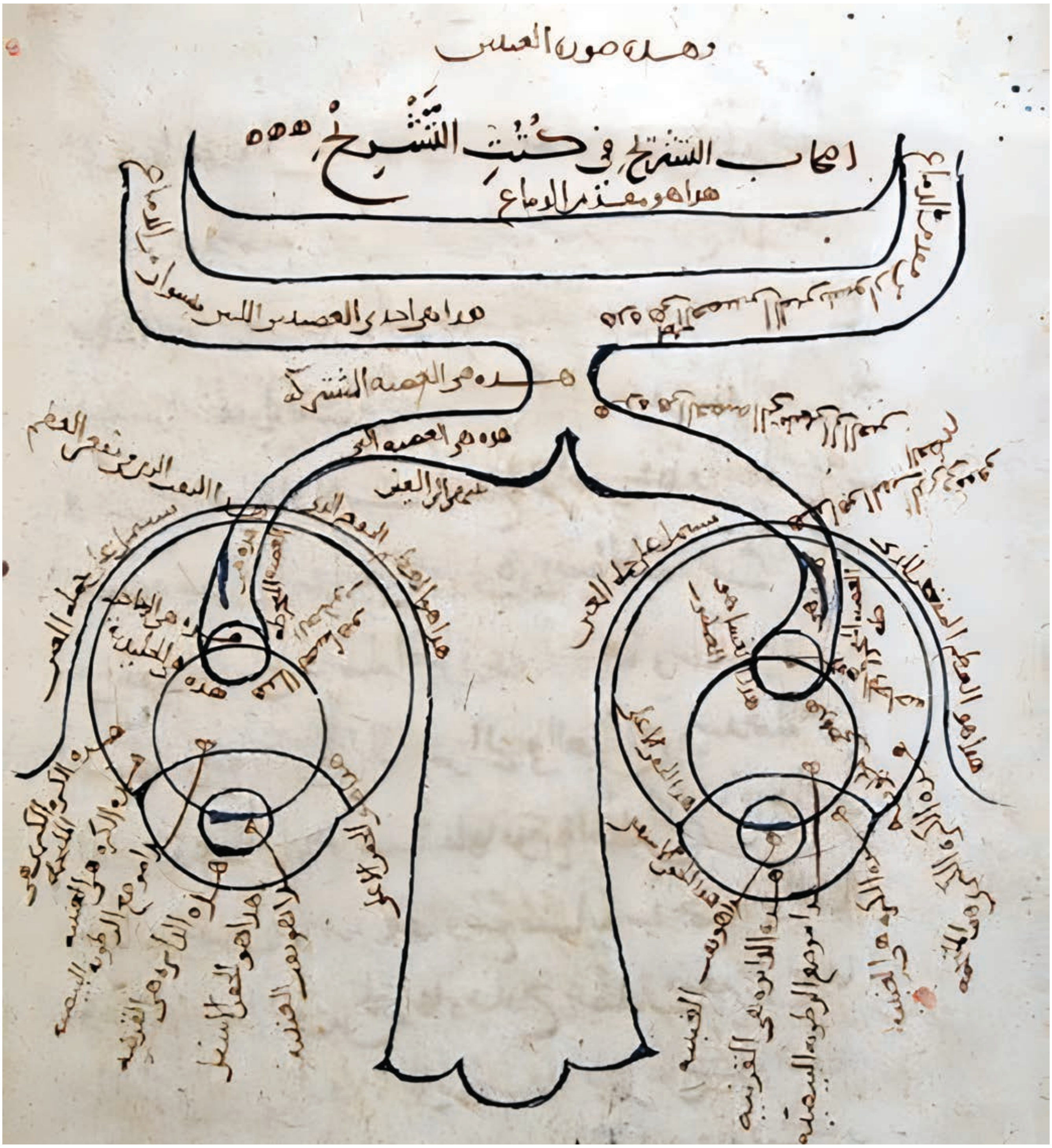


Figure 3.—The eye according to Alhacen (Ibn al-Haytham). MS Fatih 3212, vol. 1, fol. 81b, Süleimaniye Mosque Library, Istanbul.

Plato

With Plato, as with Euclid, Francisco de Melo shares the belief in the extramission hypothesis. An affinity that Melo emphasises in his text after remarking that the eyes of many living beings glow in the dark. Melo is, in some sense, very close to the idea of *visual fire* that we can find in the Platonists. But Plato's most significant influence on Melo's theory of vision is related to the role of light. Contrary to Alhacen's approach, light is not at the centre of Melo's theory, but it is essential for vision to occur. For Plato, vision is only possible if the visual fire combines with light, an insight that Melo, with some adaptations, also embraces. In the view of Aristotle, light is a state that requires the presence of some luminous body so that a *potentially transparent homogeneous medium* (that can be found especially in air and water) could become *actually transparent*. As opposed to Plato, for Aristotle this actualisation is a qualitative change, there is no movement involved. On this topic, Melo seems to be closer to Plato, although his position is somewhat ambiguous.

Galen

Francisco de Melo's interests were not only focused on the nature of visual rays, the importance of light, or the more abstract concepts in the background. The anatomical details of the eye were also the object of his attention. For Melo, the eye is made up of three tunics (Cornea, Uvea and Arachnoid) and three humours (Albugineous Humour, Vitreous Humour and Crystalline). Melo justifies the absence of the Conjunctiva, which Galen (2nd century AD) includes in his description of the eye, because it is an external part that does not interfere in the process that leads to vision, despite its important function of connecting the eye to the bone in the head. In the manuscripts of the *Corollary to Euclid's Perspective* that have survived, the figure representing the eye is missing. However, one can understand, by Melo's description, that the anatomy of the eye that he proposes is not only inspired by Galen's but also includes other later contributions, in particular the one that Alhacen popularised in his most famous book, *De aspectibus/Kitāb al-manāzir* (Figure 3).

Contrary to what happens with the Conjunctiva, about which they have at least a formal discrepancy, Melo and Galen agree that the seat of vision is located in or around the Crystalline (the *lens* of the eye), an idea that Aristotle also defended.

Theodosius of Bithynia

Since Melo's approach is very geometric, many of the propositions involving the eye are actually results about spheres. To justify the relative position of the components of the eye, or the shape of the common sections of the humours that make it up, Melo resorts to geometry. Sometimes through results that he himself demonstrates, in other occasions using propositions from Theodosius' *Sphaerics*. An example of the first case is the Lemma in which Melo proves the following:

> If two unequal circles intersect, each will be divided into unequal arcs, and the smaller arc of the larger circle will be contained within the smaller circle. In the same way, two unequal spheres will not be cut into equal parts, and the smaller section of the larger one will be contained within the smaller one.

Melo's demonstration makes use of some results from Euclid's *Elements* and also of Proposition Fifteen, which he has proved earlier. The referred proposition can serve us as an example for the second case, since in order to show that a particular line passes through the centre of certain spheres (that are abstract representations of some components of the eye), Melo uses results from Theodosius (2nd century BC) to shortcut the argument. According to some authors, including Thomas Heath, Theodosius was not a particularly original mathematician. Heath goes so far as to describe him as nothing but a *laborious compiler*. Nevertheless, the theorems and proofs from Theodosius' textbook appear in important works on geometric optics, not only in that of Francisco de Melo. Despite Witelo never mentioning Theodosius in his treatise on optics (Witelo's *Perspectiva*), the similarity of various results and proofs to those we can find in *Sphaerics* indicates that he had a thorough knowledge of Theodosius' work.

Main References

Aristotle, *On Dreams*, 2.459a23–28.

Harry Edwin Burton, *The optics of Euclid*. JOSA, 35(5):357–372, 1945.

Thomas Little Heath, *A history of Greek mathematics*. Cambridge University Press, 2013, vol. 2.

David C. Lindberg, *Theories of Vision from al-Kindi to Kepler*. Chicago: University of Chicago Press, 1976.

Francisco de Melo: obras matemáticas, ed. by Bernardo Mota and Henrique Leitão. Biblioteca Nacional de Portugal/Centro de Estudos Clássicos, 2014, vol. I

A. Mark Smith, *Alhacen's Theory of Visual Perception: A Critical Edition, with English Translation and Commentary, of the First Three Books of Alhacen's De Aspectibus, the Medieval Latin Version of Ibn Al-Haytham's Kitab Al-Manazir*. American Philosophical Society, 2001, vol. I.

A Mark Smith, *From sight to light*. University of Chicago Press, 2014.

Witelonis Perspectivae Liber Primus: Book I of Witelo's Perspectiva, ed. by Sabetai Unguru. Wroclaw, Warszawa, Kraków, Gdansk: Ossolineum, 1977.

REPORT

7^{th} Workshop New Trends in Quaternions and Octonions — NTQO 2023

by Alexandra Baptista*, Ana Mendes*, Milton Ferreira* and Nelson Vieira**

The Seventh International Workshop New Trends in Quaternions and Octonions — NTQO 2023, took place from October 27th to 28th, 2023, at the School of Technology and Management (ESTG) of the Polytechnic of Leiria (IPLeiria) and at the M|I|MO Museum located in the center of Leiria.

The event was organized by the Department of Mathematics of the ESTG and the Center for Research and Development in Mathematics of the University of Aveiro (CIDMA-UA), in collaboration with the Center of Mathematics of the University of Minho (CMAT-UMinho), the Center of Mathematics and Applications of the University of Beira Interior (CMA-UBI), and the International Center for Mathematics (CIM).

This series of workshops aims to bring together researchers from pure and applied mathematics, physics, scientific computing, and engineering, to present recent advances in the research on quaternions and octonions, and their applications. In this edition were presented several short communications, and four invited lectures by Alessandro Perotti (University of Trento, Italy), Aleš Ude (Jožef Stefan Institute, Slovenia), João Pimentel Nunes (Instituto Superior Técnico, Portugal) and José María Pérez Izquierdo (University of La Rioja, Spain), see

https://sites.google.com/view/ntqo2023

* Department of Mathematics of the School of Technology and Management of the Polytechnic of Leiria.
Emails: alexandra.nascimento@ipleiria.pt; aimendes@ipleiria.pt; milton.ferreira@ipleiria.pt
** Dep. of Mathematics of the University of Aveiro and CIDMA-UA. Email: nvieira@ua.pt

AN INTERVIEW WITH

Richard A. Davis

by **Miguel de Carvalho***

Richard A. Davis (born 11 September 1952) is a world authority in the fields of Time Series Analysis and Extreme Value Theory. He holds the position of Howard Levene Professor of Statistics at Columbia University. He earned his PhD in Mathematics from the University of California at San Diego in 1979. Throughout his academic career, Davis has served in various capacities at institutions such as MIT, Colorado State University, and Columbia University, in addition to having visiting roles at a range of other universities (e.g., Hans Fischer Senior Fellow at the Technical University of Munich 2009–12; Villum Kan Rasmussen Visiting Professor at the University of Copenhagen 2011–13; Chalmers Jubilee Professor at Chalmers University of Technology 2019).

* School of Mathematics, University of Edinburgh; Departament of Mathematics, Universidade de Aveiro

Davis is a Fellow of the Institute of Mathematical Statistics and the American Statistical Association. He has also been recognized as an elected member of the International Statistical Institute. In 2016, he served as the president of the Institute of Mathematical Statistics, and from 2010 to 2012, he held the position of Editor-in-Chief for the Bernoulli Journal. He co-authored the widely acclaimed books *Time Series: Theory and Methods* and *Introduction to Time Series and Forecasting* alongside Peter Brockwell, and also developed the ITSM2000 software for time series analysis. Davis also co-edited the *Handbook in Financial Time Series* and the *Handbook of Discrete-Valued Time Series*. In 1998, he and W. T. M. Dunsmuir were awarded the Koopmans Prize for their contributions to Econometric Theory.

A mentor to more than 30 PhD students, his research spans time series, applied probability, extreme value theory, and heavy-tailed modeling, with a focus on network models and spatial-temporal modeling.

Richard, can you tell us about your early life?

I grew up in a small town outside of Ann Arbor, Michigan. Our family lived on a lake, where boating and related activities occupied a big part of my childhood at least during the summers. My mother was born and raised in the New York City borough of Brooklyn and she headed west for college to attend the University of Michigan (UM). There, she met my father, married, and they ended up settling on the outskirts of a small town called Pinckney to raise a family. Schools in these rural areas were limited in their academic standards and my mother coming from New York City, where education tended to be highly valued, she made arrangements for her children (4 boys of which I am the youngest) to attend the UM Laboratory School in Ann Arbor. This was not an easy commute as our home was located nearly 20 miles away. The UM school was a training ground for K-12 grade teachers and curriculum development at UM's School of Education. So we were lucky especially in later grades to have professors and student teachers from the school of education to experiment with curriculum and modern teaching techniques. So I think I really got my academic start going to this special school.

How important was this stage for your passion for Mathematics?

Primary schools in rural areas in the 1950s were not particularly good and by making the move to the University School, I was exposed to a more modern and engaging educational experience. The teachers were extremely good, especially the math teachers who either had their PhDs (or were pursuing PhDs) in math education. They were inspiring and novel in their approach to teaching mathematics (as well as other subjects) for which I seem to have some skill. In particular one math teacher used our class as a testing ground for his forthcoming book on algebra and geometry, and as my brother reminded me recently, I took great pride in finding the most mistakes (misprints?) in the book. In any event, these young and energetic math teachers clearly activated my interest in mathematics.

Did a particular moment significantly influence your interest in Mathematics?

Although I loved playing math games and was very good at them, I never imagined making a career out of math. I began my college studies at Michigan State University (MSU), which might seem kind of strange growing up in Ann Arbor, the home of UM. I started college as an undeclared major and at some point in time, the School of Education, which was at capacity for majors, opened up their program briefly for new majors. Since my mother was an elementary school teacher and it was time to declare a major area of study, I chose, in a moment of indecision, to declare math education as my major. This seemed like a safe choice, but teaching at the high school level was something I really didn't want to do as a career choice. Math education was a short-lived selection. During my high school and college years, I was heavily involved with sailboat racing. I had worked for a sail making company during high school and my brother and I raced sailboats all over the country and even overseas a few times. So I became very good at it. I was sailing almost every single day and sailed for the MSU sailing team. In the summer of 1972, I competed for MSU in the collegiate national championships held on Mission Bay in San Diego. Coming from Michigan, this was the most idyllic place I had ever seen. After the championship regatta, I immediately applied to the University of California at San Diego (UCSD), mainly to further my sailing career. I transferred to UCSD in the middle of my junior year and began racing for the UCSD team. Since UCSD did not have an education major, I began as a full-fledge math major. I did the usual

topics of a math major, real and complex analysis, logic, partial differential equations, and algebra, but definitely no probability or statistics!

Graduate School was your next step — why?

After graduating from UCSD, I decided to continue to live in San Diego in order to further my sailing career and also because I just didn't really know what to do next. I was admitted into the PhD program at UCSD, and was supported as a teaching assistant (TA). During the first year of graduate school, I took the basic core PhD courses such as abstract algebra, complex analysis, and applied math. After doing well in year 1, I was trying to decide what to do for year 2 — I really had no clue — whereupon my office partner and fellow graduate student, Gail Gong, told me I should take statistics. She heard there's a future in statistics with plenty of job opportunities. So I followed Gail's advice and after talking to the first year graduate advisor, who was a logician and had no clue about statistics, I signed up to take the PhD level mathematical statistics course.

And Rosenblatt, who would become your supervisor, was teaching that course?

Yes, indeed. I hadn't taken any courses in probability and statistics previously so I decided to meet with Murray Rosenblatt before the summer break to see what sort of preparation I could do. I would be spending the summer sailing in regattas, one of which was the World Championships in France, so I would have to study while training and preparing for the upcoming racing season. Murray, who was very supportive and encouraging, suggested I have a look at Cramér's book, *Mathematical Methods of Statistics*. So I read the book over the summer, which was quite the challenge especially for someone without a rudimentary background in probability and statistics. I made it through the first quarter of the course. During the second term, Murray asked me if I would be willing to read a paper on extreme value theory written by Ross Leadbetter. Of course, I said yes, but really didn't know exactly what he wanted me to do with the paper. I was naive and had no clue about how the publication and refereeing process for research papers worked. I came to learn that

the Leadbetter paper was being considered for a special volume that Murray was editing and that he was using me as a referee. So in a way, this might have been a screening test of sorts to see if I had any ability to read papers and provide reasonable feedback. Leadbetter's paper contained some open questions, and so I started to think about solving them. And then a month or so later, Murray asked, "Do you want to work with me?" I was shocked that anybody would be interested in having me as their student and so I just said yes immediately. I later heard from other students, who were much more connected than me, that Murray was well-known in the probability and statistics community and that I had made a good decision without having done any sort of due diligence. I was just happy that someone was interested in being my advisor. As it turns out, there was a second professor also interested in having me as a student. In my second year of graduate school, I also took a year long course in noncommutative ring theory. Why I did this, I have no idea! Not more than a couple weeks after I had committed to Murray, the algebraicist teaching noncommutative ring theory asked if I wanted to do noncommutative ring theory with him! The timing was close, but I was relieved that Murray had already approached me-- noncommutative ring theory was not going to be for me. So that's how I got started working in probability and statistics. It was purely by accident and certainly not preordained in any way. And it turned out to be fantastic!

What other books or people might have influenced you beyond Rosenblatt and beyond the books you've mentioned as well?

There was a very strong group of probabilists at UCSD including Ron Getoor, Michael Sharpe, Adriano Garsia, and Murray Rosenblatt, who attracted a significant number of PhD students interested in probability theory. I tended to hang out with this group of students, even though I wasn't a hardcore probabilist, at least relative to the others. I took measure-theoretic probability from Michael Sharpe, who in my book, was the best professor I ever had. To this day, I still think about his lecturing style and try to emulate him as much as possible. Sharpe would come into class with just a piece of chalk, no notes, and he would lay out the most beautiful and perfect lecture every time. He was a master of the subject.

The books that influenced me early on in graduate school were essentially texbooks: Rudin's *Real and Complex Analysis*; Breiman's *Probability*, and the two volumes of Feller's books *An Introduction to Probability Theory and its Applications*. I can't really recall any book in statistics, other than Cramér's, of course, that made a major impact on me at that time. Rather it was more reading papers and technical reports in extreme value theory especially those by Leadbetter and collaborators and a couple reports by Breiman and Stone.

What came after UCSD?

After UCSD, my first academic position was a two year appointment as an instructor in applied mathematics at MIT. In actuality, I was part of a statistics subgroup consisting of 5 to 6 people that was embedded in the mathematics department and headed up by Herman Chernoff.

The MIT opportunity was an amazing experience because I didn't really have much exposure to applied statistics. And Chernoff, who was one of the giants in statistics, had wonderful insight and could translate a real world question into a meaningful statistical one with real skill. Hanging around Chernoff was just an unbelievable experience. He would bring in experts from around MIT and elsewhere for seminars, whether it be leading economists or astronomers or climate scientists and he would brainstorm about their problems with our small group of statisticians.

Why Colorado?

After my two-year experience at MIT, it was time to leave and fortunately, I had many attractive options. I ended up choosing the Statistics Department at Colorado State University (CSU), mainly for two reasons. The first was that my brother was already living in Colorado and it would be a great opportunity to reconnect with him. Second, CSU was developing a very strong group in applied probability. Peter Brockwell was the senior guy of that group that included Sid Resnick and Simon Tavaré, who was hired the same year as me. So there was the four of us in this group for a number of years. It was a closely-knit group in the sense that we not only worked together on various research projects, but our families became life long friends as well. Sid and I embarked on a variety of problems in extreme value theory and Peter and I began writing our books on time series analysis. It was an exciting and productive time for me. Peter, Sid, and I had a number of National Science Foundation grants during this period. It was wonderful that Peter and Sid took me under their wings, and showed, by example, how to become a professional academic. They always treated me as an equal and never as a junior partner in our research endeavors. I don't think they ever viewed themselves as mentors to me, but rather kept their eye out on me so that I didn't screw up.

How did you get to lead the department in CSU?

We had a series of really great chairs in the statistics department, and after Sid left for Cornell in 1987, this was a blow to me both personally and professionally. We maintained our research program for a few more years, but it was difficult to sustain when we were not at the same institution. Sid and I had a great collaboration as we both brought different perspectives to the table and learned from each other. The papers we wrote in the 80's have aged well — they are still widely cited today. Around the same time that Sid left CSU, Peter returned to Australia. His future status about living in either Australia or Colorado was uncertain for the next 15 years or so. The bottom line is that there was a lot of uncertainty about the future of applied probability and time series at CSU during this period.

Fortunately, we were able to lure to CSU Richard Tweedie, an Australian, who was a powerhouse in Markov chains. Also, perhaps less known, was that Richard was also an amazing applied statistician who was a player in some of the major issues of the time in applied statistics. So Richard filled an important leadership gap in the department and became the chair of the department shortly after his arrival. After one term as chair, Richard stepped down due primarily to health issues. The Dean of Natural Sciences talked to me about succeeding Richard, and although my research career was going great, I agreed to take a stab at being chair in 1997.

So Columbia was the next step?

Yes, Columbia was the next stop, but this came 10 years later! In early 2006, some faculty from the Statistics Department at Columbia approached me about the possibility of joining the department. Now I am mostly a country boy and never envisioned living in a moderately sized city let alone one as big as NYC. They suggested I come out to give a talk in the department and see what it was like. I had become somewhat disillusioned with the lackluster support I was receiving from CSU's administration, so was more open to the idea of a possible move. Of course, my wife had to be comfortable moving to New York, which would represent a major change in our life style. After visiting NY and seeing what Columbia had to offer, my wife said she was willing to make the move. After giving my go ahead to Columbia, they made me an offer a few months later and I accepted — end of story.

How did you manage to balance research with chairing the Departments at CSU and Columbia?

This may seem strange, but I think some of my most productive research years occurred while I was chair of the departments at CSU and Columbia! I am not sure that I have a good reason for this. There is a saying attributed to Benjamin Franklin that, "if you want something done, ask a busy person," and I think there is a lot of truth in this. I reduced my teaching load substantially while being chair which allowed me to visit various collaborators for 2-week periods. I could get a lot done during these getaways, while being away from everyday hassles in running a department, and then I would continue to finish up these projects upon returning home. This model worked well for me. Now when I have less responsibilities, I feel less productive — it's strange!

One final curiosity. Was your first encounter with the regular variation related or unrelated with extreme value theory?

No, it's definitely related to extreme value theory. And of course, I knew about this in graduate school, but I didn't have such a great appreciation for it until I interacted with Resnick at CSU.

Thank you so much, Richard. It has been a great pleasure and honor.

Acknowledgements

An OpenAI speech recognition model (*whisper*) was used to automatically convert the raw recording into text. The source code is available from:

https://github.com/openai/whisper

REPORT

V Non-Associative day in Barcelos

Instituto Politécnico do Cávado de Ave (Barcelos, Portugal)

by **Natália Maria Rego**[*,**]

The *V Non-Associative Day in Barcelos* belongs to a series of activities aiming to bring together researchers interested in non-associative algebras and related topics, with a central goal of increasing the quality of research, promoting interaction among researchers, and discussing new directions for the future.

The Organizing Committee of this event was constituted by members of CMUP in collaboration with the following members of the research centers CMA-UBI and CMUC. Moreover, these members represented Portuguese higher education institutions.

Natália Rego, **Teresa Abreu**
CMUP – Instituto Politécnico do Cávado de Ave

Ivan Kaygorodov
University of Beira Interior; CMA-UBI

Amir Fernandez Ouaridi
CMUC – Universidade de Coimbra

The workshop was held at the *Polytechnic Institute of Cávado and Ave*, on January 20 and had the participation of seven invited speakers,

This event was supported by CIM – Centro Internacional de Matemática.

* In behalf of the Organizing Committee

** Instituto Politécnico do Cávado e do Ave. Polytechnic Institute of Cavado and Ave. , 4750-810 Vila Frescainha, Portugal
Email: nrego@ipca.pt

Program

10.00
Saïd Benayadi
Université de Lorraine-Metz, France
On pseudo-Euclidean Lie algebras whose Levi-Civita product is left Leibniz

11.00
Ignacio Bajo
Universidade de Vigo, Spain
Nilpotent pseudo-Hermitian quadratic Lie algebras

12.00
Lunch

15.00
Jorge Garces
Universidad Politécnica de Madrid
On the strict topology of the multipliers of a JB-algebra*

16.00
Daniel Fox
Universidad Politécnica de Madrid, Spain
Partial and quantitative associativity conditions and trace-forms

17.00
Coffee Break

17.30
Esther García González
Universidad Rey Juan Carlos, Spain
The filtration associated to an abelian inner ideal

18.00
Pedro Lopes
CAMGSD—Universidade de Lisboa, Portugal
Quandles and their profiles

19.00
Mykola Khrypchenko
CMUP—Universidade do Porto, Portugal
Transposed Poisson structures on Block Lie algebras and superalgebras

21.00
Dinner

Geometry on surfaces and Higgs bundles

by **Peter B. Gothen**[*,**]

Abstract.—There are three complete plane geometries of constant curvature: spherical, Euclidean and hyperbolic geometry. We explain how a closed oriented surface can carry a geometry which locally looks like one of these. Focussing on the hyperbolic case we describe how to obtain all hyperbolic structures on a given topological surface, and how to parametrise them. Finally we introduce Higgs bundles and explain how they relate to hyperbolic surfaces.

1 Introduction

The idea of considering geometry on a surface is well known to inhabitants of Planet Earth. Indeed, as any explorer knows, spherical geometry is appropriate. In this geometry distance is measured along arcs of great circles. These are the *geodesics* of spherical geometry, just like the geodesics of plane Euclidean geometry are segments of straight lines.

The spherical surface and the Euclidean plane are both *complete*, meaning that any geodesic can be extended indefinitely. Moreover they both have constant curvature, positive in the case of the sphere, and zero in the case of the plane. There is also a complete 2-dimensional geometry of constant negative curvature, namely the hyperbolic plane (which we shall introduce below).

The sphere is an example of a *closed surface*, i.e., a surface which is compact and has no boundary (as opposed to a closed disk, for example). Topologically, closed orientable surfaces are classified by the genus g, a non-negative integer: a surface of genus g can be realised inside 3-space as a g-*holed torus* as illustrated in Figure 1. We have seen that the genus zero surface supports spherical geometry but what about the other surfaces? Our first main goal in this article is to explain how the torus (genus one) supports a complete geometry which locally looks like the Euclidean plane, while a surface of genus $g \geq 2$ can be given a complete locally hyperbolic geometry. This involves considering certain special subgroups of the matrix group $\mathrm{SL}(2,\mathbb{R})$. We shall then see how the algebra and geometry of the matrix group $\mathrm{SL}(2,\mathbb{R})$ interact in interesting ways, and how this sheds light on the question of which subgroups give rise to hyperbolic surfaces.

Our second main goal is to explain how considering the set of all possible hyperbolic structures on a fixed topological surface of genus $g \geq 2$ leads to interesting and beautiful mathematics. Thus we introduce moduli spaces and explore some of their properties.

Finally, we shall give an introduction to Higgs bundles. We shall show how they can be used to shed new light on the theme of hyperbolic structures on surfaces and indicate their role in recent generalisations of some of the results explained earlier in the article.

The paper is mostly expository, only the final Section 8.4 includes some results in which the author has been involved.

For reasons of space the references are by no means complete, but we hope the interested reader will be able to use them as a starting point for further exploration.

* CMUP e Departamento de Matemática, Faculdade de Ciências da Universidade do Porto.
Email: pbgothen@fc.up.pt
** Partially supported by CMUP under the projects UIDB/00144/2020, UIDP/00144/2020, and the project EXPL/MAT-PUR/1162/2021 funded by FCT (Portugal) with national funds.

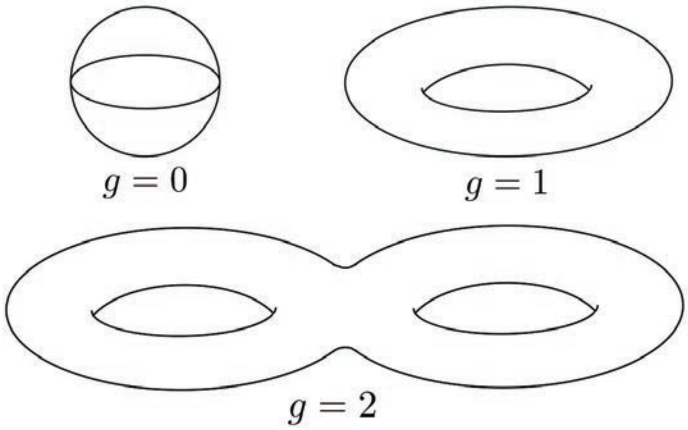


Figure 1.— Genus of a surface

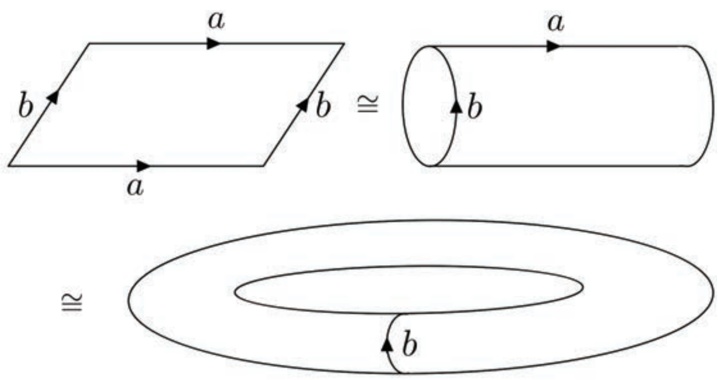


Figure 2.—A Euclidean surface

2 Euclidean surfaces

We want to explain how to do Euclidean geometry on a closed surface, in a way which makes the generalisation to the hyperbolic case natural.

2.1 The Euclidean plane

Using Cartesian coordinates we identify the Euclidean plane $\mathbb{E}^2$ with the coordinate plane $\mathbb{R}^2$. Distance is determined by calculating the length of a parametrised curve $\alpha : [a,b] \to \mathbb{R}^2$ in the usual way: $l(\alpha) = \int_a^b |\alpha'(t)|dt$. This is usually expressed by saying that in Cartesian coordinates $(x,y) \in \mathbb{R}^2$ on $\mathbb{E}^2$ the Euclidean element of arc length ds is given by

$$ds^2 = dx^2 + dy^2. \qquad (1)$$

Moreover, the distance preserving transformations form a group, $\mathrm{Isom}(\mathbb{E}^2)$, which is called the isometry group of $\mathbb{E}^2$. An example of isometries are *translations*. Using coordinates $\mathbb{E}^2 \cong \mathbb{R}^2$, the translation $A : \mathbb{E}^2 \to \mathbb{E}^2$ by the vector $\mathbf{a} \in \mathbb{R}^2$ can be written

$$A(P) = P + \mathbf{a}.$$

2.2 Euclidean surfaces

The Euclidean plane $\mathbb{E}^2$ is obviously not a closed surface. However, we can build a closed surface by taking a parallellogram in the plane and gluing its opposite sides, as illustrated in Figure 2: sides labeled with the same letter are to be glued with the orientation indicated by the arrows. We can carry out this process in 3-space — in a non-distance preserving way! — to convince ourselves that the resulting surface is really a topological torus.

More formally, we take linearly independent vectors $\mathbf{a}$ and $\mathbf{b}$ generating the sides labeled a and b. Let A and B be the translations by the vectors $\mathbf{a}$ and $\mathbf{b}$, respectively, and consider the subgroup

$$\Gamma = \langle A, B \rangle \subseteq \mathrm{Isom}(\mathbb{E}^2)$$

generated by them inside the isometry group of $\mathbb{E}^2$. This is just the group of translations by vectors of the form $n\mathbf{a} + m\mathbf{b}$, where $m, n \in \mathbb{Z}$. Since translations commute, the generators of Γ satisfy the single relation

$$[A, B] = I,$$

where $[A,B] := ABA^{-1}B^{-1}$ is the *commutator* and I is the identity. The *orbit space*

$$\mathbb{E}^2/\Gamma$$

is obtained identifying points $P, Q \in \mathbb{E}^2$ if there is a $\gamma \in \Gamma$ such that $Q = \gamma(P)$. Its points correspond to *orbits* $\Gamma \cdot Q = \{\gamma(Q) \mid \gamma \in \Gamma\}$. Thus each point of the interior of the paralellogram generated by $\mathbf{a}$ and $\mathbf{b}$ corresponds to a unique point of $\mathbb{E}^2/\Gamma$, and pairs of points on opposite sides are identified via the corresponding translation, thus realising the desired gluing. Hence $\mathbb{E}^2/\Gamma$ is a locally Euclidean surface, which is topologically a torus.

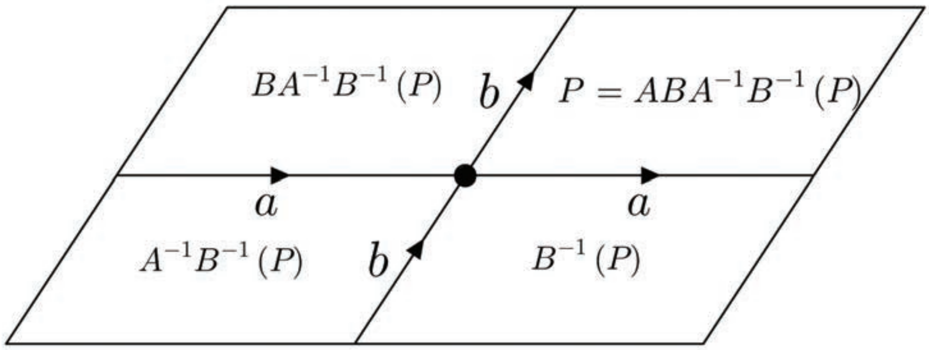


Figure 3.—Four copies of P

As illustrated in Figure 3 the four vertices of the parallelogram P get identified in $\mathbb{E}^2/\Gamma$. Moreover, there are four copies of the parallelogram meeting there, which fit together because of the relation $[A, B] = I$; the Euclidean metric is not distorted because the sum of the internal angles of the parallelogram is exactly 2π.

From a more abstract point of view, a key point is that the group Γ has the following property: for every point P in $\mathbb{E}^2$, there is an open neighbourhood U such that $\gamma(U) \cap U = \varnothing$ for all γ different from the identity. We say the action of Γ on $\mathbb{E}^2$ is *properly discontinuous*. This property ensures that for each $Q \in U$ its orbit $\Gamma \cdot Q = \{\gamma(Q) \mid \gamma \in \Gamma\}$ has a unique representative (namely Q itself) in U, so that U works as a *coordinate patch* for $\mathbb{E}/\Gamma$ around $\Gamma \cdot P$. This, together with the fact that the elements of Γ are isometries, means that $\mathbb{E}^2/\Gamma$ has a well defined distance function: indeed, the arc length of a parametrised curve in $\mathbb{E}^2/\Gamma$ can be calculated using the formula (1) which is invariant under isometries.

Note that we can also construct non-compact surfaces in this way. For example, if we take Γ to be the subgroup generated by a single translation, we obtain a cylinder. This is a locally Euclidean surface which, unlike the Euclidean torus, can be easily visualised in 3-space by rolling up a sheet of paper.

The so-called *Killing–Hopf Theorem* implies that any complete connected locally Euclidean surface can be represented as $\mathbb{E}^2/\Gamma$, where Γ acts freely and properly discontinuously on $\mathbb{E}^2$ (see, for example, Stillwell [16]).

3 Hyperbolic surfaces

3.1 The hyperbolic plane

We start by describing the *hyperbolic plane* $\mathbb{H}^2$. Hyperbolic geometry is different from spherical and Euclidean geometry in that it is not possible to embed (smoothly) $\mathbb{H}^2$ in Euclidean 3-space in a distance preserving way. Instead we consider the *upper half plane model*, defined by

$$\mathbb{H}^2 = \{z = x + iy \in \mathbb{C} \mid y > 0\}$$

with element of arc length

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

In the model $\mathbb{H}^2$ geodesics are open arcs of semicircles orthogonal to the real axis $\mathbb{R} = \{y = 0\} \subset \mathbb{C}$ together with open half-lines orthogonal to $\mathbb{R}$. Note that the hyperbolic plane is complete, so these curves do in fact have infinite hyperbolic length. Moreover, orientation preserving isometries can be represented by *Möbius transformations*

$$z \mapsto A \cdot z = \frac{az + b}{cz + d},$$

where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R})$$

is a real 2×2-matrix of determinant one. As examples we can take

$$A = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix}$$

which gives a hyperbolic translation whose axis is the imaginary axis in $\mathbb{H}^2$, and

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

which gives a hyperbolic rotation about $i \in \mathbb{H}^2$ by the angle 2θ.

We note that A and $-A$ define the same Möbius transformation, so the group of orientation preserving isometries is really the quotient group $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\{\pm I\}$. We shall mostly ignore this distinction in what follows but it will become relevant in Section 7 below.

We finish this section by commenting on the topology of $\mathrm{SL}(2,\mathbb{R})$. Identifying the set of all 2×2-matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $\mathbb{R}^4$, the group $\mathrm{SL}(2,\mathbb{R})$ is the subset cut out by the equation $ac - bd = 1$. Thus it has a topology inherited from $\mathbb{R}^4$. In fact the Implicit Function Theorem applied to this equation shows that $\mathrm{SL}(2,\mathbb{R})$ is a 3-dimensional *Lie group*, meaning that it can be covered by local coordinate systems in 3-space and that the group operations are differentiable in these coordinates.

3.2 Hyperbolic surfaces

As we shall see, a closed orientable topological surface of genus g can be given a hyperbolic structure for any $g \geq 2$. In the case of $g = 2$, take an octagon with gluing instructions to create a surface as illustrated in Figure 4. If we cut the octagon along the diameter indicated, we see that indeed the resulting surface has genus 2, as desired.

In order to get a nice *hyperbolic* surface, the octagon should be taken in the hyperbolic plane (it will look very different from that of 4). And, in a manner analogous to the Euclidean case, we require that pairs of sides which are to be glued have the same length. Moreover, the vertices of the octagon all get identified to one point in the surface, so the internal angles should add up to 2π. This condition sounds strange to our Euclidean wired brains but, it is a fact that such an octagon exists.[1]

In order to write the surface as $\mathbb{H}^2/\Gamma$ for a suitable subgroup $\Gamma \subset \mathrm{SL}(2,\mathbb{R})$ we take hyperbolic translations $\boldsymbol{A_i}$ and $\boldsymbol{B_i}$ giving the required identifications, and let Γ be the group generated by these translations. The octagon then becomes a fundamental domain for the action of Γ. The condition that the interior angles add up to 2π is equivalent to the identity

$$[\boldsymbol{A}_1, \boldsymbol{B}_1][\boldsymbol{A}_2, \boldsymbol{B}_2] = \boldsymbol{I}$$

in $\mathrm{SL}(2,\mathbb{R})$. In general, we let Γ_g be the abstract group

$$\Gamma_g = \langle a_1, b_1, \dots, a_g, b_g \mid \prod_{i=1}^{g} [a_i, b_i] = 1 \rangle.$$

This group is known as a *surface group*.[2] In view of the genus 2 example it is hopefully not a surprise that genus g surfaces can be obtained from subgroups of $\mathrm{SL}(2,\mathbb{R})$ isomorphic to Γ_g. In order to study all such subgroups we consider homomorphisms $\rho : \Gamma_g \to \mathrm{SL}(2,\mathbb{R})$ (often also called *representations*). We say that ρ is *Fuchsian* if it is injective and its image is discrete, i.e., consists of isolated points.[3] When ρ is Fuchsian it can be proved that the action of Γ_g on $\mathbb{H}^2$ is properly discontinuous. Hence the orbit space

$$S_\rho := \mathbb{H}^2/\rho(\Gamma_g),$$

is a nice hyperbolic surface of genus g with charts coming from $\mathbb{H}^2$. Conversely, the Killing–Hopf Theorem again tells us that any closed orientable hyperbolic surface is of this form.[4]

However, it is certainly not true that any homomorphism $\rho : \Gamma_g \to \mathrm{SL}(2,\mathbb{R})$ defines a closed hyperbolic surface: for example, the trivial homorphism clearly does not! This leaves us with the following

Question: Let $\rho : \Gamma_g \to \mathrm{SL}(2,\mathbb{R})$ be a representation. How can we tell if ρ defines a closed hyperbolic surface?

[1] In fact, in hyperbolic geometry the sum of the internal angles of a polygon depends on its area!

[2] The group Γ_g can be identified wth the fundamental group of a topological surface of genus g.

[3] Recall that topological notions make sense viewing $\mathrm{SL}(2,\mathbb{R}) \subseteq \mathbb{R}^4$.

[4] As already noted, we should really consider representations to $\mathrm{PSL}(2,\mathbb{R})$. However, it turns out that representations defining closed hyperbolic surfaces can always be lifted to $\mathrm{SL}(2,\mathbb{R})$.

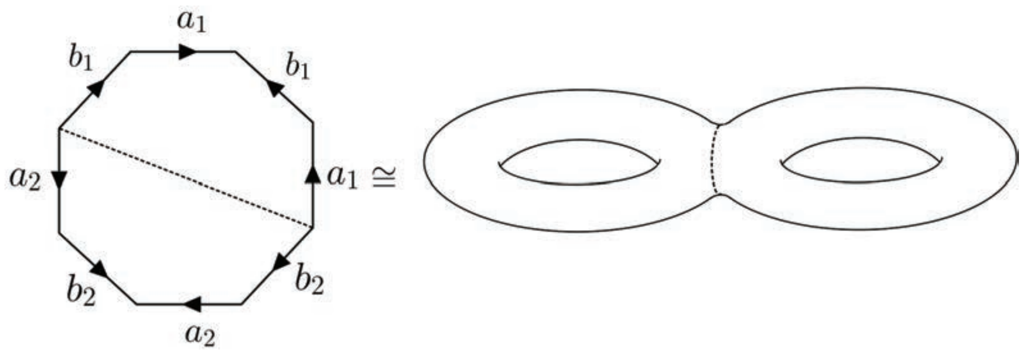


Figure 4.—Genus 2 surface from an octagon

4 Topology and algebra of $\mathrm{SL}(2,\mathbb{R})$

In order to answer the question at the end of the last section we shall define an invariant of representations $\rho\colon \Gamma \to \mathrm{SL}(2,\mathbb{R})$. For that we shall need to understand how the topology and algebra of $\mathrm{SL}(2,\mathbb{R})$ interact.

The subgroup $\mathrm{SO}(2) \subseteq \mathrm{SL}(2,\mathbb{R})$ of rotation matrices

$$E(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

can be identified with a circle.

The map $E\colon \mathbb{R} \to \mathrm{SO}(2)$, $\theta \mapsto E(\theta)$ wraps the real line around the circle, and it satisfies $E(0) = I$ and $E(\theta_1 + \theta_2) = E(\theta_1)E(\theta_2)$. In other words, E is a group homomorphism from the additive group $\mathbb{R}$ to $\mathrm{SO}(2)$.

Now, thinking of $\mathrm{SO}(2)$ inside $\mathrm{SL}(2,\mathbb{R})$, we want to extend this picture and find a group $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ containing $\mathbb{R}$, with a surjective group homomorphism $p\colon \widetilde{\mathrm{SL}}(2,\mathbb{R}) \to \mathrm{SL}(2,\mathbb{R})$ which restricts to $E\colon \mathbb{R} \to \mathrm{SO}(2)$, i.e., making the diagram

$$\begin{tikzcd} \mathbb{R} \arrow[r] \arrow[d, "E"] & \widetilde{\mathrm{SL}}(2,\mathbb{R}) \arrow[d, "p"] \\ \mathrm{SO}(2) \arrow[r] & \mathrm{SL}(2,\mathbb{R}) \end{tikzcd}$$

commutative (the horizontal maps are inclusions). In fact it follows from general theory that such a group exists and is essentially unique; it is known as the *universal covering group* of $\mathrm{SL}(2,\mathbb{R})$. We shall explain how it can be constructed explicitly, following [13, §1.8], using the action of $\mathrm{SL}(2,\mathbb{R})$ on the hyperbolic plane $\mathbb{H}^2$.

So let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{R}).$$

Write

$$j(A,z) = cz + d$$

for the denominator of $A \cdot z$. Note that

$$j(E(\theta), i) = i\sin\theta + \cos\theta = e^{i\theta},$$

which indicates that this function can used to keep track of the phase θ. For each fixed A, we can consider the holomorphic function

$$\begin{aligned} \mathbb{H}^2 &\to \mathbb{C} \setminus \{0\} \\ z &\mapsto j(A,z) = cz + d. \end{aligned}$$

Observe that $cz + d \neq 0$ for $z \in \mathbb{H}$. Therefore, since $\mathbb{H}^2$ is simply connected, there is a continuous determination of the logarithm of $j(A,z) = cz + d$, i.e., a continuous map $\phi\colon \mathbb{H}^2 \to \mathbb{C}$ such that

$$e^{\phi(z)} = cz + d.$$

We want to make the point that such a ϕ can be explicitly calculated: simply choose a value θ for the argument $\arg(ci + d)$, write $ci + d = re^{i\theta}$ and let $\phi(i) = \log(r) + i\theta$. Then

$$\phi(z) - \phi(i) = \int_\gamma \frac{dz}{z} = \int_0^1 \frac{c(z-i)dt}{c(i + t(z-i)) + d}$$

(here γ parametrises the segment joining $ci + d$ to $cz + d$). Note that ϕ is not unique, but it is uniquely determined by the choice of $\phi(i)$. Thus any two determinations ϕ differ by an integer multiple of $2\pi i$.

Now define $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ as the set of pairs (A,ϕ), where $A \in \mathrm{SL}(2,\mathbb{R})$ and $\phi\colon \mathbb{H}^2 \to \mathbb{C}$ is any continuous determination of the logarithm of $j(A, z)$. The product on $\widetilde{\mathrm{SL}}(2, \mathbb{R})$ is defined by

$$(A_1, \phi_1) \cdot (A_2, \phi_2) = (A_1 A_2, \tilde{\phi}),$$

where

$$\tilde{\phi}(z) := \phi_1(A_2 \cdot z) + \phi_2(z).$$

It is an easy calculation to check that

$$j(A_1 A_2, z) = j(A_1, A_2 \cdot z) j(A_2, z)$$

which implies that indeed

$$e^{\tilde{\phi}(z)} = j(A_1 A_2, z)$$

as required. It is not hard to check that this defines a group structure on $\widetilde{\mathrm{SL}}(2, \mathbb{R})$. For example, for $A = I$, the identity matrix, we can take $\phi(z) = 0$ and $(A, 0)$ is the neutral element. Moreover, $(A, \phi)^{-1} = (A^{-1}, \tilde{\phi})$, where

$$\tilde{\phi}(z) = -\phi(A^{-1} \cdot z). \tag{2}$$

The projection $p : \widetilde{\mathrm{SL}}(2, \mathbb{R}) \to \mathrm{SL}(2, \mathbb{R})$ is of course just $(A, \phi) \mapsto A$. The inclusion $\mathbb{R} \hookrightarrow \widetilde{\mathrm{SL}}(2, \mathbb{R})$ is given by $\theta \mapsto (E(\theta), \phi_\theta)$, where ϕ_θ is the determination of $\log(j(E(\theta), z))$ which satisfies $\phi_\theta(i) = i\theta$ (recall that $j(E(\theta), i) = e^{i\theta}$).

Proposition 1.— The kernel of $p : \widetilde{\mathrm{SL}}(2, \mathbb{R}) \to \mathrm{SL}(2, \mathbb{R})$ consists of pairs (I, ϕ), where I is the identity matrix and ϕ is a constant function taking values in $2\pi\mathbb{Z} \subset \mathbb{R}$.

Proof.— Clearly $p(A, \phi) = I$ if and only if $A = I$. Moreover, $j(I, z) = 1$, so ϕ is a determination of the logarithm of the constant function $z \mapsto 1 \in C$, i.e., it is a constant $\phi \in 2\pi\mathbb{Z} \subset \mathbb{R}$. ∎

5 The Toledo Invariant

Let $\rho : \Gamma \to \mathrm{SL}(2, \mathbb{R})$ be a representation. We shall associate an integer invariant to ρ. This invariant is known as the Toledo invariant, even though it was actually introduced by Milnor [14], and sometimes is referred to as the Euler number. Write

$$A_i = \rho(a_i), \quad B_i = \rho(b_i)$$

for $i = 1, \ldots, g$. Choose *lifts* $\tilde{A}_i$ and $\tilde{B}_i$ in $\widetilde{\mathrm{SL}}(2, \mathbb{R})$ such that $p(\tilde{A}_i) = A_i$ and $p(\tilde{B}_i) = B_i$, and define the *Toledo invariant* of ρ to be

$$\tau(\rho) = \frac{1}{\pi} \prod_{i=1}^{g} [\tilde{A}_i, \tilde{B}_i].$$

In view of the relation defining Γ_g, the product $\prod_{i=1}^{g} [\tilde{A}_i, \tilde{B}_i]$ is in the kernel of p. Hence Proposition 1 shows that the Toledo invariant is an even integer.[5]

From the description of $\widetilde{\mathrm{SL}}(2, \mathbb{R})$ of the preceding section, it is easy to check that the Toledo invariant is well defined, i.e., that it does not depend on the choice of lifts: the main point is that the ambiguity in the choice of ϕ is canceled by (2), because each lift occurs together with its inverse in the commutator. Moreover, the Toledo invariant of a representation defined by matrices A_i and B_i *can be explicitly calculated.*

6 Goldman's theorem

A celebrated inequality due to Milnor [14] states that

$$|\tau(\rho)| \leq 2g - 2$$

for every representation $\rho : \Gamma_g \to \mathrm{SL}(2, \mathbb{R})$. The following beautiful result shows that representations with maximal Toledo invariant (known as *maximal representations*) have a special geometric significance.

Theorem 2 (Goldman [7]).— A representation $\rho : \Gamma_g \to \mathrm{SL}(2, \mathbb{R})$ is Fuchsian if and only if $|\tau(\rho)| = 2g - 2$.

Remark 1.— One might wonder about the significance of the sign of the Toledo invariant. If we conjugate a representation ρ by the outer automorphism of $\mathrm{SL}(2, \mathbb{R})$ given by conjugation by a reflection we obtain a representation $\bar{\rho}$ with $\tau(\bar{\rho}) = -\tau(\rho)$. In fact, the hyperbolic surface $S_{\bar{\rho}}$ is obtained from S_ρ by a change of orientation, i.e., by composing all charts with a reflection in $\mathbb{H}^2$.

7 The moduli space of representations

Let us now take a global view and consider all representations of Γ_g in $\mathrm{SL}(2, \mathbb{R})$ simultaneously. The *representation space* for representations of Γ_g in $\mathrm{SL}(2, \mathbb{R})$ is the set of homomorphisms $\mathrm{Hom}(\Gamma_g, \mathrm{SL}(2, \mathbb{R}))$. It is natural to consider ρ_1 and ρ_2 equivalent if they differ by overall conjugation by an element of $\mathrm{SL}(2, \mathbb{R})$, corresponding to a change of basis in $\mathbb{R}^2$. It also turns

[5] Odd Toledo invariants arise from representations $\rho : \Gamma_g \to \mathrm{PSL}(2, \mathbb{R})$ which do not lift to $\mathrm{SL}(2, \mathbb{R})$.

out that two hyperbolic structures on the same topological surface are isometric by an isometry which can be continuously deformed to the identity if and only if the corresponding Fuchsian representations are equivalent in this sense. Thus the *moduli space* of representations is defined to be the orbit space

$$\mathcal{R}(\Gamma_g, \mathrm{SL}(2,\mathbb{R})) = \mathrm{Hom}(\Gamma_g, \mathrm{SL}(2,\mathbb{R}))/\mathrm{SL}(2,\mathbb{R})$$

under the conjugation action.[6]

A homomorphism $\rho : \Gamma_g \to \mathrm{SL}(2,\mathbb{R})$ is determined by $2g$ matrices

$$A_i = \rho(a_i),\ \ B_i = \rho(b_i), \quad i = 1, \dots, g$$

satisfying the single relation $\prod[A_i, B_i] = I$. Hence $\mathrm{Hom}(\Gamma_g, \mathrm{SL}(2,\mathbb{R}))$ can be identified with the subspace of $\mathbb{R}^{6g}$ cut out by the 3 scalar equations given by $\prod[A_i, B_i] = I$ (the equation takes values in the 3-dimensional group $\mathrm{SL}(2,\mathbb{R})$). It follows that it is a variety of dimension $6g-3$. The conjugation action by $\mathrm{SL}(2,\mathbb{R})$ reduces the dimension by 3, and so the moduli space has dimension

$$\dim \mathcal{R}(\Gamma_g, \mathrm{SL}(2,\mathbb{R})) = 6g - 6.$$

The Toledo invariant separates the moduli space into subspaces

$$\mathcal{R}_d \subseteq \mathcal{R}(\Gamma_g, \mathrm{SL}(2,\mathbb{R}))$$

corresponding to representations with invariant d. Goldman [8] showed that the $\mathcal{R}_d$ are in fact connected components of the moduli space, except in the maximal case $|d| = 2g-2$. It turns out that $\mathcal{R}_{2g-2}$ has 2^{2g} connected components. However, these components get identified after projecting onto

$$\mathcal{R}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R})),$$

which thus has just one connected component with Toledo invariant $2g-2$. This is not surprising because, by Goldman's Theorem 2, the subspace $\mathcal{R}_{2g-2}$ is exactly the locus of Fuchsian representations and, moreover, any two Fuchsian representations into $\mathrm{SL}(2,\mathbb{R})$ define the same hyperbolic surface if and only if they coincide after projecting to $\mathrm{PSL}(2,\mathbb{R})$. Accordingly, the corresponding connected component $\mathcal{T} \subseteq \mathcal{R}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$ is known as the *Fuchsian locus*. As we have seen, it parametrises all hyperbolic structures on the topological surface S_g up to a natural equivalence. It is a classical result that the space of such hyperbolic structures can be identified with $\mathbb{R}^{6g-6}$. In the next section we shall explain how a parametrisation of this space can be obtained using Higgs bundles.

8 Higgs bundles

We now describe how the results of the previous section can be understood using non-abelian Hodge theory, a subject founded by Hitchin [11] and Simpson [15].

8.1 Riemann surfaces and holomorphic bundles

A *Riemann surface* X is a topological surface together with a family of local charts which together cover the surface, and are such that changes of coordinates are holomorphic functions between open sets in $\mathbb{C}$. An example of this is the *Riemann sphere* $\hat{\mathbb{C}} = \mathbb{C} \cup \infty$: we use the standard coordinate z in $\mathbb{C}$ and around $\infty \in \hat{\mathbb{C}}$ we use the coordinate $w = 1/z$. Thus the change of coordinates $T : \mathbb{C} \setminus \{0\} \to \mathbb{C} \setminus \{0\}$ given by $w = T(z) = 1/z$ is holomorphic in the domain where both z and w are defined.

In particular, if we have a hyperbolic surface $S_g \cong \mathbb{H}^2/\Gamma$ for a Fuchsian representation of Γ, then the local coordinates in $\mathbb{H}^2$ give S_g the structure of a Riemann surface: indeed the changes of coordinates are Möbius transformations of $\mathbb{H}^2$, which are certainly holomorphic. We write X_ρ for the Riemann surface constructed from a Fuchsian representation ρ in this way.

Note that not all holomorphic maps define isometries of $\mathbb{H}^2$, so the concept of Riemann surface is less rigid than that of hyperbolic surface. However, the famous Uniformisation Theorem, due to Köbe and Poincaré, asserts that any Riemann surface can be represented as a hyperbolic surface. This means, in particular, that the space of all Riemann surfaces with the same underlying topological surface of genus g (up to a suitable equivalence) can be identified with the Fuchsian locus $\mathcal{T}$. When thought of in this way, it is known as *Teichmüller space*.

A *holomorphic line bundle* $L \to X$ on a Riemann surface X is a holomorphic family of 1-dimensional complex vector spaces parametrised by X. Thus, for each $p \in X$ we have a 1-dimensional complex vector space L_p, which varies holomorphically with p. The simplest example is the *trivial bundle* $L = X \times \mathbb{C} \to X$; here the map is projection onto X and the fibre

[6] In order to get a Hausdorff quotient, one should in fact exclude representations whose action on $\mathbb{R}^2$ is not semisimple.

$L_p = \{p\} \times \mathbb{C}$ for $p \in X$ with its vector space structure coming from $\mathbb{C}$. Locally on X, a holomorphic line bundle is required to look like the product $U \times \mathbb{C}$, where $U \subseteq X$ is open. We say that L is trivialised over U. This means that a holomorphic line bundle can be given by an open covering $\{U_\alpha\}$ of X and *trivialisations*

$$L_{|U_\alpha} \cong U_\alpha \times \mathbb{C}$$

for each α. This gives rise to *transition functions*

$$g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathbb{C} \setminus \{0\}$$

which compare the isomorphisms $L_p \cong \mathbb{C}$ given by the trivialisations over U_α and U_β, respectively.

More important than the line bundles themselves are their *sections*. These are holomorphic maps $s : X \to L$ such that $s(p) \in L_p$ for all $p \in X$. A section of the trivial bundle $U \times \mathbb{C}$ over U is of course nothing but a map $s : U \to \mathbb{C}$, and if we have local trivialisations of a line bundle L as above, a holomorphic section s corresponds to a collection of holomorphic maps $s_\alpha : U_\alpha \to \mathbb{C}$ which glue correctly, i.e., satisfy the condition

$$s_\alpha(p) = g_{\alpha\beta}(p) s_\beta(p)$$

for $p \in U_\alpha \cap U_\beta$. As an illustrative example, we take the *canonical bundle* $K \to X$. Its sections are *holomorphic differentials*. In a local coordinate z on X a holomorphic differential, say α, can be written

$$g(z)dz$$

for a holomorphic function $g(z)$ and if $h(w)dw$ is the representation of α in another holomorphic coordinate $w = T(z)$, then

$$g(z)dz = h(T(z))d(T(z)) = h(T(z))T'(z)dz.$$

Thus a holomorphic differential can be represented by a collection of holomorphic functions locally defined on coordinate charts which transform according to the preceding rule. It turns out that the vector space of holomorphic differentials on a closed Riemann surface X of genus g, usually denoted by $H^0(X, K)$, is finite dimensional, of dimension $2g-2$. More generally, the vector space $H^0(X, L)$ of holomorphic sections of any holomorphic line bundle $L \to X$ is finite dimensional. Any holomorphic line bundle has a topological invariant called its *degree*; in case L has a non-zero holomorphic section, this is the number of zeroes of such a section, counted with multiplicity. For example, the degree of the canonical bundle is $2g-2$. The fact that this is the same as the dimension of $H^0(X, K)$ is a consequence of a fundamental result known as the Riemann–Roch formula.

We can perform the usual operations of linear algebra, like taking duals and tensor products, fibrewise on line bundles. Thus, if L and M are line bundles with transition functions $g_{\alpha\beta}$ and $h_{\alpha\beta}$, respectively, the tensor product $L \otimes M$ has transition functions $g_{\alpha\beta}h_{\alpha\beta}$ (pointwise multiplication in $\mathbb{C} \setminus \{0\}$) and the dual bundle L^* has transition functions $g_{\alpha\beta}^{-1}$.

8.2 Higgs bundles

A $\mathrm{PSL}(2, \mathbb{R})$-*Higgs bundle* on X consists of three pieces of data

$$(L, \beta, \gamma)$$

where, $L \to X$ is a holomorphic line bundle, and $\beta \in H^0(X; K \otimes L)$ and $\gamma \in H^0(X; K \otimes L^*)$ can be seen as holomorphic differentials which take values in the line bundles L and L^*, respectively.

In a manner analogous to the conjugation action on representations, there is a natural notion of isomorphism of Higgs bundles, and the set of isomorphism classes of $\mathrm{PSL}(2, \mathbb{R})$-Higgs bundles forms the *moduli space* $\mathscr{M}(X, \mathrm{PSL}(2, \mathbb{R}))$. It is a complex algebraic variety of complex dimension $3g-3$. We note that in order to get a reasonable moduli space it is necessary to restrict to so-called semistable Higgs bundles. This is analogous to the way in which one restricts to semisimple representations in the moduli space of representations.

The *Non-abelian Hodge Theorem* (due to Corlette, Donaldson, Hitchin and Simpson) for this situation states the following.

Theorem 3.— There is a real analytic isomorphism

$$\mathscr{R}(\Gamma_g, \mathrm{PSL}(2, \mathbb{R})) \cong \mathscr{M}(X, \mathrm{PSL}(2, \mathbb{R})).$$

This is a remarkable theorem for many reasons. Here we just point out that while the character variety $\mathscr{R}$ is **real** and depends only on the topological surface of genus g (through its fundamental group), the moduli space $\mathscr{M}$ depends on the Riemann surface structure X given to the topological surface and has a **complex** structure.

For fixed d we denote by $\mathscr{M}_d$ the subspace of $\mathrm{PSL}(2, \mathbb{R})$-Higgs bundles (L, β, γ) with $\deg(L) = d$. Then we have $\mathscr{R}_d \cong \mathscr{M}_d$ under the non-abelian Hodge Theorem. In particular, the Fuchsian locus $\mathscr{T}$ corresponds to $\mathscr{M}_{2g-2}$.

8.3 Hitchin's parametrisation of $\mathcal{T}$

A particular class of $\mathrm{PSL}(2,\mathbb{R})$-Higgs bundles can be obtained by taking $L = K$. Then γ is a section of the line bundle $K \otimes K^*$ which is naturally isomorphic to the trivial line bundle on X. In other words, γ is simply a holomorphic function on X, so we can set $\gamma = 1$ (the constant function). Moreover, β is a section of $K^2 = K \otimes K$. In other words it is a *quadratic differential*, so it can locally be written as $\beta(z) = b(z)(dz)^2$, where $b(z)$ satisfies an appropriate transformation rule under changes of coordinates. The vector space $H^0(X, K^2)$ of quadratic differentials on X has complex dimension $3g - 3$ which equals the dimension of the moduli space $\mathcal{M}(X, \mathrm{PSL}(2,\mathbb{R}))$. This construction defines a map

$$\begin{aligned} \Psi\colon\ H^0(X,K^2) &\to \mathcal{M}(X,\mathrm{PSL}(2,\mathbb{R})),\\ \beta &\mapsto (K,\beta,1). \end{aligned}$$

The semistability condition alluded to earlier implies that all Higgs bundles in $\mathcal{M}_{2g-2}$ arise in this way. Hence Ψ is an isomorphism onto its image $\mathcal{M}_{2g-2}$.

From the non-abelian Hodge Theorem we already knew that $\mathcal{M}_{2g-2} \cong \mathcal{T}$ is a connected component. But the Higgs bundle construction gives an alternative proof. Using gauge theoretic methods Hitchin also shows that $\mathcal{M}_{2g-2}$ parametrises all hyperbolic metrics on the topological surface underlying X. Moreover, under this parametrisation the Higgs bundle $(K, 1, 0)$ corresponds to the hyperbolic metric which uniformises X. Thus Hitchin's approach gives alternative proofs of Goldman's theorems and the Uniformisation Theorem.

8.4 The general Cayley correspondence

Hitchin [12] generalised the construction of the map Ψ to a map

$$\Psi\colon\ \bigoplus_i H^0(X,K^{d_i}) \to \mathcal{M}(X,G)$$

whose image is again a connected component of the *moduli space $\mathcal{M}(X,G)$ of G-Higgs bundles* for any simple split real Lie group G, nowadays known as a *Hitchin component*.[7] The domain of Ψ is a direct sum of spaces of higher holomorphic differentials on X; the integers d_i are determined by the Lie group G (in fact they are the exponents of its Lie algebra).

Similar constructions of special connected components have later been given for Hermitian groups G of non-compact tube type, such as $\mathrm{SU}(p,p)$ (see, for example, [5, 6, 2]). In this case the domain of the map Ψ turns out to be a moduli space $\mathcal{M}_{K^2}(X,G')$ of so-called K^2-twisted G'-Higgs bundles, for a certain real Lie group G' associated to G (known as its *Cayley partner*).

Both Hitchin components and Cayley components are special because they are not (as all other known components of the moduli space) detected by standard topological invariants of the underlying bundles and the Higgs fields satisfy a certain non-degeneracy condition.

Recently (see [1, 4] and the recent survey [3]) both of these constructions have been unified and generalised. The class of Lie groups G covered are characterised by the fact that their Lie algebras admit a *magical $\mathfrak{sl}_2$-triple*. This new Lie theoretic notion builds on ideas of Hitchin [12] and generalises that of a principal $\mathfrak{sl}_2$-triple introduced by Kostant. Conjecturally the generalised Cayley components obtained by this construction account for all "special" (in the sense of the previous paragraph) connected components of the moduli space and thus opens the door to a complete determination of this important topological invariant.

One important piece of supporting evidence for this conjecture comes from the area of *Higher Teichmüller Theory*. Higher Teichmüller theory has developed in parallel with the Higgs bundle story just described, and there has been a rich cross-fertilisation of ideas between the two areas. We cannot do justice to this fast-growing, rich and important area of mathematics here but refer the interested reader to [10, 17] and references therein. Very briefly, a higher Teichmüller space is a connected component of the moduli space of representations, which consists exclusively of discrete and injective representations, like the Fuchsian locus in the $\mathrm{PSL}(2,\mathbb{R})$-case. It turns out that the generalised Cayley components are indeed higher Teichmüller spaces [4, 9], and it is expected that all higher Teichmüller spaces are thus obtained.

[7] In the case of classical matrix groups this means that G is one of the groups $\mathrm{SL}(n,\mathbb{R})$, $\mathrm{Sp}(2n,\mathbb{R})$, $\mathrm{SO}(p,p)$ and $\mathrm{SO}(p,p+1)$.

References

[1] M. Aparicio-Arroyo, S. Bradlow, B. Collier, O. García-Prada, P. B. Gothen, and A. Oliveira, *$SO(p, q)$-Higgs bundles and higher Teichmüller components*, Invent. Math. **218** (2019), no. 1, 197–299.

[2] O. Biquard, O. García-Prada, and R. Rubio, *Higgs bundles, the Toledo invariant and the Cayley correspondence*, J. Topol. **10** (2017), no. 3, 795–826.

[3] S. Bradlow, *Global properties of Higgs bundle moduli spaces*, 2023, to appear.

[4] S. Bradlow, B. Collier, O. García-Prada, P. B. Gothen, and A. Oliveira, *A general Cayley correspondence and higher Teichmüller spaces*, `arXiv:2101.09377 [math.AG]`, 2021.

[5] S. B. Bradlow, O. García-Prada, and P. B. Gothen, *Surface group representations and* $\mathrm{U}(p, q)$*-Higgs bundles*, J. Differential Geom. **64** (2003), 111–170.

[6] ———, *Maximal surface group representations in isometry groups of classical hermitian symmetric spaces*, Geometriae Dedicata **122** (2006), 185–213.

[7] W. M. Goldman, *Representations of fundamental groups of surfaces*, Springer LNM 1167, 1985, pp. 95–117.

[8] ———, *Topological components of spaces of representations*, Invent. Math. **93** (1988), 557–607.

[9] O. Guichard, F. Labourie, and A. Wienhard, *Positivity and representations of surface groups*, `arXiv:2106.14584 [math.DG]`, 2021.

[10] O. Guichard and A. Wienhard, *Positivity and higher Teichmüller theory*, European Congress of Mathematics, Eur. Math. Soc., Zürich, 2018, pp. 289–310.

[11] N. J. Hitchin, *The self-duality equations on a Riemann surface*, Proc. London Math. Soc. (3) **55** (1987), 59–126.

[12] ———, *Lie groups and Teichmüller space*, Topology **31** (1992), 449–473.

[13] G. Lion and M. Vergne, *The Weil representation, Maslov index and theta series*, Progress in Mathematics, vol. 6, Birkhäuser, Boston, Mass., 1980.

[14] J. W. Milnor, *On the existence of a connection with curvature zero*, Comment. Math. Helv. **32** (1958), 216–223.

[15] C. T. Simpson, *Higgs bundles and local systems*, Inst. Hautes Études Sci. Publ. Math. **75** (1992), 5–95.

[16] J. Stillwell, *Geometry of surfaces*, Universitext, Springer-Verlag, New York, 1992. MR 1171453 (94b:53001)

[17] A. Wienhard, *An invitation to higher Teichmüller theory*, Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. II. Invited lectures, World Sci. Publ., Hackensack, NJ, 2018, pp. 1013–1039.

REPORT

Global Portuguese Mathematicians 2023

by João Gouveia*

The third edition of the *Global Portuguese Mathematicians Conference* took place this year from 26 to 28 of July at the Department of Mathematics of the University of Coimbra. Previous editions have taken place in Lisbon (2017) and Porto (2019). This Conference brings together Portuguese mathematicians, in a very broad sense, from around the world, and provides them a forum to share their research, exchange ideas, and make new connections in a friendly and supportive environment.

This edition had over 60 enrolled participants and 12 invited speakers.

The presentations covered many different areas of mathematics and provided a glimpse on the current work of the Portuguese mathematical diaspora, as can be seen by the list of invited speakers.

Besides the talks, there were plenty of opportunities for interaction, and the conference successfully provided a lively networking forum for all attendants. All speakers were invited to participate in a special number of the *Boletim da Sociedade Portuguesa de Matemática* that will be published during 2024. The next edition of the conference will take place in 2025 at the University of Lisbon. For more information please visit the conference website at www.mat.uc.pt/~gpm23/.

* Departamento de Matemática da Universidade de Coimbra. Email: jgouveia@mat.uc.pt

Invited speakers

Eloísa Grifo
University of Nebraska

António Girão
University of Oxford

Gonçalo dos Reis
University of Edinburgh

João Pereira
IMPA

Vanda Inácio de Carvalho
University of Edinburgh

André Guerra
ETH Zurich

Rita Costa
Princeton University

Nuno Romão
University of Augsburg/IHES

João Lourenço
University of Münster

Miguel Moreira
ETH Zurich

Teresa Conde
University of Stuttgart

Margarida Melo
Università Roma Tre

A tour through some Diophantine equations

by **Ariel Pacetti**[*,**]

Abstract.—The purpose of this short note is to present some results regarding the study of Diophantine equations, ranging from very old problems (well known results to most mathematicians) to some quite new results in the area. Most of the times, the techniques developed to solve particular problems are more interesting than the results themselves. The last section contains our humble contribution.

1 Introduction

The term *Diophantine equations* comes from the pioneer work of Diophantus of Alexandria, a Greek mathematician that lived sometime around 200 AD. In a series of books called *Arithmetica*, Diophantus studied solutions (over the positive rational numbers) to different systems of equations. In this short article we will mostly focuses on studying (integral or rational) solutions of a single equation of the form

$$F(x, y) = 0,$$

for some polynomial $F(x, y)$ with integral coefficients.

2 Linear Diophantine Equations

Let a, b, c be three rational integers, with the condition that the pair (a, b) is not $(0, 0)$. Consider the equation

$$aX + bY = c. \tag{1}$$

The set of solutions to (1) form a line, and it is well known that it has infinitely many rational points. However, it is not so clear what happens with its set of integral points. For example, the line

$$2X + 2Y = 1,$$

does not have any integral point (the reason being that the left hand side is always even, while 1 is odd). Similarly, for (1) to have an integral solution, it must be the case that any number dividing a and b must also divide c. It is not hard to prove that this condition is enough for the existence of solutions.

Theorem 1.— The equation (1) has an integral solution if and only if $\gcd(a, b) \mid c$. Furthermore, if it has one, it has infinitely many.

Proof.— See §5 of the very nice book [11]. ∎

Actually the proof is constructive: suppose that $\gcd(a, b) \mid c$. Using the Euclidean algorithm, one can construct integers r, s such that

$$\gcd(a, b) = a \cdot r + b \cdot s.$$

Multiplying both sides by $\frac{c}{\gcd(a,b)}$ gives a non-trivial solution $x_0 = \frac{rc}{\gcd(a,b)}$, $y_0 = \frac{sc}{\gcd(a,b)}$. Then all solutions are of the form

$$\begin{cases} x = x_0 + \kappa \frac{b}{\gcd(a,b)}, \\ y = y_0 - \kappa \frac{a}{\gcd(a,b)}. \end{cases}$$

for $\kappa \in \mathbb{Z}$. It is important to remark that computing integral solutions is much harder than finding the rational ones.

* CIDMA, University of Aveiro. Email: apacetti@ua.pt

** Aknowledgements: This work was supported by CIDMA and is funded by the Portuguese Foundation for Science and Technology,.under grant UIDB/04106/2020 (https://doi.org/10.54499/UIDB/04106/2020)

3 Conics

Consider now the case of a degree two polynomial in the variables x, y, namely a polynomial of the form

$$aX^2 + bXY + cY^2 + dX + eY + f = 0, \qquad (2)$$

where we can assume that a, b, c, d, e, f are all integers (otherwise we can multiply by the minimum common multiple of their denominators). We will also assume that the degree two polynomial is irreducible (i.e. is not the product of two degree 1 ones), as otherwise the study of its rational/integral points reduces to the study of points on the factors.

We start studying rational solutions, say of the form $(X/Z, Y/Z)$ where X, Y, Z are integers. Substituting in (2) and multiplying by Z^2, we obtain an integral point on the *projective* conic

$$aX^2 + bXY + cY^2 + dXZ + eYZ + fZ^2 = 0, \quad (3)$$

where $Z \neq 0$ (it is customary to consider solutions where at least one of the coordinates is non-zero, as they correspond to points in the projective plane).

A general conic like (3) might not have integer solutions for easy reasons, for example there are no solutions to the equation $X^2 + Y^2 + Z^2$ other than $(0, 0, 0)$ (which we do not consider). In this case, the failure for a solution to exist comes from the fact that there are no real solutions to it (this is called an *Archimedean* failure). There might be other failures.

Example 1.— The conic $X^2 + Y^2 = 3Z^2$ has no non-trivial solution.

Suppose it does have a non-trivial solution (a, b, c) and assume that $\gcd(a, b, c) = 1$ (otherwise, we can divide each a, b, c by $\gcd(a, b, c)$ obtaining a new solution with this property). Note that if we divide any square by 4, the remainder is either 0 or 1 (or in terms of congruences, $x^2 \equiv 0, 1 \pmod 4$). Then $a^2 + b^2$ while divided by 4 has reminder 0, 1 or 2. Note that it is zero precisely when $2 \mid a$ and $2 \mid b$. Similarly, the reminder of $3c^2$ while divided by 4 is 0 or 3. Then the equality $a^2 + b^2 = 3c^3$ implies that both sides are divisible by 4, so $2 \mid c$ as well, contradicting the assumption that $\gcd(a, b, c) = 1$.

Contrary to the previous example, now the failure has to do with the prime 2, it is what is called a 2-adic failure (related to non-existence of solutions over the field $\mathbb{Q}_2$ of 2-adic numbers). A similar obstruction appears for the prime $p = 3$ (we leave the details to the reader).

It is natural to wonder whether the non-existence of solutions is always due to a *local* (i.e. attached to a congruence modulo N for some integer N) or an Archimedean reason. Indeed this is the case.

Theorem 2 (Hasse-Minkowski).— An equation like (3) has a non-zero rational solution if and only if it has a real solution, and a solution modulo N for all positive integers N.

Proof.— See for example Theorem 8 in [15]. ∎

The proof presented by Serre is different from our statement, so let us add a few comments. By the Chinese remainder Theorem (see §2.3 of [11]), searching for solutions modulo a general integer N is equivalent to search for solutions modulo prime powers. Once the prime p is fixed, the existence of a solution modulo p^n for all positive integers n is equivalent to the existence of a solution over the field of p-adic numbers. This is what Serre proves in [15].

Remark 1.— The result of Hasse-Minkowski works for homogeneous polynomials of degree 2 in any number of variables (not just 3).

Remark 2.— As stated Theorem 2 seems only of a theoretical nature (as it implies verifying infinitely many conditions). However, it is easy to transform it into a finite computation (it is enough to verify the statement at primes dividing the discriminant of the quadratic form together with the case $p = 2$). See for example §5.4 of [4].

Once that we have an *algorithm* to determine whether a conic has a rational point or not, it is natural to ask how many rational points it might have. The answer is infinitely many, as a conic with a point is isomorphic to a line, as proved in the following example.

Example 2.— Let us study the case of the unit circle centered at $A = (0, 0)$ with equation

$$\mathscr{C} : X^2 + Y^2 = 1. \qquad (4)$$

Take the point $B = (1, 0)$ (which belongs to the circle). Take the tangent line at B and translate it by some non-zero rational number (for example one to the right as in Figure 1).

Call the line L. Then we get a bijective map from rational points on $\mathscr{C}$ (removing the point B) to rational points on L as follows: given a rational point C in L, consider the line going through C and B. It must intersect the circle $\mathscr{C}$ in a rational point (why?). Explicitly, if $C = (2, y)$ then the second intersection point has coordinates

$$\left(\frac{y^2 - 1}{y^2 + 1}, \frac{-2y}{y^2 + 1}\right). \qquad (5)$$

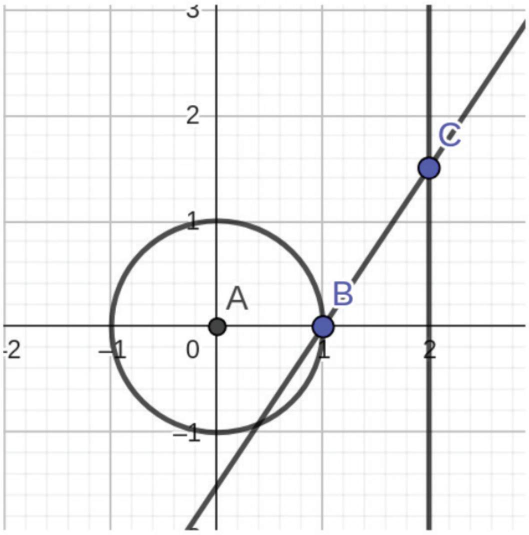


Figure 1.—Rational points circle

The inverse sends a point (x, y) in $\mathscr{C}$ to the point $(2, y/(x-1))$. The reason we need to remove the point B is that the affine line is not compact, if we add its point at infinity, then we really get a bijection between the two sets.

As done with the general equation (2), the set of rational points on the unit circle is the same as the set of integral points on the projective curve $X^2+Y^2 = Z^2$ (the so called Pitagorean triples). Writing the rational point y of the line L in the form $y = \frac{m}{n}$ (for $m, n \in \mathbb{Z}$) we recover the classical parametrization of the Pitagorean triples

$$(m^2 - n^2, -2mn, m^2 + n^2). \tag{6}$$

Remark 3.— The same construction/strategy works for a general conic as in (3) with one rational point.

Remark 4.— Over an algebraically closed field, equation (2) always has a point, hence it is isomorphic to a line. From a topological point of view, a line and a conic are the same, they both are genus 0 curves (or equivalently Riemann surfaces with no holes).

The problem of determining the set of integral points on a conic is much harder. There might be no points at all (as in Example 1), there might be finitely many (for example it is easy to verify that the circle (4) only has the four integral points $\{(\pm 1, 0), (0, \pm 1)\}$) or there might be infinitely many. For example, let d be a square-free positive integer, and consider Pell's equation

$$X^2 - dY^2 = 1. \tag{7}$$

The equation has infinitely many integral solutions, and all of them (up to a sign) can be obtained as *powers* of a particular one (see for example §7.8 of [11]). This equation appears while studying the *integers* whose inverses are also integers in the quadratic field $\mathbb{Q}(\sqrt{d})$.

4 Cubics

As mentioned before, we are mostly interested in studying hypersurfaces, i.e. solutions of a single equation $F(x_1, \dots, x_n) = 0$ (furthermore, most of the time we restrict to $n = 2$). The hypersurface

$$\mathscr{C} : F(x_1, \dots, x_n) = 0$$

is *non-singular* (or smooth) is there are no points P in $\mathscr{C}$ satisfying that $\frac{\partial F}{\partial x_i}(P) = 0$ for all $i = 1, \dots, n$. All lines are smooth, and conics given by an irreducible polynomial are smooth as well.

Suppose that $F(x, y)$ is a cubic (i.e. it has degree 3), and that the curve

$$\mathscr{C} : F(x, y) = 0$$

is non-singular. How can we determine whether it has a rational point or not?

As happened before, it is better to work with an homogeneous polynomial $F(X, Y, Z)$ in 3 variables. Its set of solutions corresponds to a cubic in the projective plane, and we are trying to determine whether it has an integral point different from $(0, 0, 0)$.

The first approach would be to use Hasse's criterion, i.e. try to search for points modulo N for different values of N. If no such a point exists, then we have proved that the curve $\mathscr{C}$ has no rational points.

Theorem 3 (Selmer).— The cubic equation

$$3X^3 + 4Y^3 + 5Z^3 = 0$$

has the only solution $(0, 0, 0)$ over $\mathbb{Q}$, but it has a nonzero solution over $\mathbb{R}$ and modulo N for all N.

Proof.— See [14]. ∎

Selmer's example shows that Hasse-Minkowski result does not hold for degrees larger than 2. Let us state some very interesting density results.

Theorem 4.— The probability that a random plane cubic curve over $\mathbb{Q}$ has a point modulo N for all positive values of N is approximately 97.256%

Proof.— See Theorem 2 of [2]. ∎

Remark 5.— Unlike conics, a cubic polynomial always has a real root, so the only failures can be local ones.

Furthermore, it was proved by Bhargava (see https://arxiv.org/pdf/1402.1131.pdf) that a positive proportion of cubics (at least 28%) fail the Hasse principle, so another approach is needed.

Assuming a deep open conjecture (namely finiteness of the Tate-Shafarevich group), there does exist an algorithm to determine if a cubic has a rational point or not. In practice, running the algorithm in some particular bad behaved cases might be very challenging.

Definition 5.— An *elliptic curve* is a non-singular cubic with a rational point.

The usual definition of an elliptic curve is that of a non-singular genus 1 curve with a rational point. It is not hard to prove (see for example §III.3 of [17]) that any rational elliptic curve can be given by a Weierstrass equation

$$E : Y^2 = X^3 + AX + B, \tag{8}$$

for $A, B \in \mathbb{Q}$ with $4A^3 + 27B^2 \neq 0$ (so the curve is smooth). The marked rational point corresponds to the solution $O = (0 : 1 : 0)$ of the homogeneous polynomial

$$ZY^2 = X^3 + AXZ^2 + BZ^3.$$

The point O is the unique point at the infinity line which we do not see while working on the affine plane.

Elliptic curves are very interesting objects. If K is any field (like $\mathbb{Q}$ or $\mathbb{C}$), the set $E(K)$ of points on E defined over K has an addition law (see §III.2 of [17]), making $(E(K), +)$ an abelian group (whose identity element is the point O).

Theorem 6 (Mordell).— The group $E(\mathbb{Q})$ is finitely generated.

A proof is given in §VIII of [17]. In particular, the fundamental theorem of finitely generated abelian groups implies that there exists a non-negative integer r such that

$$E(\mathbb{Q}) \simeq T \times \mathbb{Z}^r,$$

where T is a finite group. The number r is called the *rank* of the elliptic curve. There are very effective algorithms to compute T. Furthermore, a conjecture of Beppo Levi proven by Mazur states that there are only 15 possible groups for T (see Theorem 7.5 of [17]). Computing r (and generators for the free part) is a deep problem. Once again, assuming finiteness of the Tate-Shafarevich group, there exists a theoretical algorithm to do it.

Remark 6.— It is not known whether the value of r is bounded or not. The current largest value for it is 28, found by Elkies in 2006.

Regarding integral points, there is a very general result due to Siegel ([16]), which states the following.

Theorem 7.— If $F(x, y)$ is a polynomial of degree larger than 2 and the curve $\mathscr{C} : \{F(x, y) = 0\}$ is non-singular, then $\mathscr{C}$ has finitely many integral points.

The result is not effective (i.e. it does not give information on the number of integral points of $\mathscr{C}$ nor how to compute them). In the case of elliptic curves, the elements of T have integral coordinates. If a set of generators for $E(\mathbb{Q})$ is known then a priori one can determine all the integral points on E.

5 Larger degrees

If the polynomial $F(x, y)$ has degree larger than 3, the (non-singular) curve $\mathscr{C}$ has genus larger than 1. The following result is a deep conjecture of Mordell proven by Falting ([9]).

Theorem 8 (Faltings).— If $\mathscr{C}$ is a rational non-singular curve of genus larger than 1 then $\mathscr{C}(\mathbb{Q})$ is finite.

As Siegel's theorem, the proof is not effective. In a remarkable article, Chabauty ([3]) gave a method to bound the number of rational points when the rank of the Jacobian of $\mathscr{C}$ is smaller than its genus. An effective version of the method was obtained by Coleman in ([6]). Since then, many improvements have been obtained, making the Chabauty method a very active research area.

6 Fermat's last theorem

Without getting into details of the history behind Fermat's last theorem, in a margin of his copy of Diophantus' *Arithmetica*, Fermat wrote that a cube cannot be written as the sum of two cubes, a fourth power as a sum as two fourth powers, etc. In other words, his claim can be stated as:

Theorem 9 (Fermat's last theorem).— The equation

$$X^n + Y^n = Z^n \tag{9}$$

has no rational solutions other than the trivial ones (i.e. when one of the variables equals zero).

After the contributions of many mathematicians, Fermat's last theorem was finally proved in 1995 by Wiles (see [19]). The book [8] contains details of the history behind the problem as well as different strategies used to solve particular cases before the Frey-Hellegouarch approach used in Wiles' proof.

Historically, a major breakthrough for understanding Fermat's last theorem was Faltings' result, which implies the existence of finitely many solutions for each $n > 3$.

The case $n = 3$ is of particular interest, as it is a cubic curve, with a rational point (actually with 3 different ones up to multiplication by -1). Substituting (X, Y, Z) by $(y/9, x/3, y/9z)$ in (9) (when $n = 3$) and multiplying the equation by 27 gives the curve in Weirestrass form

$$y^2z + 9yz^2 = x^3 - 27z^3.$$

Any modern number theory software (like [18]) verifies that this curve has only three rational points, namely $(3 : 0 : 1)$, $(3 : -9 : 1)$ and $(0 : 1 : 0)$ (mapping to the points $(0 : 1 : 1)$, $(-1 : 1 : 0)$ and $(1 : 0 : 1)$ respectively), proving Fermat's last theorem when $n = 3$.

The general proof of Fermat's last theorem is very technical, but we content ourselves to stating a few ingredients of the proof: start with a putative solution (a, b, c) of (9) satisfying that $\gcd(a, b, c) = 1$ and $abc \neq 0$ (to avoid the trivial solutions). It is enough to prove the statement when n is a prime number, and when $n = 4$. The case $n = 4$ was proved by Fermat, so suppose that n is an odd prime number ℓ.

1. Attach to the solution the Frey elliptic curve
$$E : Y^2Z = X(X - a^\ell Z)(X + b^\ell Z).$$
2. Wiles proved that this elliptic curve is *modular* i.e. is related to an holomorphic function of the complex upper half plane satisfying many transformation properties (such functions are called *modular forms*). The number of equations depend on a parameter N (a positive integer) called the *level* of the modular form. For the experienced reader, the modular form has weight 2 and is invariant under the group $\Gamma_0(N)$.
3. For each value of N, the set of modular forms satisfying the relations given by the value N is actually a finite dimensional vector space. There are many algorithms to compute a basis for it (using the so called *modular symbols*). The problem is that the value of N attached to E depends on a, b and c (which are unknown).
4. Making use of the particular shape of a solution, results of Hellegouarch and Ribet imply that actually one can take (up to a congruence) $N = 2$.
5. The space of modular forms for the parameter $N = 2$ is zero, so there is no form in this space to match the curve E attached to our solution. This gives a contradiction, so the original solution (a, b, c) cannot exist.

As previously mentioned, the proof follows from the effort and contributions of many mathematicians, including Frey, Hellegouarch, Mazur, Ribet, Serre, Wiles and Taylor among others.

7 The generalized Fermat equation

Let a, b, c, p, q, r be non-zero positive integers. The so called *generalized Fermat equation* is the equation

$$ax^p + by^q = cz^r. \tag{10}$$

The case $a = b = c = 1$ and $p = q = r$ is the classical Fermat's equation. There is a big difference between equation (10) and Fermat's one, since the former defines an affine surface (instead of a projective curve). There are many examples of surfaces containing lines (like a cone, although it is a singular surface). For this reason, the number of solutions to (10) depends on whether $(1/p) + (1/q) + (1/r)$ is larger than 1, equals 1 or is smaller than 1. See ([1]) for a nice exposition in the case $a = b = c = 1$.

The first case (called spherical) corresponds to the exponents $(2, 2, r)$, $(2, q, 2)$, $(2, 3, 3)$, $(2, 3, 4)$, $(2, 4, 3)$ or $(2, 3, 5)$. In general one expects that if one solution exists, then there are infinitely many (and the solutions can be parametrized). See §14 of [5].

The second case (called parabolic) corresponds to the exponents $(2,6,3)$, $(2,4,4)$, $(4,4,2)$, $(3,3,3)$ or $(2,3,6)$. In this cases one also expects that if one solution exists, then there should be infinitely many of them (but we do not expect a parametrization). See §6 of [7] and also §6.5 of [4].

The last case (called hyperbolic) is the general one. Note that since the polynomial giving (10) is not homogeneous, we cannot *assume* that our solution is *primitive* (i.e. $\gcd(x,y,z) = 1$). This phenomenom gives raise to the existence of many *unwanted* solutions.

Here is an example taken from [7]: consider the equation

$$x^3 + y^3 = z^4.$$

(corresponding to equation (10) with parameters $(a,b,c,p,q,r) = (1,1,1,3,3,4)$). Let $z = \alpha^3 + \beta^3$, $x = \alpha z$, $y = \beta z$ for α, β arbitrary integers. These are all solutions! (though all of them but finitely many are not primitive). For this purpose, one focus on studying only primitive solutions. Here is a very nice general result.

Theorem 10 (Darmon-Granville).— If $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$ then equation (10) has finitely many primitive solutions.

The proof (see [7]) depends on Mordell's conjecture (Theorem 8), so it is not effective. It is expected that once (a,b,c) is fixed, the set of primitive solutions (where the exponents (p,q,r) vary) is still finite. Here is an explicit version of what we expect to be true.

Conjecture 1.— Any primitive solution to

$$x^p + y^q = z^r,$$

with $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$ is either the solution $1^p + 2^3 = 3^2$, or it belongs to a finite list.

In other words, if we vary the exponents (p,q,r) with the condition that the equation is hyperbolic, then the union of all solutions is a finite set. There is an explicit candidate for the finite list (based on numerical computations) which we omit for space reasons. They all have the property that one of p, q or r equals 2. A conjecture of Beal (with a prize of 1 million USD for its resolution) actually states that there are no solution if $\min\{p,q,r\} > 2$.

8 Our contribution to the problem

Together with my former student Lucas Villagra Torcomian, we study the particular generalized Fermat equations:

$$x^4 + dy^2 = z^p, \tag{11}$$

and

$$x^2 + dy^6 = z^p, \tag{12}$$

for different values of d. In [12], following the modular method used in the proof of Fermat's last theorem, we gave an algorithm that for fixed d, proves (in many instances) the non-existence of solutions for any large value of the exponent p (assuming it is a prime number). Here are a few particular instances of the results proven in [12].

Theorem 11.— There are no non-trivial primitive solutions of $x^4 + 5y^2 = z^p$ if p is any prime number larger than 499.

The result should hold for small values of p as well (say larger than 13), but getting this bound requires a huge computational effort that is unfeasible nowadays.

Theorem 12.— There are no non-trivial primitive solutions of $x^2 + 6y^6 = z^p$ if p is any prime number larger than 563.

When $d < 0$ equations (11) and (12) become harder to study. However, in [13] we proved some partial results like the following.

Theorem 13.— Let $p > 19$ be a prime number such that $p \neq 97$ and $p \equiv 1, 3 \pmod 8$. Then $(\pm 7, \pm 20, 1)$ are the only non-trivial primitive solutions of the equation $x^4 - 6y^2 = z^p$.

The aforementioned results depend on a computation for each value of the parameter d. Recently, in [10] we obtained the following asymptotic result.

Theorem 14.— Let d be a prime number congruent to 3 modulo 8 and such that the class number of $\mathbb{Q}(\sqrt{-d})$ is not divisible by 3. Then there are no non-trivial primitive solutions of the equation

$$x^4 + dy^2 = z^p,$$

for p large enough.

A similar result was obtained for the equation $x^2 + dy^6 = z^p$, namely that if d is a prime number congruent to 19 modulo 24 and such that the class number of $\mathbb{Q}(\sqrt{-d})$ is not divisible by 3, then the equation $x^2 + dy^6 = z^p$ does not have non-trivial primitive solutions for p large enough.

References

[1] Michael Bennett, Preda Mihăilescu, and Samir Siksek. The generalized Fermat equation. In *Open problems in mathematics*, pages 173–205. Springer, [Cham], 2016.

[2] Manjul Bhargava, John Cremona, and Tom Fisher. The proportion of plane cubic curves over $\mathbb{Q}$ that everywhere locally have a point. *Int. J. Number Theory*, 12(4):1077–1092, 2016.

[3] Claude Chabauty. Sur les points rationnels des courbes algébriques de genre supérieur à l'unité. *C. R. Acad. Sci. Paris*, 212:882–885, 1941.

[4] Henri Cohen. *Number theory. Vol. I. Tools and Diophantine equations*, volume 239 of *Graduate Texts in Mathematics*. Springer, New York, 2007.

[5] Henri Cohen. *Number theory. Vol. II. Analytic and modern tools*, volume 240 of *Graduate Texts in Mathematics*. Springer, New York, 2007.

[6] Robert F. Coleman. Effective Chabauty. *Duke Math. J.*, 52(3):765–770, 1985.

[7] Henri Darmon and Andrew Granville. On the equations $z^m = F(x,y)$ and $Ax^p + By^q = Cz^r$. *Bull. London Math. Soc.*, 27(6):513–543, 1995.

[8] Harold M. Edwards. *Fermat's last theorem*, volume 50 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996. A genetic introduction to algebraic number theory, Corrected reprint of the 1977 original.

[9] G. Faltings. Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. *Invent. Math.*, 73(3):349–366, 1983.

[10] Franco Golfieri Madriaga, Ariel Pacetti, and Lucas Villagra Torcomian. Asymptotic results for the equations $x^4 + dy^2 = z^p$ and $x^2 + dy^6 = z^p$, 2022.

[11] Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery. *An introduction to the theory of numbers*. John Wiley & Sons, Inc., New York, fifth edition, 1991.

[12] Ariel Pacetti and Lucas Villagra Torcomian. $\mathbb{Q}$-Curves, Hecke characters and some Diophantine equations. *Math. Comp.*, 91(338):2817–2865, 2022.

[13] Ariel Pacetti and Lucas Villagra Torcomian. $\mathbb{Q}$-curves, Hecke characters and some Diophantine equations II. *Publ. Mat.*, 67(2):569–599, 2023.

[14] Ernst S. Selmer. The Diophantine equation $ax^3 + by^3 + cz^3 = 0$. *Acta Math.*, 85:203–362 (1 plate), 1951.

[15] J.-P. Serre. *A course in arithmetic.* Springer-Verlag, New York-Heidelberg, 1973. Translated from the French.

[16] Carl L. Siegel. Über einige Anwendungen diophantischer Approximationen [reprint of Abhandlungen der Preußischen Akademie der Wissenschaften. Physikalisch-mathematische Klasse 1929, Nr. 1]. In *On some applications of Diophantine approximations*, volume 2 of *Quad./Monogr.*, pages 81–138. Ed. Norm., Pisa, 2014.

[17] Joseph H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics*. Springer, Dordrecht, second edition, 2009.

[18] The PARI Group, Univ. Bordeaux. *PARI/GP version 2.13.4*, 2022.Available from http://pari.math.u-bordeaux.fr/.

[19] Andrew Wiles. Modular elliptic curves and Fermat's last theorem. *Ann. of Math. (2)*, 141(3):443–551, 1995.

LxDS Spring School 2023

by Telmo Peixe*

The LxDS-Lisbon Dynamical Systems Group, CEMAPRE and CMAF-CIO organized the LxDS Spring School 2023 on the days 29-31 of May 2023 in dynamical systems, which took place at ISEG-Lisbon School of Economics & Management, Universidade de Lisboa.

The school consisted of three courses in dynamical systems, which were given by specialists of recognized international merit. Namely,

- *Diophantine approximations and the space of lattices* (**Nicolas Chevallier**, Université de Haute Alsace),
- *Statistical properties for certain dynamical systems* (**Silvius Klein**, Pontifícia Universidade Católica do Rio de Janeiro),
- *Game dynamics* (**Josef Hofbauer**, University Vienna).

The school had about 20 participants, some of them international.

In addition to the courses the school also had a session of oral presentations, in which some of the PhD students and researchers presented their most recent work.

Due to financial support provided by CIM, it was possible to support the participation of three PhD students covering their travel, lodging and meals during the school days.

The organizing committee:

João Lopes Dias
Universidade de Lisboa, ISEG, CEMAPRE

José Pedro Gaivão
Universidade de Lisboa, ISEG, CEMAPRE

Pedro Miguel Duarte
Universidade de Lisboa, FCUL, CMAFCIO

Telmo Peixe
Universidade de Lisboa, ISEG, CEMAPRE

* Department of Mathematics, ISEG, Universidade de Lisboa. Email: telmop@iseg.ulisboa.pt

REPORT

Particle Systems and PDEs XI

by **Ana Jacinta Soares*** and **Patrícia Gonçalves****

The International Conference on Particle Systems and Partial Differential Equations XI was hosted at Instituto Superior Técnico (IST), University of Lisbon, from November 6th to 10th, 2023.

This marked the eleventh edition of the conference, following seven editions at the University of Minho, Braga, from 2012 to 2016, and 2021 to 2022, one edition at the University of Nice Sophia Antipolis, France, in 2017, another at the University of Palermo, Italy, in 2018, and one previously at IST Lisbon in 2019.

The main objective of the conference was to gather leading active researchers in the fields of particle systems and partial differential equations, providing an opportunity to present their latest findings, promote the discussion and exchange of ideas, and encouraging the establishment of new scientific collaborations.

The scientific program included two mini-courses presented by distinguished researchers, Michael Loss from the Georgia Institute of Technology, Atlanta, U.S.A., and Pablo Ferrari from the University of Buenos Aires, Argentina. Additionally, it included twenty-seven talks by invited speakers, along with two poster sessions. The conference drew a diverse audience of fifty participants from fourteen countries, encompassing numerous young researchers and Ph.D. students.

The conference was a joint initiative organized by the University of Lisbon (CAMGSD, CMAFcIO), University of Minho (CMAT) and the University of Nice Sophia Antipolis, France. Financial support was provided by CAMGSD, CMAFcIO, CMAT, CIM, ERC, FCT.

For further details, please visit https://sites.google.com/view/pspde/

* CMAT e Dep. de Matemática, Univ. do Minho. Email: ajsoares@math.uminho.pt
** CAMGSD e Dep. de Matemática do IST, Univ. de Lisboa. Email: pgoncalves@tecnico.ulisboa.pt

Organizing committee

Cédric Bernardin
Université Côte d'Azur

Patrícia Gonçalves
CAMGSD, University of Lisbon

Maria João Oliveira
Universidade Aberta and CMAFcIO

Gunter Schütz
CAMGSD, University of Lisbon

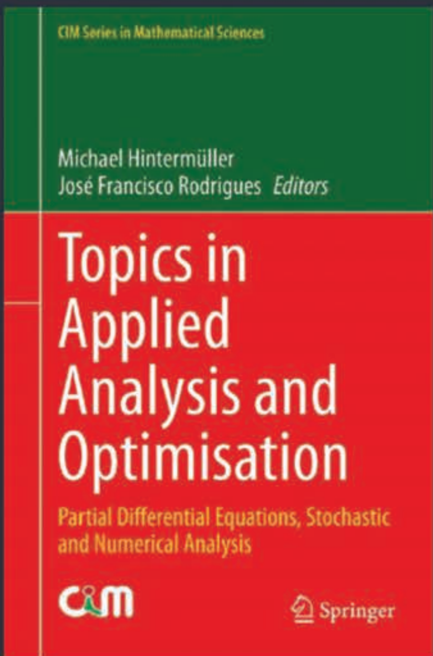
Ana Jacinta Soares
CMAT, University of Minho

Ali Zahra
CAMGSD, University of Lisbon

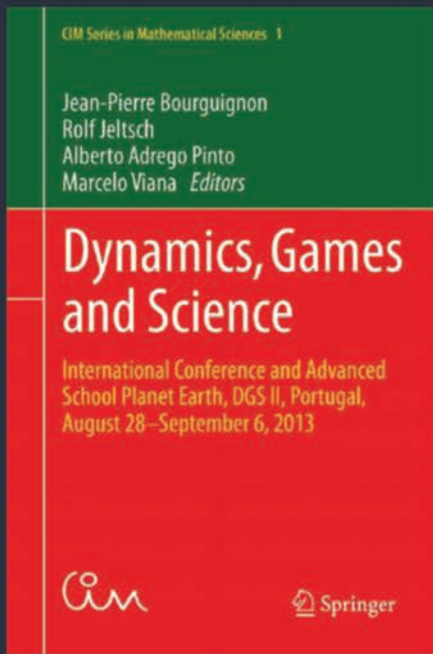
CIM Series in Mathematical Sciences

Series Editors: I. Fonseca, I.M.N. Figueiredo

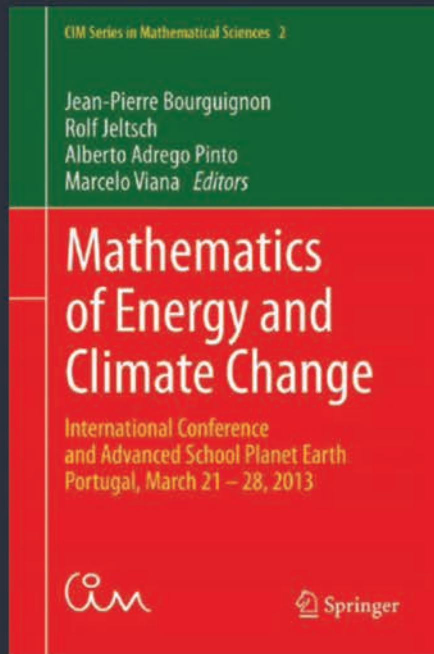
The CIM Series in Mathematical Sciences is published on behalf of and in collaboration with the Centro Internacional de Matemática (CIM) in Portugal. Proceedings, lecture course material from summer schools and research monographs will be included in the new series.



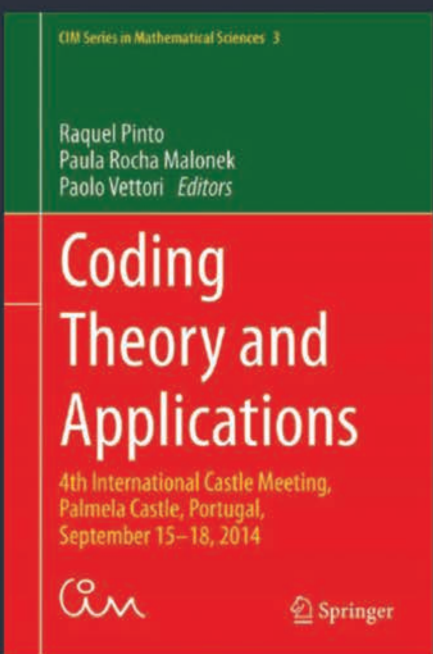
1st ed. 2019, XV, 396 p. 55 illus., 45 illus. in color



1st ed. 2015, XVIII, 772 p. 141 illus., 94 illus. in color



1st ed. 2015, XVI, 430 p. 120 illus., 78 illus. in color



1st ed. 2015, IX, 365 p. 37 illus., 25 illus. in color.

1st ed. 2015, XI, 465 p.

On the geometry of learning from data: Bayes meets Hilbert

by **Miguel de Carvalho***

Abstract.—Bayes' theorem is a central result of Statistics and related fields, such as Artificial Intelligence and Machine Learning. In this note, we offer a gentle introduction to a geometric interpretation of Bayesian inference that allows one to think of priors, likelihoods, and posteriors as vectors in an Hilbert space. The given framework can be conceptualized as a geometry of learning from data, and it can be used to construct measures of agreement between these vectors. Conceptually, the geometry is tantamount to that of Pearson correlation, but where an inner product is considered over the parameter space—rather than over the sample space.

1 Introduction

This note builds on ideas from two prominent thinkers: Thomas Bayes (c. 1701–1761) and David Hilbert (1862–1943).[1] While their lives never overlapped temporally, this note shows how the work of Hilbert can be used to reinterpret Bayes' theorem and Bayesian inference from a geometric viewpoint—as well as other key statistical concepts on what we regard as a geometry of learning from data.

The Bayesian paradigm is a well-known statistical inference approach that can be used for learning from data about a parameter of statistical interest using Bayes theorem. Let $Y_1, \ldots, Y_n$ be a sequence of independent and identically distributed (iid) random variables in a measurable space $(\Omega, \mathscr{A})$ that are drawn from parametric density function

$$f_\theta(y) \equiv f(y \mid \theta),$$

with $y \in \Omega$ and $\theta \in \Theta$. The sets Ω and Θ are respectively known as *sample space* and *parameter space*.

The key goal of Bayesian inference is to learn about the distribution of the parameter θ given the data $y = (y_1, \ldots, y_n)$. It follows from Bayes theorem that,

$$p(\theta \mid y) = \frac{\pi(\theta)\ell(\theta)}{\int_\Theta \pi(u)\ell(u)\, \mathrm{d}u}. \tag{1}$$

where $\ell(\theta) = \prod_{i=1}^n f_\theta(y_i)$ is the likelihood function, and $\pi(\theta)$ is the prior density function. The density $p(\theta \mid y)$ is known as posterior density and it summarizes what we learn about θ after observing y.

The prior density can understood as a way adding prior knowledge about θ to the analysis—say, from an expert opinion, from a census, and so on—or simply as a way to "initiate the inferential machine." Quoting [9]:

> The choice of a prior distribution is necessary (as you would need to initiate the inferential machine) but there is no notion of the "optimal" prior distribution. Choosing a prior distribution is similar in principle to initializing any other sequential procedure (e.g., iterative optimization methods […] etc.). The choice of such initialization can be good or bad in the sense of the rate of convergence of the procedure to its final value, but as long as the procedure is guaranteed to converge, the choice of prior does not have a permanent impact.

[1] The key concepts and methods from this note relate with the ideas and principles in [3], which was awarded with the 2018 Lindley Prize from the International Society of Bayesian Analysis.

* School of Mathematics, University of Edinburgh; Department of Mathematics, Universidade de Aveiro
Email: Miguel.deCarvalho@ed.ac.uk

And indeed, the posterior can be shown to converge to the true value, under rather general conditions on the prior distribution—a result known in statistical parlance as the Bernstein—von Mises theorem [11, Theorem 10.1].

The remainder of this note is organized as follows. In §2 we note that there's an hidden geometry underlying Eq. (4) that can be used to rethink Bayesian inference and to develop measures of agreement between prior, likelihood, and posterior. In §3 we illustrate how that geometry can be used for shedding light on other statistical inference concepts.

Before we get started a disclaimer is in order. To make the presentation of the key ideas more accessible, we will often use visualizations based on Cartesian representations. Yet, it is important to remember that these representations are mainly heuristic and hence should be interpreted with care.

2 The geometry of Bayesian inference

2.1 Abstract geometry

We first clarify the sense in which the term geometry will be used throughout this note. The following definition of abstract geometry can be found in [7, p. 17].

Definition 1 (Abstract geometry).— An abstract geometry $\mathscr{A}$ consists of a pair $\{\mathscr{P}, \mathscr{L}\}$, where the elements of set $\mathscr{P}$ are designed as points, and the elements of the collection $\mathscr{L}$ are designed as lines, such that:

1. For every two points $A, B \in \mathscr{P}$, there is a line $l \in \mathscr{L}$.
2. Every line has at least two points.

Our abstract geometry of interest is $\mathscr{A} = \{\mathscr{P}, \mathscr{L}\}$, where $\mathscr{P} = L_2(\Theta)$ is the the space of square integrable functions, and the set of all lines is

$$\mathscr{L} = \{g + kh : g, h \in L_2(\Theta), k \in \mathbf{R}\}. \qquad (2)$$

Hence, in our setting points can be, for example, prior densities, posterior densities, or likelihoods, as long as they are in $L_2(\Theta)$. While not all priors and likelihoods are in $L_2(\Theta)$, the framework discussed herein may extend beyond $L_2(\Theta)$ with some modifications, while still allowing similar geometric interpretations as the ones provided below. See [3, §3] for details.

2.2 Bayes geometry

2.2.1 *The marginal likelihood is an inner product*

Suppose the goal of the inference is over a parameter θ which takes values on $\Theta \subseteq \mathbf{R}^p$. We use the geometry of the Hilbert space $\mathscr{H} = (L_2(\Theta), \langle\cdot,\cdot\rangle)$, with inner-product[2]

$$\langle g, h\rangle = \int_\Theta g(\theta)h(\theta)\,\mathrm{d}\theta, \quad g, h \in L_2(\Theta). \qquad (3)$$

Adopting the geometric terminology used in linear spaces, we denote the elements of $L_2(\Theta)$ as vectors, and assess their *magnitudes* through the use of the norm induced by the inner product in (3), i.e., $\|\cdot\| = (\langle\cdot,\cdot\rangle)^{1/2}$.

The starting point for constructing our geometry is the observation that Bayes theorem can be written using the inner-product in (2.2.1) as follows

$$p(\theta \mid y) = \frac{\pi(\theta)\ell(\theta)}{\langle\pi, \ell\rangle}, \qquad (4)$$

where $\langle\pi, \ell\rangle = \int_\Theta f(y \mid \theta)\pi(\theta)\,\mathrm{d}\theta$ is the so-called marginal likelihood. The inner product in (3) naturally leads to considering π and ℓ that are in $L_2(\Theta)$, which is compatible with a wealth of parametric models and proper priors.

As can be seen from Fig. 1, by considering p, π, and ℓ as vectors with different magnitudes and directions, Bayes' theorem essentially describes the method of reshaping the prior vector in order to derive the posterior vector. The likelihood vector amplifies or diminishes the magnitude of the prior vector, and appropriately adjusts its direction, in a way that will be clearly defined in the subsequent discussion.

The marginal likelihood $\langle\pi, \ell\rangle$ is simply the inner product between the likelihood and the prior, and thus can be interpreted as an assessment of the concordance between the prior and the likelihood. To provide a more tangible understanding, let's define the *angle measure* between the prior and the likeli-

[2] In mathematical terminology, the assertion that $\mathscr{H}$ constitutes a Hilbert space is frequently referred to as the Riesz–Fischer theorem. For a proof see [2, p. 411].

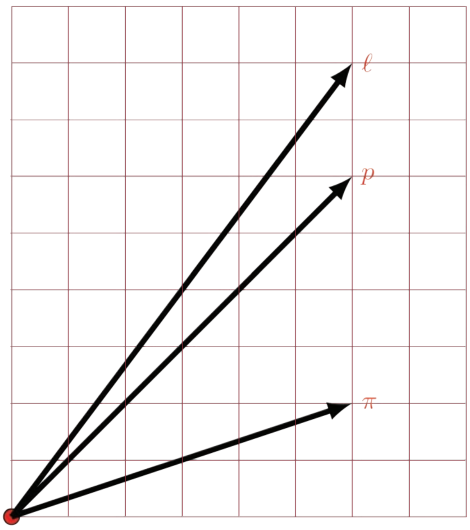


Figure 1.—Cartesian representation of vectors of interest in a Bayesian analysis.

hood as

$$\pi \angle \ell = \arccos \frac{\langle \pi, \ell \rangle}{\|\pi\| \|\ell\|}. \tag{5}$$

Since π and ℓ are nonnegative, the angle between the prior and the likelihood can only be acute or right, i.e., $\pi \angle \ell \in [0, 90^\circ]$. The closer $\pi \angle \ell$ is to 0°, the greater the agreement between the prior and the likelihood. Conversely, the closer $\pi \angle \ell$ is to 90°, the greater the disagreement between prior and likelihood. In the limiting case where $\pi \angle \ell = 90^\circ$—which implies the prior and the likelihood have all of their mass on disjoint sets—we say that the prior is orthogonal to the likelihood. Bayes theorem does not allow for a prior to be orthogonal to the likelihood as $\pi \angle \ell = 90^\circ$ implies that $\langle \pi, \ell \rangle = 0$, thus yielding a division by zero in (4).

2.2.2 *Compatibility*

The object we aim to focus next is given by a standardized inner product

$$\kappa_{\pi,\ell} = \frac{\langle \pi, \ell \rangle}{\|\pi\| \|\ell\|}. \tag{6}$$

The quantity $\kappa_{\pi,\ell} \in (0, 1]$ assesses the extent to which an expert's viewpoint aligns with the data, thereby offering an intuitive measurement of the concordance between the prior and the data.

Extending the principle in (6), for any two points in the geometry under consideration we define their compatibility as a standardized inner product.

Definition 2 (Compatibility).— The compatibility between points in the geometry under consideration is defined as

$$\kappa_{g,h} = \frac{\langle g, h \rangle}{\|g\| \|h\|}, \quad g, h \in L_2(\Theta). \tag{7}$$

Particular instances include (6) as well as:

- κ_{π_1,π_2}: which assesses the level of agreement between two experts, with respective priors π_1 and π_2.
- $\kappa_{\pi,p}$: which is a metric of the sensitivity of the posterior to the prior specification.

Example 1 (Beta-Bernoulli model).— Let

$$\begin{cases} Y_i \mid \theta \overset{\text{iid}}{\sim} \text{Bern}(\theta), & i = 1, \dots, n, \\ \theta \sim \text{Beta}(a, b). \end{cases} \tag{8}$$

Then, $\theta \mid y \sim \text{Beta}(a^\star, b^\star)$ with $a^\star = n_1 + a$ and $b^\star = n - n_1 + b$, where $n_1 = \sum_{i=1}^n y_i$.

The compatibility between prior and likelihood for this beta–Bernoulli model is

$$\kappa_{\pi,\ell} = \frac{B(a^\star, b^\star)}{\{B(2a-1, 2b-1)B(2n_1+1, 2(n-n_1)+1)\}^{1/2}},$$

for $a, b > 1/2$, with $B(a, b) = \int_0^1 u^{a-1}(1-u)^{b-1}\, du$.[3] To assess how compatible the priors $\pi_1 \sim \text{Beta}(a_1, b_1)$ and $\pi_2 \sim \text{Beta}(a_2, b_2)$ are, we obtain

$$\kappa_{\pi_1,\pi_2} = \frac{B(a_1 + a_2 - 1, b_1 + b_2 - 1)}{\{B(2a_1 - 1, 2b_1 - 1)B(2a_2 - 1, 2b_2 - 1)\}^{1/2}}.$$

for $a_1, a_2, b_1, b_2 > 1/2$.

[3] The geometry underlying compatibility can be reframed within an Hellinger affinity context so to allow for any $a, b > 0$. See [3, §3].

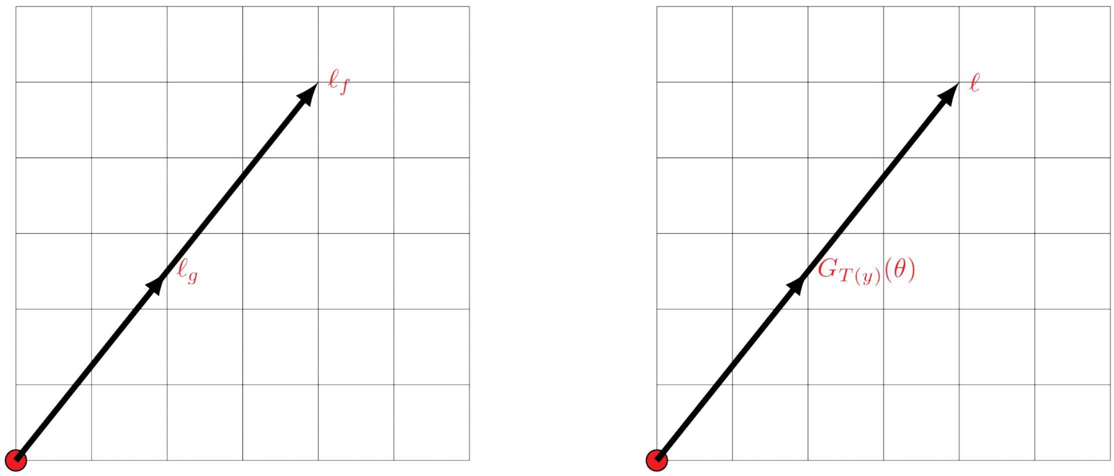


Figure 2.—Cartesian representation underlying the strong likelihood principle (left) and sufficiency (right). See §§ 3.2 and 3.3.

3 Further perspectives and insights

The roadmap for this section is as follows. §3.1 notes that a variational representation of the posterior density naturally fits our geometry. §§3.2 and 3.3 are related with collinearity; it follows from §2, whenever the symbol "∝" is used in a Bayesian setting it simply implies that two likelihoods, priors or posteriors are collinear. Finally, §3.4 notes the similarities between the geometry of compabitility and that of Pearson correlation.

3.1 Donsker–Varadhan representation

The celebrated Donsker–Varadhan representation shows that the posterior density is the solution to a variational problem with search domain $\mathscr{P}(\Theta)$; here and below, $\mathscr{P}(\Theta)$ is the space of probability density functions that can be defined over Θ and $l(\theta) = \log \ell(\theta)$ is the log likelihood. Specifically, the Donsker–Varadhan representation is given by

$$p(\theta \mid y) = \arg \min_{q \in \mathscr{P}(\Theta)} [-\mathrm{E}_q\{l(\theta)\} + \mathrm{KL}(q, \pi)], \qquad (9)$$

where E_q and KL are respectively the prior expectation and Kullback–Leibler divergence, that is,

$$E_q\{l(\theta)\} = \int_\Theta l(\theta)\, q(\theta)\, \mathrm{d}\theta,$$

$$\mathrm{KL}(q, \pi) = \int_\Theta q(\theta) \log\{q(\theta)/\pi(\theta)\}\, \mathrm{d}\theta.$$

A geometric interpretation of (3.1) follows from elementary properties of inner products,

$$\begin{aligned} p(\theta \mid y) &= \arg \min_{q \in \mathscr{P}(\Theta)} -\langle q, l \rangle + \langle q, \log(q/\pi) \rangle \\ &= \arg \max_{q \in \mathscr{P}(\Theta)} \langle q, l \rangle - \langle q, \log(q/\pi) \rangle \\ &= \arg \max_{q \in \mathscr{P}(\Theta)} \langle q, \mathrm{DV}_q \rangle, \end{aligned} \qquad (10)$$

where DV_q is what we refer to as the *Donsker–Varadhan likelihood ratio*,

$$\mathrm{DV}_q(\theta) \equiv \log[\ell(\theta)/\{q(\theta)\pi(\theta)\}]. \qquad (11)$$

Loosely, (10) implies that the posterior density is the density in $\mathscr{P}(\Theta)$ which is most lined up with the Donsker–Varadhan likelihood ratio in (11).

3.2 Collinearity, I: likelihood principle

Let ℓ_f and ℓ_g be the likelihoods based on observing $y \sim f$ and $y^* \sim g$, respectively. The strong likelihood principle states that if

$$\ell_f(\theta) = f(\theta \mid y) \propto g(\theta \mid y^*) = \ell_g(\theta),$$

then the same inference should be drawn from both samples. According to our geometry, this means that likelihoods with the same direction yield the same inference. For instance, the Bernoulli likelihood of the model from Example (1) is

$$\ell_f(\theta) = \prod_{i=1}^n \theta^{y_i}(1-\theta)^{n-y_i} = \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i},$$

wheras that of the Binomial model for $n_1 = \sum_{i=1}^n y_i$ is

$$\ell_g(\theta) = \binom{n}{n_1} \theta^{n_1}(1-\theta)^{n-n_1},$$

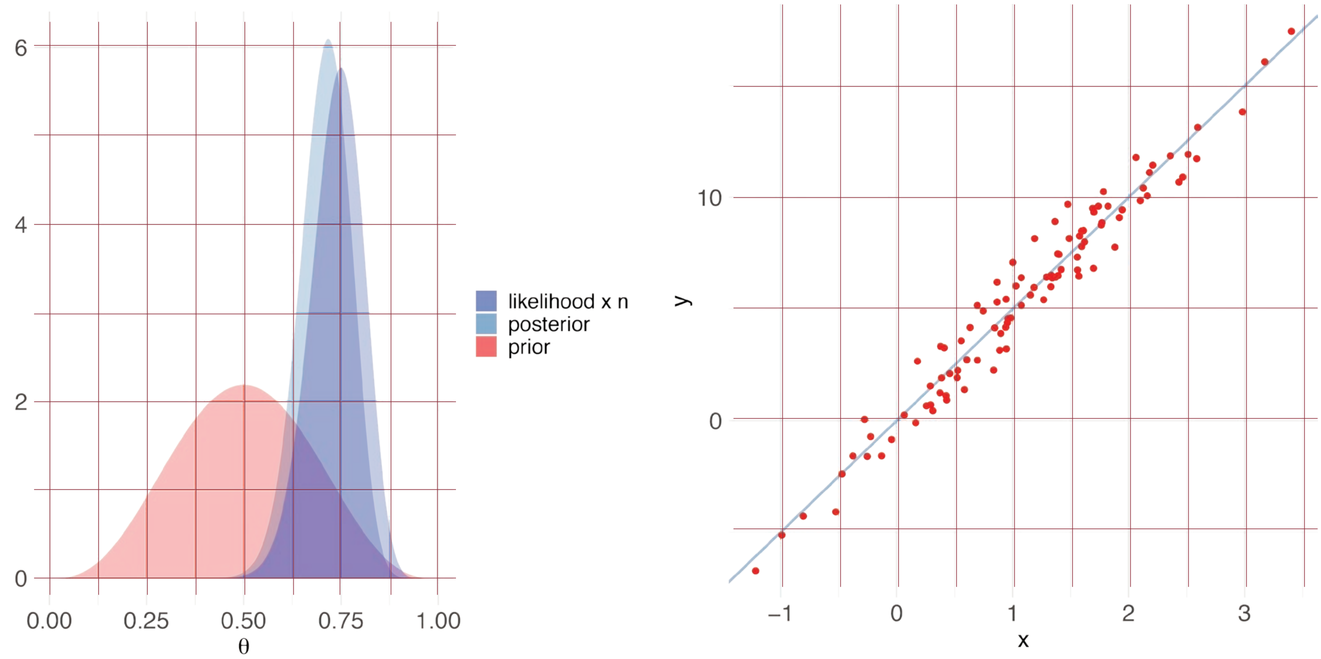


Figure 3.—Left: Prior, posterioar, and likelihood for beta–binomial specification from Example 1 with $(a,b) = (4, 4)$, $n = 40$, and $n_1 = 30$ so that, for example, $\kappa_{\pi,l} = 0.41$. Right: Simulated data from bivariate normal distribution with $\rho_{X,Y} = 0.98$.

with $\binom{a}{b}$ denoting the binomial coefficient. Trivially,

$$\ell_f(\theta) \propto \ell_g(\theta),$$

and hence ℓ_f and ℓ_g are collinear.

3.3 Collinearity, II: sufficiency

Roughly speaking, a sufficient statistic is one that contains all the information that is required to learn about θ.[4] The geometry from §2.2 can also be used to rethink a celebrated characterization of sufficient statistics in a geometric fashion.

Theorem 3 (Neyman factorization).— Suppose that $Y = (Y_1, \dots, Y_n)$ has a joint density function or a frequency function $f_\theta(y)$. Then $T(Y)$ is sufficient for θ iff there exists a function of that statistic, $G_{T(y)}(\theta)$, that is collinear to $\ell(\theta)$, that is,

$$\ell(\theta) \propto G_{T(y)}(\theta).$$

See, for instance, [6, §4] for a nongeometrical formulation of this classical result. Let's illustrate this on a well-known example.

Example 2.— Let $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. It can be easily shown that

$$\ell(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbf{1}_{[0,\theta]}(y_i) \propto \frac{1}{\theta^n} \mathbf{1}_{[0,\theta]}\{T(y)\} \equiv G_{T(y)}(\theta),$$

where $T(y) = \max\{y_1, \dots, y_n\}$ and $\mathbf{1}_A$ is the indicator function.

3.4 Compatibility vs Pearson correlation

Compatibility in Definition 2 follows the same construction principles as the Pearson correlation coefficient, which is based on the inner product

$$\langle X, Y \rangle = \int_\Omega XY \, \mathrm{d}P, \quad X, Y \in L_2(\Omega, \mathbb{B}_\Omega, P), \quad (12)$$

instead of the inner product in (3). Recall that Pearson correlation is defined as

$$\rho_{X,Y} = \frac{\operatorname{cov}(X, Y)}{\operatorname{sd}(X)\operatorname{sd}(Y)},$$

and it can be understood as a cosine of $X \angle Y$ in a similar fashion as (5)—but with "cov" and "sd" denoting the covariance (inner product) and standard deviation (norm), respectively. And indeed, just like the cosine function, $\rho_{X,Y} \in [-1, 1]$.

Compatibility is however defined for priors, posteriors, and likelihoods in $L_2(\Theta)$ equipped with the inner product (3), whereas Pearson correlation works with random variables in $L_2(\Omega, \mathbb{B}_\Omega, P)$ equipped with the inner product (12).

Fig. 3 sheds light on the different uses of compatibility and Pearson correlation. For example, $\kappa_{\pi,\ell}$ mea-

[4] Recall that a statistic $T = T(Y)$ is sufficient for θ if, $P(Y \in A \mid T = t)$ does not depend on θ, for all t in the range of T and for all sets A.

sures the agreement between likelihood and prior density, whereas $\rho_{X,Y}$ assesses the degree of linear association between random variables X and Y. The value $\kappa_{\pi,\ell} = 0.41$ is in line with the moderate overlap between prior and likelihood visible in Fig. 3. The value of $\rho_{X,Y} = 0.98$ is in line with the strong positive association between the random variables X and Y that can be seen in Fig. 3.

4 Closing remarks

This note offers a gentle introduction to geometrical aspects underlying the Bayesian paradigm that can be used for defining metrics of agreement between priors, likelihoods and posteriors as well as to rethink other concepts and results related with learning from data.

Geometrical interpretations are commonplace in Statistics and related fields—including for example that of Pearson correlation [15], least squares and LASSO (Least Absolute Shrinkage and Selection Operator) [10], and information geometry [1]; also, the geometry of multivariate analysis is well-known [13]. Many well-known geometrical insights concentrate on the *geometry of data* itself, whereas the focus of this note has been on the *geometry of learning from data*. Despite the long tradition of geometrical interpretations of statistical concepts, the view of the Bayesian paradigm along the lines of this note is relatively novel and it has been pioneered by [3] and [5].

Beyond geometry, topology and algebra hava also recently introduced a variety of insights and novel paradigms to the practice of learning from data—leading to the fields of topological data analysis [12] and algebraic statistics [4, 14].

Finally, we note that the geometrical view of the Donsker–Varadhan representation in (10) consists of a variational maximum inner product problem, and that nonvariational versions of such problems are of interest in the Machine Learning literature [8].

Acknowledgements

I am thankful to the Editor for the invitation. This note has benefited from the input and insight of a variety of discussants, including B. Barney, S. Beentjes, V. Inácio, G. Page, and V. Palacios.

References

[1] S. I. Amari, *Information Geometry and its Applications*, Springer, New York, 2016.

[2] W. Cheney, *Analysis for Applied Mathematics* Springer, New York, 2001

[3] M. de Carvalho, G. Page and B. Barney, On the geometry of Bayesian inference, *Bayesian Anal.* **14** (2019) 1013–1036.

[4] M. Drton, B. Sturmfel and S. Sullivant, *Lectures on Algebraic Statistics*, Springer, New York, 2009

[5] S. Kurtek and K. Bharath, Bayesian sensitivity analysis with the Fisher–Rao metric, *Biometrika* **102** (2015) 601–616.

[6] K. Knight, *Mathematical Statistics*, Chapman & Hall/CRC, Boca Raton, 2001.

[7] R. S. Millman and G. D. Parker, *Geometry: A Metric Approach with Models*, Springer, New York, 1991.

[8] S. Mussmann and S. Ermon, Learning and inference via maximum inner product search, *Proceedings of Machine Learning Research* **48** (2016) 2587–2596.

[9] B. J. Reich and S. K. Ghosh, *Bayesian Statistical Methods*, Chapman & Hall/CRC, Boca Raton, 2019.

[10] J. Taylor, The geometry of least squares in the 21st century, *Bernoulli* **19** (2013) 1449–1464.

[11] A. W. Van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge UK, 1998.

[12] L. Wasserman, Topological data analysis, *Annual Review of Statistics and Its Application* **5** (2018) 501–532.

[13] T. D. Wickens, *The Geometry of Multivariate Statistics*, Psychology Press, New York & London, 2014.

[14] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*, Cambridge University Press, Cambridge UK, 2009.

[15] D. Williams, *Probability with Martingales*, Cambridge University Press, Cambridge UK, 1991.

Some results on control theory for problems in fluid dynamics

by **Susana N. Gomes***

Abstract.—I will introduce some concepts in linear control theory, and how to adapt and use them to control some problems modelled by nonlinear partial differential equations appearing in a canonical fluid dynamics setting, showcasing how some simple results tailored for ODEs or linear PDEs can be explored to solve nonlinear complex problems from applications. This is a brief exposition of some of the results in [4] and [10].

Control theory is a branch of applied mathematics and systems engineering which considers dynamical systems, usually ordinary or partial differential equations (ODE/PDEs), and studies the development — or design — of algorithms whose goal is to drive these systems to a desired state while minimising any costs, delays, overshoots, or errors.

The mathematical theory of (feedback) control is outlined in [16, 23], where control of (linear and nonlinear) ordinary differential equations is considered, and where the authors introduce concepts such as controllability (any state can be reached by any starting point), stabilisability (it is possible to drive the system to have stable dynamics), and sufficient conditions for these to be possible. One can also introduce the Linear Quadratic Regulator (LQR), an example of an optimal control problem, where in addition to controlling the system, one also minimises a cost functional, usually penalising deviations from the desired state and the cost of the control. An optimal control problem is solved using the *Pontryagin maximum principle*, which is similar to the first-order optimality conditions (or Karush-Kuhn-Tucker, KKT, conditions) in traditional optimisation.

More recently, the theory of optimal control has been extended to problems modelled by PDEs [21], where one minimises a cost functional subject to the target solution solving a PDE. In this case, when applying the Pontryagin maximum principle, one needs to compute Fréchet derivatives of a Lagrangian, which involves several tools in functional analysis, and so proving existence of optimal controls is a harder task.

While the theory of feedback and optimal control has received extensive attention for systems governed by ODEs and (linear) PDEs, it was only recently that mathematicians started to target more complex systems, such as, for example, turbulence in fluid dynamics, and in this case, they often resort to the use of reduced-order models (ROM) which use techniques such as principal component analysis (PCA) to simplify the (infinite dimensional) state space into a finite dimensional and tractable vector or Hilbert space. However, in certain applications, we can obtain simplified models based on physical assumptions of the problem, and use these for control design. I will introduce an example in fluid dynamics, falling liquid films, that, by being comparatively simple to the full problem modelled by the Navier–Stokes equations, allows us to construct feedback controls that stabilise the full system with a lower computational cost and with no need for the use of ROMs.

In what follows, I will first introduce the basic concepts and results on feedback control needed to do this, followed by a short section describing the physical problem and the various models I consider. I will conclude with a survey of recent results on the control of falling liquid films, thus illustrating how sometimes one can obtain several useful (albeit numerical) results that can have an influence on practical applications, even when we cannot prove analytical results because we do not have the necessary assumptions on the problem (such as global well-posedness), and finish with some open problems.

1 A short introduction to control theory

In this section, I summarise the main results in (feedback) control theory which I will use later on. I will

* Mathematics Institute, University of Warwick. Email: Susana.gomes@warwick.ac.uk

start with an ODE example and illustrate how these results translate to PDEs.

Consider, for simplicity, the example a scalar ODE

$$\dot{y} = \lambda y, \qquad y(0) = y_0, \tag{1}$$

where the dot represents derivatives with respect to time. One can easily show that the solution of (1) is the function $y(t) = y_0 e^{\lambda t}$, and in particular, that $y(t) \to 0$ if $\lambda < 0$ and $y(t) \to \infty$ if λ is positive.

An intuitive thing to do to *stabilise* the system (i.e., to drive it towards the solution $y(t) = 0$), is to introduce a control (or forcing) term to equation (1), i.e., rewrite the controlled equation as

$$\dot{y} = \lambda y + f, \qquad y(0) = y_0. \tag{2}$$

We can then choose f in such a way that the solution is *stabilised*, and it is easy to see that it suffices to use a simple *proportional feedback control*: $f(t) = -\alpha y(t)$ for some positive constant α so that $\lambda - \alpha < 0$. In this case, we say that the control $f(t)$ stabilises the solution to the ODE. The term feedback is used because the control uses information on the current state of the system; the control is called *proportional* since it is proportional to the current solution.

Building up on this idea, we can consider the problem of controlling a *system* of ODEs, i.e., a problem of the form

$$\dot{\mathbf{y}} = A\mathbf{y} + \mathbf{f}, \qquad \mathbf{y}(0) = \mathbf{y_0}, \tag{3}$$

where now $\mathbf{y}$, $\mathbf{y_0}$, $\mathbf{f} \in \mathbb{R}^d$ and A is a $d \times d$ matrix. It can be shown that in this case, when $\mathbf{f} = \mathbf{0}$, if all *the eigenvalues* of A have negative real part, the solution is *asymptotically stable*, i.e., $\mathbf{y}(t) \to \mathbf{0}$ as $t \to \infty$. The analogue of the previous control here is to use $\mathbf{f}(t) = -\alpha\mathbf{y}(t) = -\alpha \boldsymbol{I}\mathbf{y}(t)$, where $\boldsymbol{I}$ is the $d \times d$ identity matrix. A simple calculation can be used to find the smallest α necessary to stabilise the system, namely, choose α such that the eigenvalues of $\boldsymbol{A} - \alpha\boldsymbol{I}$ all have negative real part.

While this is an easy thing to do, often in applications we can use information about the problem to obtain more efficient controls. Alternatively, it can be necessary to apply controls only to certain variables. This can be achieved by modifying the problem statement as follows:

$$\dot{\mathbf{y}} = A\mathbf{y} + B\mathbf{f}, \qquad \mathbf{y}(0) = \mathbf{y}_0. \tag{4}$$

Here, $\boldsymbol{B}$ is a $d \times \boldsymbol{M}$ matrix that encodes some information about how one applies the controls—for example, one can have $\boldsymbol{M}$ *control actuators* (where each column of $\boldsymbol{B}$ represents the effect of one control), or have different controls affect some rows of the system and not others. In this case, the controls are $\mathbf{f} \in \mathbb{R}^M$ (i.e. there are $\boldsymbol{M}$ of them). Note that we can have $\boldsymbol{M} = d$ and $\boldsymbol{B} = \boldsymbol{I}$, which is the case outlined above. It can be shown that under some assumptions on the matrices A and B (namely, the Kalman rank condition [23]), one can find a matrix $\boldsymbol{K}$ such that the eigenvalues of $\boldsymbol{A}+\boldsymbol{BK}$ all have negative real part, and therefore the controls $\mathbf{f} = \boldsymbol{K}\mathbf{y}$ stabilise the system. The matrix $\boldsymbol{K}$ can be computed using a pole placement algorithm [12] or by solving a linear-quadratic regulator problem [23].

In several complex systems relevant to applications, the interest is to control nonlinear dynamics, and we instead have a nonlinear system of ODEs,

$$\dot{\mathbf{y}} = \mathscr{N}(\mathbf{y}) + B\mathbf{f}, \qquad \mathbf{y}(0) = \mathbf{y}_0, \tag{5}$$

where $\mathscr{N}$ is some nonlinear function of y. Similar controllability or stabilisability results can be obtained (under assumptions on $\mathscr{N}$ such as Lipschitz continuity) by considering a linearisation of the nonlinear operator and using Lyapunov function type arguments [23].

Finally, for several applications there is interest in controlling (linear or nonlinear) *partial differential equations* (PDEs); for example a reaction-diffusion equation for the evolution of a population, tumour growth or other biological and chemical applications. Such PDEs take the general form

$$u_t = \mathscr{L}u + \mathscr{N}(u) + f, \tag{6}$$

along with appropriate initial and boundary conditions. The subscript t denotes time derivative, and $\mathscr{L}$, $\mathscr{N}$ are linear and nonlinear spatial differential operators, respectively. By projecting this equation to an appropriate basis (e.g., taking Fourier transforms), one can write the PDE as an *infinite-dimensional* system of ODEs such as (5). Alternatively, one can also discretise the problem (e.g. using finite differences) to rewrite it as a finite dimensional system of equations. This approach is commonly known as "discretise then optimise". Passing to the PDE limit is not straightforward, even for linear PDEs [23]. However, in certain cases, this is possible; this is done for several linear PDEs (see [21]), and I will show a particular case of a nonlinear PDE in the next section.

2 Falling liquid films and how to control them

I will now introduce the problem of a falling liquid film, which is a canonical setting in fluid dynamics

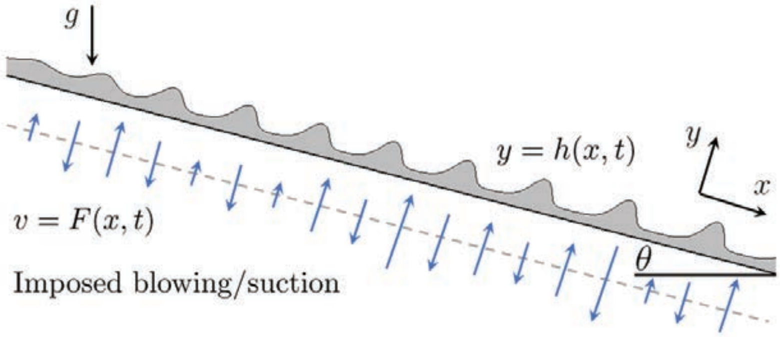


Figure 1.—Diagram of a thin film flowing down an inclined plane and allowing for blowing and suction controls. The dynamics of the interface $y=h(x,t)$ are controlled by some fluid parameters, as well as the inclination angle θ and the imposed control values $v=F(x,t)$ at $y=0$ in the coordinate system shown in the figure.

with applications such as coating of LCD screens or manufacturing of microchips.

2.1 A hierarchy of models for falling liquid films

Falling liquid films are thin films of a viscous fluid flowing down an inclined plane, as shown in Figure 1. This problem has been studied extensively both theoretically (accurate model development, see, for example, [3, 11, 15]) and experimentally ([6]) and provides a set of models which is amenable to control development. The goal here is to control the interface towards a desired shape; for example, while the uncontrolled system evolves towards a travelling wave such as the one depicted in Figure 1, or more complex, and even chaotic, solutions, in applications such as LCD screen coating one would want the interface to be flat, whereas for microchip cooling we would desire a wavy interface with a suitable profile, to enhance heat transfer. For the models I will show, the flat solution will correspond to $h(x,t) = 1$ unless otherwise stated. To control the resulting interface, we will allow for fluid to be inserted or removed from the system via slots at the plate that the film is flowing over, as depicted in Figure 1, and this will appear as a boundary condition, or as a coefficient in the different models we will consider. We will see how to design the controls, i.e., how to prescribe how much fluid is inserted or removed from the system at each slot, as well as how many of these controls we need, using variations of the feedback control theory outlined in the previous section.

This physical problem is modelled by the (two-dimensional) Navier–Stokes equations;[1] in particular by modelling the interaction between the fluid and the air via the interface at $y = h(x,t)$. After an appropriate non-dimensionalisation, the system parameters are reduced to two non-dimensional groupings: the Reynolds number Re measuring the relative importance between inertia and viscosity, and the capillary number Ca which measures the importance of surface tension. The Navier–Stokes equations consist of the momentum equations for u, v, and p the stramwise (parallel to the plane) and transverse (perpendicular to the plane) velocities, and pressure, respectively.

$$Re(u_t + uu_x + vu_y) = -p_x + 2 + u_{xx} + u_{yy}, \quad (7)$$

$$Re(v_t + uv_x + vv_y) = -p_y - 2\cot\theta + v_{xx} + v_{yy}, \quad (8)$$

which are coupled to the continuity equation given by

$$u_x + v_y = 0. \quad (9)$$

In addition, the system is completed by its boundary conditions. We consider periodic boundaries in the x-direction[2], no-slip and fluid injection/removal at the wall,

$$u = 0, \qquad v = F(x,t), \quad (10)$$

the nonlinear dynamic stress balance (or momentum

[1] It is possible to generalise the problem to three dimensions, but this is much more computationally expensive, and for the purposes of this problem, a 2D description is often enough.

[2] This is a modelling assumption, which simplifies the analytical computations that follow. If the domain is sufficiently long, this is a good enough approximation, but different approaches can consider different boundary conditions.

jump) at the interface, $y = h(x,t)$,

$$(v_x + u_y)(1 - h_x^2) + 2h_x(v_y - u_x) = 0, \tag{11}$$

$$\begin{aligned} p - \frac{2(v_y + u_x h_x^2 - h_x(v_x + u_y))}{1 + h_x^2} &= \\ = -\frac{1}{Ca}\frac{h_{xx}}{(1 + h_x^2)^{3/2}}, \end{aligned} \tag{12}$$

and finally the kinematic boundary condition

$$h_t = v - u h_x. \tag{13}$$

The uncontrolled system admits a uniform flat film solution known as the Nusselt solution [11], given by $h(x,t) = 1$ and a semi-parabolic in y horizontal fluid velocity, which can be used to obtain simplified models.

It is well-known that full models such as the Navier–Stokes equations are computationally expensive to simulate, and therefore if one wants to solve it for several values of the relevant parameters (or, for example, perform optimal control using these models), it becomes prohibitively expensive. However, in the case of thin liquid films, the mean interface height is much smaller than the length of the domain, L, and this makes it possible to define a *long wave* parameter $\epsilon = 1/L \ll 1$. This disparity of scales facilitates a multiscale approach to derive from first principles hierarchies of simplified models. [3] To be able to derive these models, we need the following assumptions:

(A1) (long-wave assumption) the geometrical aspect ratio ϵ is small;

(A2) The Reynolds number Re is $\mathcal{O}(1)$;

(A3) Surface tension is sufficiently strong to appear at leading order, i.e., the capillary number is small, and $Ca = \mathcal{O}(\epsilon^2)$ is the appropriate distinguished limit;

(A4) The controls F are small $F = \mathcal{O}(\epsilon)$, implying weak injection or removal of fluid via the control actuators.

Using assumptions **(A1)**-**(A4)** and asymptotic analysis techniques, Thompson *et al.* [20] derived two different long-wave models for falling liquid films using this type of control (long-wave models for uncontrolled falling liquid films were explored earlier in the literature, see [11]). Both models satisfy a mass conservation equation

$$h_t + q_x = F(x,t), \tag{14}$$

which is coupled with an equation for the flux $q(x,t) = \int_0^h u(x,y,t)\,\mathrm{d}y$. In the first model, the *Benney equation*, they obtain an explicit expression for $q(x,t)$ and the model is a single PDE for the interfacial height $h(x,t)$:

$$\begin{aligned} q(x,t) = \frac{h^3}{3}\left(2 - 2h_x \cot\theta + \frac{h_{xxx}}{Ca}\right) + \\ + Re\left(\frac{8h^6 h_x}{15} - \frac{2h^4 F}{3}\right). \end{aligned} \tag{15}$$

The second model is the *weighted residuals model*, which describes the evolution of the interfacial height $h(x,t)$ and the flux $q(x,t)$:

$$\begin{aligned} \frac{2Re}{5}h^2 q_t + q = \frac{h^3}{3}\left(2 - 2h_x \cot\theta + \frac{h_{xxx}}{Ca}\right) + \\ + Re\left(\frac{18q^2 h_x}{35} - \frac{34hqq_x}{35} + \frac{hqF}{5}\right). \end{aligned} \tag{16}$$

We note that the controls appear as an inhomogeneous term $F(x,t)$ in the mass conservation equation (14), and this structure plays a crucial role in the efficiency of these controls.

Due to the asymptotic reduction, these models only provide us with the interface height $h(x,t)$ and downstream flux $q(x,t)$ and do not directly provide the solution to the Navier–Stokes equations (i.e. u, v, and p). However, if needed, these can be recovered from h and q, thus allowing for comparison with direct numerical simulations of the Navier–Stokes equations when necessary.

The above long-wave models are significantly more accessible computationally than the full Navier–Stokes equations, but they are still highly nonlinear. This means that it is hard (if not impossible) to treat them analytically, and to the best of my knowledge there are no analytical results beyond linear stability analysis of the flat solution (and some results on solitary waves for some special cases) [11]. Because of this, there is some interest in applying further simplifications in order to make analytical progress. For very small but nonlinear perturbations of the flat solution, one can perform weakly nonlinear analysis to derive a Kuramoto-Sivashinsky (KS) equation [11, 19]. The KS equation is a fourth-order nonlinear PDE having the same form as (6), and is given by

$$\eta_t + \nu\eta_{xxxx} + \eta_{xx} + \eta\eta_x = f(x,t),$$

where η is a small perturbation of a flat interface and $\nu > 0$ is a parameter that encodes some of the

[3] We often call these reduced-order models, but I will not use this terminology, to avoid confusion with ROMs obtained via, e.g., PCA.

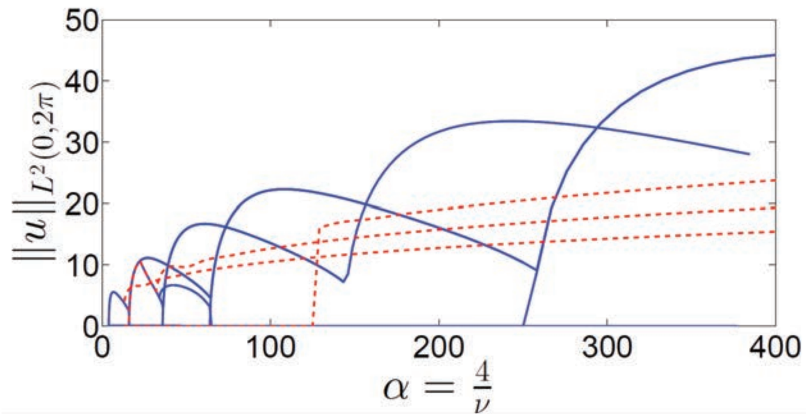


Figure 2.—Bifurcation diagram of the solutions of the KS equation. Full blue lines correspond to steady state solutions while dashed red lines are travelling waves. Not all branches are included, and most solutions depicted here are unstable.

geometry of the problem[4]. This problem is posed with periodic boundary conditions and we have $x \in [0, 2\pi]$. In this case, a flat interface corresponds to $\eta = 0$.

The KS equation appears in several applications and is widely studied since it is one of the simplest model PDEs exhibiting spatiotemporal chaotic behaviour. Over the last few decades, existence and uniqueness of solutions have been explored [17], different types of attractors have been characterised [5], and the route to chaos for solutions of the KS equation have been reported [13], to show a small subset of the range of interesting analytical and computational results that can be achieved even at this lowest member of the model hierarchy. It is possible to compare the results from these models to direct numerical simulations of the Navier–Stokes equations, and some relevant comparisons can be seen in [6]. While the long-wave models provide a very good approximation of the full system, in most cases the KS equation solution differs significantly from it (see Figure 2 in [4]). However, its simplicity and existing analytical results have allowed us to develop efficient controls (see [2, 8]) which were then extended to controlling long-wave models [19] and eventually the full model [4, 10]. For the rest of this article, I will summarise our results in this direction.

2.2 Feedback control of falling liquid films

I will start outlining our results towards control of falling liquid films by showing the (analytical and numerical) results on controlling the KS equation. As mentioned above, while there is significant model error when considering this PDE to model interfaces of falling liquid films, the analytical insights can provide us with enough information to motivate control development on the more complicated long wave models, and eventually design controls that drive the solution to the full system towards a desired state.

The controlled KS equation, rewritten so that controls reflect a finite number of control actuators that inject and remove fluid through slots is given by

$$\eta_t + \nu\eta_{xxxx} + \eta_{xx} + \eta\eta_x = \sum_{j=1}^{M} \delta(x - x_j) f_j(t). \quad (17)$$

For the uncontrolled problem, it is easy to check that if $\nu < 1$, the zero solution is linearly unstable. Without the nonlinear term $\eta\eta_x$, the solution would grow exponentially in time; however, the nonlinearity promotes exchange of energy between Fourier modes and instead we see a "zoo" of solutions, from steady states, to travelling waves, but more generally we observe chaotic behaviour. This can be seen in Figure 2, where we plot the bifurcation diagram of possible solutions of the KS equation, with steady states depicted in full blue lines, and travelling waves by red

[4] Most of the geometry of the problem, however, is encoded in the change of variables used to arrive at this equation; in particular, the solutions of this equation sit on a moving frame, and so even "steady-state" solutions correspond to travelling waves of the original problem.

dashed lines. The y axis plots the L^2 norm of different solutions (here u should be replaced by η). The figure is taken from [8].

Armaou and Christofides showed in [2] that the zero solution of the KS equation in small domains (ν close to 1) can be controlled using $M = 5$ control actuators. More recently, we were able to show that we can stabilise any unstable solution (any of the branches depicted in Figure 2) of the KS equation using as many control actuators as unstable modes in the system (see [8, 9]).

To show this, it is useful to consider a discretisation of the KS equation. Let any solution be written as

$$\eta_t = \eta_0(t) + \sum_{k=0}^{\infty} \eta_k^s(t)\sin(kx) + \eta_k^c(t)\cos(kx).$$

We can then write the KS equation as an infinite system of ODEs for the coefficients η_k^* (where $*$ stands for c or s). Defining $\boldsymbol{\eta} = (\eta_0, \eta_1^s, \eta_1^c, \dots)$, this system is written as:

$$\dot{\boldsymbol{\eta}} = \mathscr{A}\boldsymbol{\eta} + \mathscr{N}(\boldsymbol{\eta}) + \boldsymbol{B}\boldsymbol{F},$$

where $\mathscr{A}$ is a diagonal matrix whose entries are $-\nu k^4 + k^2$, $\mathscr{N}$ is given by a convolution, $\boldsymbol{B}$ includes the discretisation of the control actuators ($\boldsymbol{B}_{kj} = \int_0^{2\pi} \delta(x - x_j)\sin(kx)\,dx$, equivalently for the coefficient corresponding to $\cos(kx)$), and $\boldsymbol{F}$ encodes the contol action.

Proposition 1.— Let $\bar{\eta}$ be a linearly unstable steady state or travelling wave solution of the KS equation (17) and let $2\ell + 1$ be the number of unstable eigenvalues of the operator $\mathscr{A}$, *i.e.*, $\ell + 1 \geq 1/\sqrt{\nu} > \ell$. Additionally, let A_u be the $M \times M$ submatrix consisting of coefficients corresponding to unstable modes, and define $\boldsymbol{B}_u$ similarly. If $M = 2\ell + 1$, then there exists a matrix $\boldsymbol{K} \in \mathbb{R}^{M\times M}$ such that all of the eigenvalues of the matrix $A_u + \boldsymbol{B}_u\boldsymbol{K}$ have negative real part, and the state feedback controls $\boldsymbol{F} = \boldsymbol{K}(\boldsymbol{\eta} - \bar{\boldsymbol{\eta}})$ stabilise $\bar{\eta}$.

Proof.— I will only sketch the proof of this result; for more details see [7]. First, consider the problem of controlling the system of M ODEs

$$\dot{\mathbf{y}} = A_u\mathbf{y} + \boldsymbol{B}_u\boldsymbol{F}.$$

If each control actuator has a different location (i.e. $x_i \neq x_j$), then it is easy to show that the columns of $\boldsymbol{B}_u$ are linearly independent, and therefore it can be shown that the matrices A_u and $\boldsymbol{B}_u$ satisfy the Kalman rank condition and the system is controllable. Therefore, we can guarantee that there exists a matrix $\boldsymbol{K}$ such that $A_u + \boldsymbol{B}_u\boldsymbol{K}$ has negative eigenvalues. We can then use an algorithm such as pole placement [12] to find $\boldsymbol{K}$ – in particular, we will choose $\boldsymbol{K}$ such that all eigenvalues of $A_u + \boldsymbol{B}_u\boldsymbol{K}$ have real part smaller than $-\inf|\eta_x|/2$.

Now we define the perturbation $v = \bar{\eta} - \eta$ and write a PDE for v:

$$v_t + \nu v_{xxxx} + v_{xx} + v v_x + (\bar{\eta}v)_x = \sum_{j=0}^{M} \delta(x - x_j) f_j(t).$$

Multiplying this equation by v and integrating by parts, we obtain, formally,

$$\frac{1}{2}\frac{d\|v\|^2}{dt} = \int_0^{2\pi} v(\mathscr{A} + \boldsymbol{B}\boldsymbol{K})v\,dx + \\ + \int_0^{2\pi} v^2 v_x + v(\bar{\eta}v)_x\,dx.$$

The integral of $v^2 v_x$ vanishes due to periodic boundary conditions. Furthermore, we can show that the term $\int v(\eta v)_x\,dx$ is bounded by $\inf|\eta_x|\|v\|^2/2$, and therefore it can be shown from the choice of eigenvalues that the right-hand side is bounded by $-\lambda\|v\|^2$ where λ is the largest eigenvalue of $A_u + \boldsymbol{B}_u\boldsymbol{K}$, showing that $\|v\|^2$ is a Lyapunov function for this system, and therefore $v = 0$ is a stable solution, meaning $\eta = \bar{\eta}$ is stabilised using the controls $\boldsymbol{F} = \boldsymbol{B}\boldsymbol{K}v = \boldsymbol{B}\boldsymbol{K}(\eta - \bar{\eta})$.

It can also be shown (see [8]) that the controls are robust to uncertainty in the problem parameters, as well as to small changes in the number of controls used. For an example of a controlled solution see Figure 3.

Motivated by the similar linear stability properties between the KS equation and the Benney equation (the simplest long-wave model), we studied the control problem for two long-wave models: the Benney equation and the weighted residual model in Thompson *et al.* [19]. We started by showing that in the unrealistic scenario where one can observe the whole interface and *actuate everywhere*, the simplest proportional controls of the form

$$f(x,t) = -\alpha(h(x,t) - 1), \tag{18}$$

for some constant $\alpha > 0$ to be determined, efficiently drive the system towards the flat solution $h(x,t) = 1$ (or indeed any desired solution $H(x,t)$, by replacing 1 by $H(x,t)$). The critical value

$$\alpha_c = \frac{16Ca(Re - \frac{5}{4}\cot\theta)}{75}$$

can be computed from linear stability analysis of the Benney equation or the weighted residuals model,

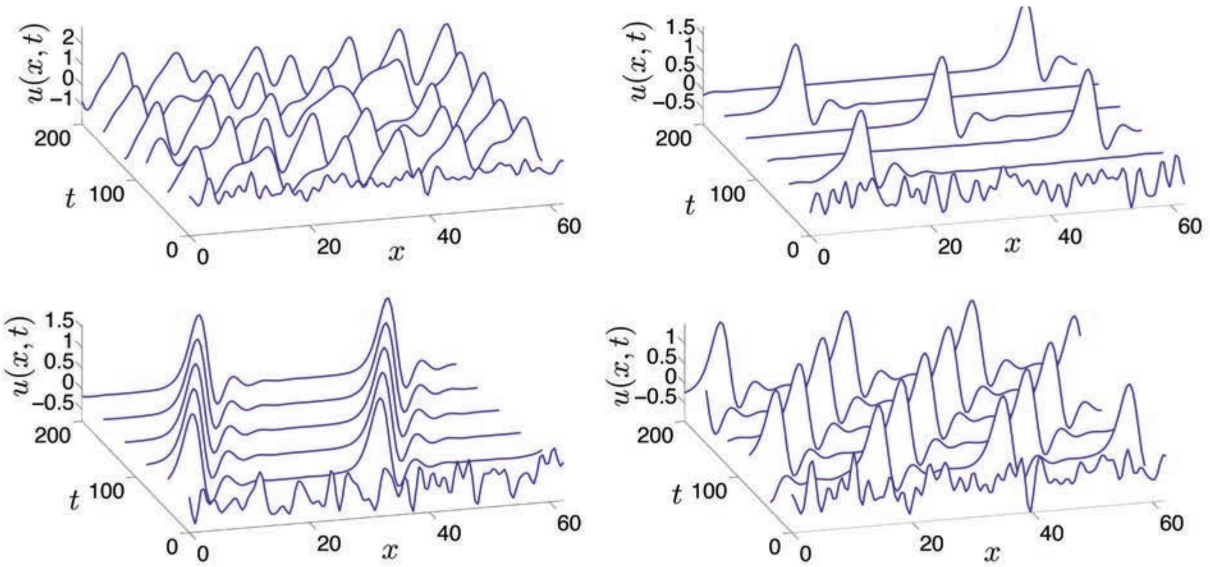


Figure 3.—Control of the KS equation for $v = 0.01$. Uncontrolled solution showing chaotic behaviour (top left), and controlled solution towards: a 1-pulse travelling wave (top right), a 2-pulse travelling wave (bottom left), and a 3-pulse travelling wave (bottom right). We used $M = 21$ equidistant controls.

and it depends only on the Reynolds and capillary numbers. Using linear stability analysis, we can also calculate the number of unstable modes (see [10]) to be

$$M = 1 + 2\ell = \\ = 1 + 2\left\lfloor \frac{L}{2\pi}\sqrt{Ca\left(\frac{8}{5}Re - 2\cot\theta\right)}\right\rfloor . \qquad (19)$$

It is also shown in [19] that the critical α for the Benney equation is sufficient to obtain linear stability of the weighted residuals model and indeed the full Navier–Stokes equations, by solving an Orr–Sommerfeld system. As mentioned before, in this case, because of the nonlinearities of the system, we cannot prove that linear stability of the controlled solutions guarantees that the solution of the long wave models or the Navier–Stokes equations will indeed be stabilised. However, we can confirm nonlinear stability of these solutions by numerical simulations of the initial value problem.

Similarly to the KS equation, one can compute point actuated controls assuming we can observe the whole interface (using pole placement or solving an LQR problem), and unsurprisingly controls of this type also stabilise the flat solution. A more interesting (and realistic) case is when we not only actuate at a finite number of locations, but can also only observe the interface at a finite number of points. In this case, in [19] we use proportional feedback controls of the form

$$f(x,t) = -\alpha \sum_{j=1}^{M} \delta(x - x_j)(h(x_j - \phi, t) - 1), \quad (20)$$

where $\delta(\cdot)$ is the Dirac delta function, the control actuators are located at the positions x_j, $j = 1, \dots, M$, and observations of the interface are made at $x = x_j - \phi$ for some displacement ϕ. Figure 4 shows predictions of whether these controls stabilise the nonlinear dynamics for $L = 64$, $\theta = \pi/3$, $Re \approx 15$ and $Ca \approx 0.001$ (3 unstable modes) using $M = 3, 5, 7$, or 9 and $P = M$ observers with a displacement ϕ from the corresponding actuator. We observe that positive ϕ, i.e. observations upstream of actuation, are beneficial; this makes sense intuitively, since if we observe upstream, we can predict where the wave will be by the time the control effects reach it.

Again, linear stability does not guarantee the solution of the nonlinear equation will be stabilised, but for most cases, we can confirm numerically that this is the case. I will show examples of this when applied to the full model (the Navier–Stokes equations) in the next section.

2.3 Applying the controls to the full model

Now that we have efficient controls that stabilise the KS equation and the long wave models, we are ready to apply these to the full Navier-Stokes equations. As mentioned previously, the full system is quite complex, and hard to simulate. To test the controls, we perform direct numerical simulations (DNS) of the Navier-Stokes equations using the open-source software Gerris [14] and its extension Basilisk, which solve the Navier-Stokes equations on an adaptive quadtree grid using a volume-of-fluid approach.

The control strategies developed in the previous

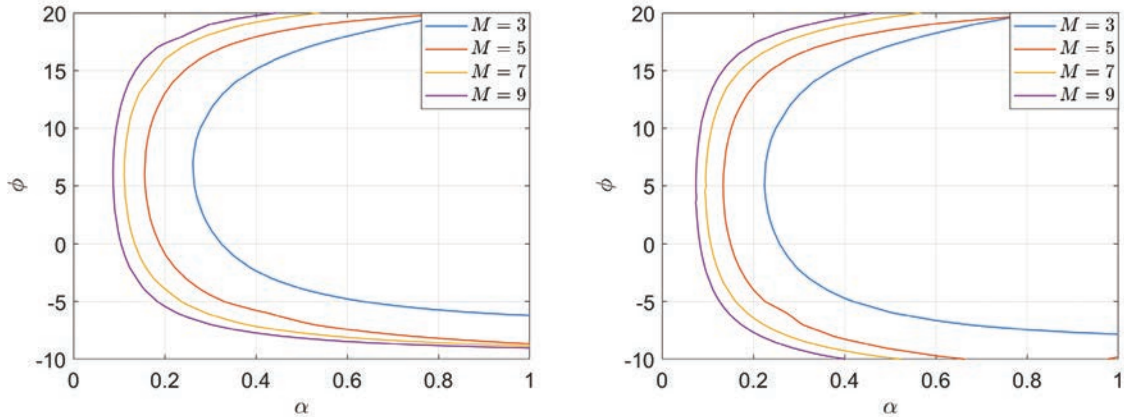


Figure 4.—Regions of stability of controls of the form (20) for several number of controls M and displacement ϕ. Left: Benney equation, and right: weighted residuals model. Inside each curve, we predict the controls to stabilise the flat solution, while outside we predict them to not be sufficient for stability.

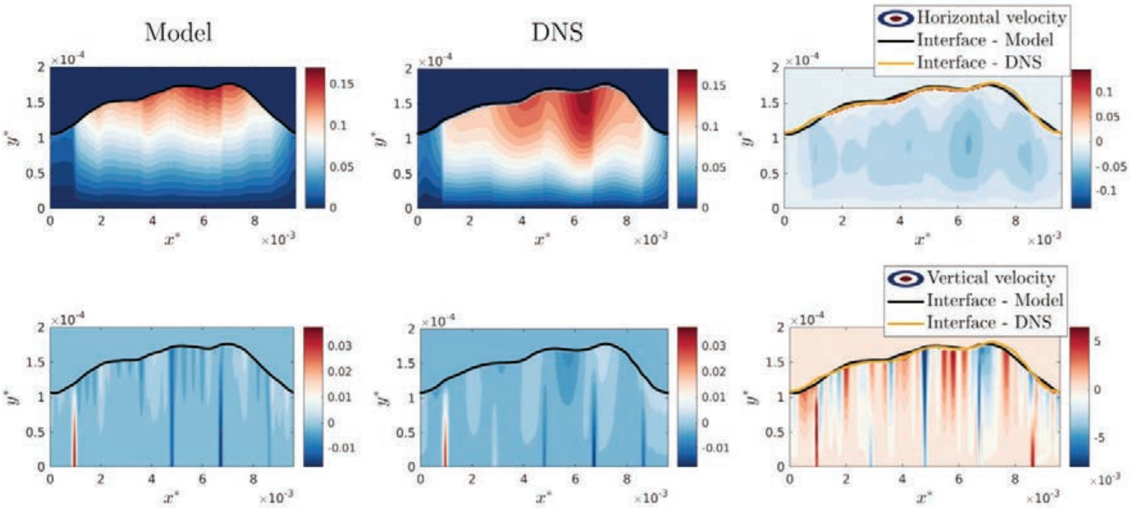


Figure 5.—Comparison between the solution, (u,v), of the Navier–Stokes equations using approximations obtained from the weighted residuals model (left) and DNS (middle), and the difference between the two (right) for the horizontal velocity u (top) and the vertical velocity v (bottom), immediately after application of controls. For details on the parameters used, see [4].

section are efficient in stabilising the flat solution for the Benney equation and the weighted residuals model, and linear stability analysis predicts they also (linearly) stabilise the full problem. Naively, we could try to use them directly in the Navier-Stokes equations; however, we observe that we cannot simply "translate" the controls directly to the full problem, i.e., simply take the numerical value from the simplified models and apply it to the Navier-Stokes equations: while they seem to work on the first few time steps, after a while the differences between the full problem and the simplified models become too big, the controls stop working, and the solution eventually returns to the original uncontrolled state. This is to be expected, since there are physical effects that appear at the DNS level which are not fully resolved in the weighted residuals model because of the physical assumptions we made to derive the models.

To illustrate this, we show a comparison in Figure 5 between the solution, (u, v), of the Navier-Stokes equations using approximations obtained from the model (left) and DNS (middle), as well as the difference between the two (right) for the horizontal velocity u (top) and the vertical velocity v (bottom), immediately after application

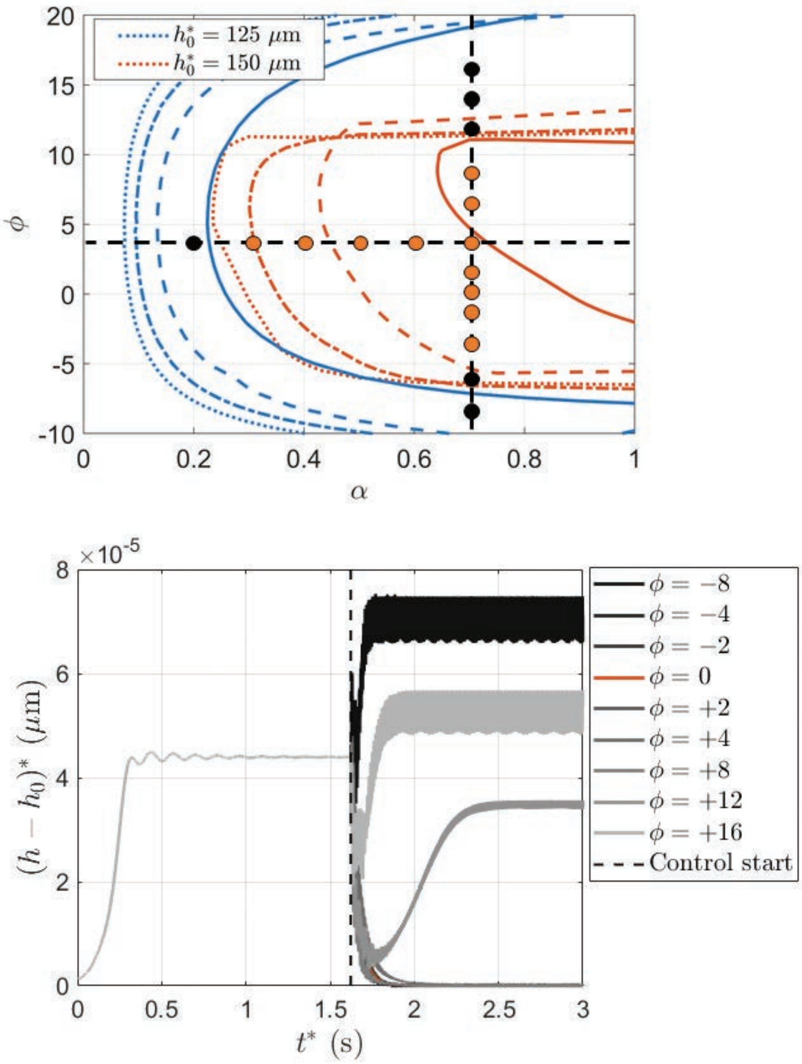


Figure 6.—Stability predictions (top) and direct numerical simulations (bottom) for the controlled solution of the Navier–Stokes equations for several values of α.

of controls. We can see that even though the error on the horizontal velocity is small, there are significant differences in the vertical velocity.

However, we can use the linear stability analysis predictions (such as the control strategy in (20) with the predictions for α visible in Figure 4) and apply these controls based on observations of the numerical solution obtained via the direct numerical simulations.

We tested this methodology for several cases, with $L = 64$, $\theta = \pi/3$ and including $Re \approx 10$, $Ca \approx 0.001$ (case 1), and $Re \approx 17$, $Ca \approx 0.001$ (case 2), with the results shown in Figure 6, where we show the stability predictions for case 1 in blue, and case 2 in orange (top figure). The different curves correspond to a different number of control actuators and observers varying from $M = 3$ (full lines) to $M = 9$ (dotted lines). In the bottom figure, we pick case 2 and $M = 5$ controls, and apply controls with varying ϕ and fixed α. Each curve corresponds to a dot on the vertical line in the top figure, where orange dots signify a stabilised solution, while black dots correspond to a failed control. We see that the direct numerical simulations confirm that controls predicted to linearly stabilise the weighted residuals model do indeed stabilise the full problem. We also performed similar tests for fixed ϕ and varying α with similar results.

Our final result concerns applying the controls derived for the long wave models to direct numerical simulations of the Navier-Stokes equations. As above, we use the control "rule" derived from linear stability analysis (in this case, solving an LQR Problem for the weighted residuals model), but where we use observations of the DNS solution. Motivated by the success of the controls in Figure 6, we expect the same philosophy to be applicable. We tested a range of Reynolds

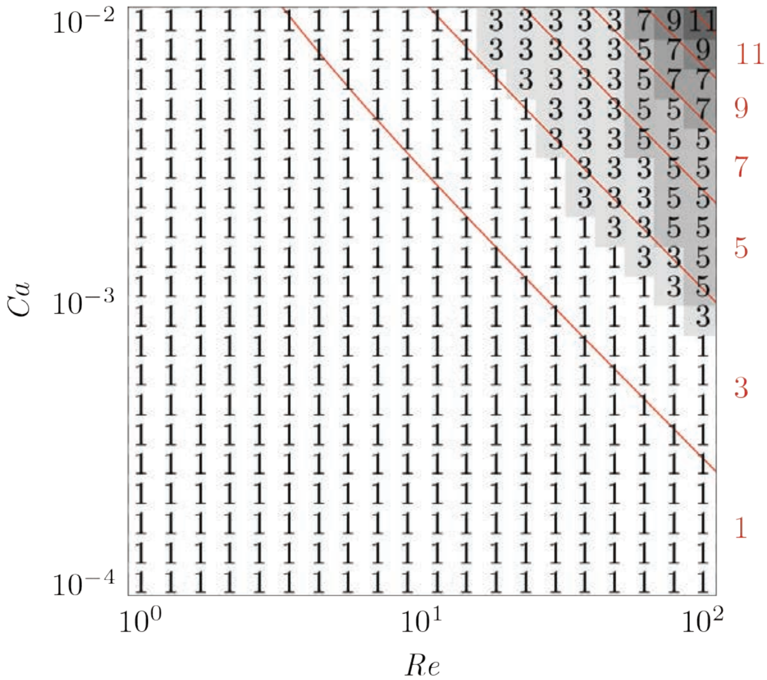


Figure 7.—The minimum number of actuators required to stabilise the Navier-Stokes film compared to the number of unstable modes of the linearised weighted- residual system (red). The number of controls needed to stabilise the uniform film never exceeds the number of unstable modes of the linear system M as given in (19). The ranges for the two parameters cover a broad range of different fluids. For more details, see [10].

numbers Re and capillary numbers Ca which correspond to several physically motivated fluids (see [10] for more details). For each case, we predicted the number of unstable modes — and therefore the number of necessary controls — using (19) and computed the matrix K from the weighted residuals model. We then applied the controls to the full model using observations from the DNS. The results are summarised in Figure 7: the red lines show the predicted number of controls, and the numbers in each square show how many controls were needed to stabilise the flat solution. We observe that in almost every case, we did not need as many controls as linear stability suggests, thus showing the efficiency of the controls we designed.

3 Discussion

I presented a control methodology based on a hierarchy of models, which I used to control a canonical problem in fluid dynamics: falling liquid films. This is a complex problem for which control is hard due to the computational and analytical complexity of the models involved. Using a hierarchy of models allowed me to start from a weakly nonlinear model (the Kuramoto-Sivashinsky equation), where it is possible to derive controls analytically that stabilise the flat solution, and any other unstable solution.

While the KS equation is not a very good approximation to the original problem, the results at this level provide crucial information to guide us in the right direction for controlling the more accurate long wave models (Benney equation and weighted residuals model), and eventually the Navier–Stokes equations.

The results I presented are based on linear feedback control theory, and can be thought of as a "discretise-then-optimise" framework. Other approaches can be used; for example, we can first optimise and then discretise, as seen, e.g., in [1], or we can use optimal control methodologies (see [22]). We can also use other forms of control such as electric fields [22] or temperature [18].

This illustrates how simple mathematical models are key players in mathematical studies and help us push conceptual boundaries to the point where the developed methodologies can be applied higher up in the model hierarchy. Often in control theory, several problems are not explored enough due to their nonlinearity, which makes analytical progress impossible, and I hope this example shows the value of mathematical modelling and numerical simulation working together with control theory to advance our understanding of complex phenomena in fluid dynamics.

References

[1] R. Al Jamal and K. Morris. Linearized stability of partial differential equations with application to stabilization of the kuramoto–sivashinsky equation. *SIAM Journal on Control and Optimization*, 56(1):120–147, 2018.

[2] A. Armaou and P. D. Christofides. Feedback control of the kuramoto–sivashinsky equation. *Phys. D*, 137(1-2):49–61, 2000.

[3] D. J. Benney. Long waves on liquid films. *J. Math. Phys.*, 45(1-4):150–155, 1966.

[4] R. Cimpeanu, S. N. Gomes, and D. T. Papageorgiou. Active control of liquid film flows: beyond reduced-order models. *Nonlinear Dyn.*, 104(1):267–287, 2021.

[5] P. Collet, J.-P. Eckmann, H. Epstein, and J. Stubbe. A global attracting set for the kuramoto-sivashinsky equation. 1993.

[6] F. Denner, A. Charogiannis, M. Pradas, C. N. Markides, B. G. M. van Wachem, and S. Kalliadasis. Solitary waves on falling liquid films in the inertia-dominated regime. *J. Fluid Mech.*, 837:491–519, 2018.

[7] S. N. Gomes, S. Kalliadasis, D. T. Papageorgiou, G. A. Pavliotis, and M. Pradas. Controlling roughening processes in the stochastic kuramoto–sivashinsky equation. *Phys. D*, 348:33–43, 2017.

[8] S. N. Gomes, D. T. Papageorgiou, and G. A. Pavliotis. Stabilizing non-trivial solutions of the generalized kuramoto–sivashinsky equation using feedback and optimal control. *IMA J. Appl. Math.*, 82(1):158–194, 2017.

[9] S. N. Gomes, M. Pradas, S. Kalliadasis, D.T. Papageorgiou, and G. A. Pavliotis. Controlling spatiotemporal chaos in active dissipative-dispersive nonlinear systems. *Phys. Rev. E*, 92(2):022912, 2015.

[10] O. A. Holroyd, R. Cimpeanu, and S. N. Gomes. Linear quadratic regulation control for falling liquid films. *arXiv preprint arXiv:2301.11379*, 2023.

[11] S. Kalliadasis, C. Ruyer-Quil, B. Scheid, and M. G. Velarde. *Falling liquid films*, volume 176. Springer Science & Business Media, 2011.

[12] J. Kautsky, N. K. Nichols, and P. Van Dooren. Robust pole assignment in linear state feedback. *International Journal of control*, 41(5):1129–1155, 1985.

[13] D. T. Papageorgiou and Y. S. Smyrlis. The route to chaos for the kuramoto-sivashinsky equation. *Theoretical and Computational Fluid Dynamics*, 3(1):15–42, 1991.

[14] S. Popinet. Gerris: a tree-based adaptive solver for the incompressible euler equations in complex geometries. *J. Comput. Phys.*, 190(2):572–600, 2003.

[15] G. I. Sivashinsky and D. M. Michelson. On irregular wavy flow of a liquid film down a vertical plane. *Prog. Theor. Phys.*, 63(6):2112–2114, 1980.

[16] E. D. Sontag. *Mathematical control theory: deterministic finite dimensional systems*, volume 6. Springer Science & Business Media, 2013.

[17] E. Tadmor. The well-posedness of the kuramoto–sivashinsky equation. *SIAM Journal on Mathematical analysis*, 17(4):884–893, 1986.

[18] A. B. Thompson, S. N. Gomes, F. Denner, M. C. Dallaston, and S. Kalliadasis. Robust low-dimensional modelling of falling liquid films subject to variable wall heating. *Journal of Fluid Mechanics*, 877:844–881, 2019.

[19] A. B. Thompson, S. N. Gomes, G. A. Pavliotis, and D. T. Papageorgiou. Stabilising falling liquid film flows using feedback control. *Phys. Fluids*, 28(1):012107, 2016.

[20] A. B. Thompson, D. Tseluiko, and D. T. Papageorgiou. Falling liquid films with blowing and suction. *J. Fluid Mech.*, 787:292–330, 2016.

[21] F. Tröltzsch. *Optimal control of partial differential equations: theory, methods, and applications*, volume 112. American Mathematical Soc., 2010.

[22] Alexander W. Wray, Radu Cimpeanu, and Susana N. Gomes. Electrostatic control of the navier-stokes equations for thin films. *Phys. Rev. Fluids*, 7:L122001, 2022.

[23] J. Zabczyk. *Mathematical control theory: an introduction*. Birkhäuser, 1992.

Atlantic Conference in Nonlinear PDEs Dispersive and Elliptic Equations and Systems

by Simão Correia*, Hugo Tavares**, James Kennedy*** and Gianmaria Verzini****

The event *Atlantic Conference in Nonlinear PDEs: Dispersive and Elliptic Equations and Systems* took place at Instituto Superior Técnico, Universidade de Lisboa from October 30 to November 3, 2023, having 74 participants. This conference was sponsored by the Portuguese government via FCT – Fundação para a Ciência e Tecnologia, I.P, through the research centers CAMGSD (grant UID/MAT/04459/2020), CEMAPRE and GFM (UID/00208/2020) and through the project NoDES (PTDC/MAT-PUR/1788/2020). It was also supported by CIM (Portugal), FLAD – Fundação Luso Americana para o Desenvolvimento and Dipartimento de Matematica - Politecnico di Milano (Italy). The event was organized by Hugo Tavares and Simão Correia (IST), James Kennedy (FCUL) and Gianmaria Verzini (Politecnico di Milano), with the help of the local assistants Gabriel Moraes, Francisco Agostinho and Pêdra Andrade (IST).

The structure of the event comprise plenary talks, contributed talks (divided in two parallel session, one on Dispersive and the other on Elliptic PDEs) and poster sessions, bringing together the distinct but parallel communities working and/or interested in Elliptic and Dispersive Partial Differential Equations. It joined experts and early career mathematicians from both sides in an atmosphere that fostered an exchange of ideas and forged new collaborations and perspectives.

The *plenary lectures* were delivered by **Thomas Bartsch** (Universität Gießen), **Jean-Baptiste Casteras** (Universidade de Lisboa), **Mónica Clapp** (Universidad Nacional Autónoma de México), **Raphaël Côte** (Université de Strasbourg), **Luca Fanelli** (Ikerbasque and Universidad del País Vasco, BCAM), **Stefan Le Coz** (Université de Toulouse), **Felipe Linares** (IMPA), **Yvan Martel** (Université de Versailles), **Ademir Pastor** (Universidade Estadual

* CAMGSD, Instituto Superior Técnico. Email: simao.f.correia@tecnico.ulisboa.pt
** CAMGSD, Instituto Superior Técnico. Email: hugo.n.tavares@tecnico.ulisboa.pt
*** GFM, Faculdade de Ciências da Universidade de Lisboa. Email: jbkennedy@ciencias.ulisboa.pt
**** Politecnico di Milano. Email: gianmaria.verzini@polimi.it

de Campinas), **Dmitry Pelinovsky** (McMaster University), **Svetlana Roudenko** (Florida International University), **Enrico Serra** (Politecnico di Torino), **Jorge Silva** (Universidade de Lisboa), **Didier Smets** (Sorbonne Université), **Susanna Terracini** (Universitá de Torino), **Luis Vega** (Basque Center for Applied Mathematics), **Zhi-Qiang Wang** (Utah State University), **Tobias Weth** (Goethe-Universität Frankfurt), while the *contributed talks* were given by: **Pêdra Andrade** (Instituto Superior Técnico – Universidade de Lisboa) **Lukas Bengel** (Karlsruhe Institute of Technology), **Filippo Boni** (Università degli Studi di Napoli "Federico II"), **William Borrelli** (Politecnico di Milano), **Luccas Campos** (UFMG – Universidade Federal de Minas Gerais), **Andreia Chapouto** (UCLA), **Simone Dovetta** (Politecnico di Torino), **Amin Esfahani** (Nazarbayev University), **Francesco Esposito** (University of Calabria), **Luiz Gustavo Farah** (Universidade Federal de Minas Gerais), **Filippo Giuliani** (Politecnico di Milano), **Julia Henninger** (Karlsruhe Institue of Technology), **Marco Morandotti** (Politecnico di Torino), **Giuseppe Negro** (Instituto Superior Técnico, Universidade de Lisboa), **Matteo Rizzi** (Justus Liebig University), **Makson Santos** (Instituto Superior Técnico, Universidade de Lisboa), **Delia Schiera** (Instituto Superior Técnico, Universidade de Lisboa), **Jacopo Schino** (North Carolina State University), **Panayotis Smyrnelis** (University of Athens), **Frédéric Valet** (University of Bergen), **Jianjun Zhang** (Chongqing Jiaotong University). *Posters* were presented by **Francisco Agostinho** (Instituto Superior Técnico, Universidade de Lisboa), **Laura Baldelli** (Polish Academy of Sciences), **Nicolò Cangiotti** (Politecnico di Milano), **Mariem Dhifet** (University of Monastir), **Umberto Guarnotta** (University of Enna "Kore"), **Zhengni Hu** (Universität Gießen), **Rahma Jlel** (University of Monastir), **Gabriel Moraes** (Universidade de Maringá) and **Sebastian Ohrem** (Karlsruhe Institute of Technology).

For more information, see the website
https://sites.google.com/view/atlanticpdes/

ENUMATH 2023

by Adélia Sequeira* and Ana Silvestre**

The ENUMATH 2023 Conference was held at the Instituto Superior Técnico (IST), in Lisbon, Portugal. It was the 14th of a series of conferences that started in Paris (1995), followed by Heidelberg (1997), Jyväkylä (1999), Ischia Porto (2001), Prague (2003), Santiago de Compostela (2005), Graz (2007), Uppsala (2009), Leicester (2011), Lausanne (2013), Ankara (2015), Bergen (2017) and Egmond aan Zee (2019).

The central goal of the Local Organizing Committee, composed by **Adélia Sequeira** (Chair), **Ana Silvestre** (Co-Chair) and **Jorge Tiago**, from IST and CEMAT, University of Lisbon, **Telma Guerra**, from IPSetúbal and CEMAT, **João Janela**, ISEG and CEMAPRE, University of Lisbon, **Marília Pires**, CIMA, University of Évora, and **Svilen S. Valtchev**, IPLeiria and CEMAT, was to fulfill the objectives of the ENUMATH conferences, namely to provide a forum for presenting and discussing novel and fundamental advances in numerical mathematics and challenging scientific and industrial applications on the highest level of international expertise.

The Scientific Program of ENUMATH 2023 included

* Dep. of Mathematics and CEMAT, Instituto Superior Técnico, Universidade de Lisboa.
Email: adelia.sequeira@tecnico.ulisboa.pt

** Dep. of Mathematics and CEMAT, Instituto Superior Técnico, Universidade de Lisboa.
Email: ana.silvestre@math.tecnico.ulisboa.pt

Plenary talks:

Habib Ammari
ETH, Zurich, Switzerland
From condensed matter theory to sub-wavelength physics

Paola Francesca Antonietti
MOX, Politecnico di Milano, Italy
Mathematical and numerical modeling of neurodegenerative diseases

Peter Bastian
University of Heidelberg, Germany
Multithreaded multilevel spectral domain decomposition

Mária Lukácová-Medvidová
University of Mainz, Germany
What is a limit of numerical methods for compressible flows?

Jean-Marie Mirebeau
University of Paris-Sud, France
Discretization of anisotropic PDEs using Voronoi's reduction of positive quadratic forms

Daniel Peterseim
University of Augsburg, Germany
Numerical solution of nonlinear eigenvector problems

José A. Carrillo de la Plata
University of Oxford, UK
Primal Dual methods for Wasserstein gradient flow

Carola-Bibiane Schönlieb
University of Cambridge, UK
From differential equations to deep learning for image analysis

Luís Oliveira e Silva
IST, University of Lisbon, Portugal
Challenges in numerical modeling of extreme plasma physics in the laboratory and in astrophysics

Alessandro Veneziani
University of Emory, USA
The role of applied mathematics in the design of coronary stents

Sara Zahedi
KTH, Royal Institute of Technology, Sweden
Conservative cut finite element methods

Minisymposia (some of them with several sessions),

Contributed talks and

Poster presentations.

The winner of the Best Poster Award (sponsored by CIM) was **Charlotte Milano**, from Reims Mathematical Laboratory (LMR), University of Reims Champagne Ardenne, France, with a poster entitled *Numerical Methods for electromagnetic cartography in medical imaging*. Two honorable mentions were awarded to **Lisa Grandjean**, also from the University of Reims Champagne Ardenne, and **Alessio Fumagalli**, from Politecnico di Milano, Italy.

Overall, ENUMATH 2023 was an inspiring meeting, both for scientific interactions and informal discussions, involving leading experts and young scientists from 31 different countries, with special emphasis on contributions from Europe.

As in the previous editions of the conference, the participants were invited to submit a short paper for the ENUMATH 2023 Proceedings to be published by Springer, as a volume of the series Lecture Notes in Computational Science and Engineering.

3rd Women in Mathematics Meeting

by Sílvia Barbeiro* and Susana Moura**

The third edition of the conference Women in Mathematics Meeting (WM23) was held at the University of Coimbra, from the 24th to the 26th of July 2023.

The motivation of this series of conferences is to promote the role of women in Mathematics and to develop a supportive community as well as motivate and inspire new generations of women mathematicians through high quality scientific presentations.

The WM23 meeting brought together 68 participants, coming from various countries, including several experts in their fields of research as well as early-stage post-docs and graduate students. Besides the scientific talks and the posters session, the program included two special sessions addressing gender gap in Mathematics and the role of the Mathematical Societies in promoting gender parity. The conference was also an opportunity to showcase the exhibition Women of Mathematics from around the world—a gallery of portraits.

The attendance of the event was free of charge, and the organization acknowledges the financial support from the following institutions: Centre for Mathematics of the University of Coimbra (CMUC), Centre of Mathematics of the University of Minho (CMAT), International Center for Mathematics (CIM), Center for Mathematical Analysis, Geometry and Dynamical Systems (CAMGSD).

Further details can be found at

https://ucpages.uc.pt/events/wm23/

* CMUC, Department of Mathematics, University of Coimbra. Email: silvia@mat.uc.pt
** CMUC, Department of Mathematics, University of Coimbra. Email: susana.moura@uc.pt

Organizing Committee

Ana Jacinta Soares
CMAT, Univ. Minho
Marina Ferreira
CMUC, Univ. Coimbra
Milene Santos
CMUC, Univ. Coimbra
Sílvia Barbeiro
CMUC, Univ. Coimbra
Susana Moura
CMUC, Univ. Coimbra

Scientific Committee

Joana Nunes da Costa
CMUC, Univ. Coimbra
Margarida Mendes Lopes
CAMGSD, Univ. Lisbon
Sofia Castro
CMUP, Univ. Porto

Invited Speakers

Ana Cristina Ferreira
CMAT, Univ. Minho
Ana Paula Dias
CMUP, Univ. Porto
Diana Rodelo
CMUC & Univ. Algarve
Eloísa Grifo
Univ. Nebraska
Fátima Silva Leite
ISR, Univ. Coimbra
Magda Rebelo
CMA, Nova Univ. Lisbon

Special Session I — *Gender gap in Mathematical, Computing and Natural Sciences*

Colette Guillopé
LAMA, CNRS et Université Paris-Est Créteil

Special Session II — *The role of the Mathematical Societies in promoting gender parity*

Moderator
Ana Cristina Casimiro
Nova University of Lisbon

Panel members
Alessandra Bernardi
University of Trento — Secretary of the Italian Mathematical Union
Eva A. Gallardo Gutiérrez
Complutense University of Madrid — President of the Royal Spanish Mathematical Society
José Carlos Santos
University of Porto — President of the Portuguese Mathematical Society

Exhibition — *Women of Mathematics from around the world — a gallery of portraits*

PEDRO NUNES LECTURES

RICHARD A. DAVIS

Richard Davis received his Ph.D. degree in Mathematics from the University of California at San Diego in 1979 and has held academic positions at MIT, Colorado State University. He was Hans Fischer Senior Fellow at the Technical University of Munich (2009-12), Villum Kan Rasmussen Visiting Professor at the University of Copenhagen (2011-13), and Chalmers Jubilee Professor at Chalmers University of Technology. His research interests include time series, applied probability, extreme value theory, heavy-tailed modeling with applications to network models, and spatial-temporal modeling. Richard Davis is a fellow of the Institute of Mathematical Statistics and the American Statistical Association, and is an elected member of the International Statistical Institute.

USING SAMPLE SPLITTING FOR ASSESSING GOODNESS OF FIT IN TIME SERIES

DECEMBER, 13 — 2023

UNIVERSIDADE DE AVEIRO EDIFÍCIO CENTRAL E DA REITORIA
15:00 — SALA DOS ATOS ACADÉMICOS

STATISTICAL LEARNING OF MULTIVARIATE EXTREMES

DECEMBER, 15 — 2023

ACADEMIA DAS CIÊNCIAS DE LISBOA
15:00 — SALA DAS SESSÕES

CIM CENTRO INTERNACIONAL DE MATEMÁTICA

SPM SOCIEDADE PORTUGUESA DE MATEMÁTICA