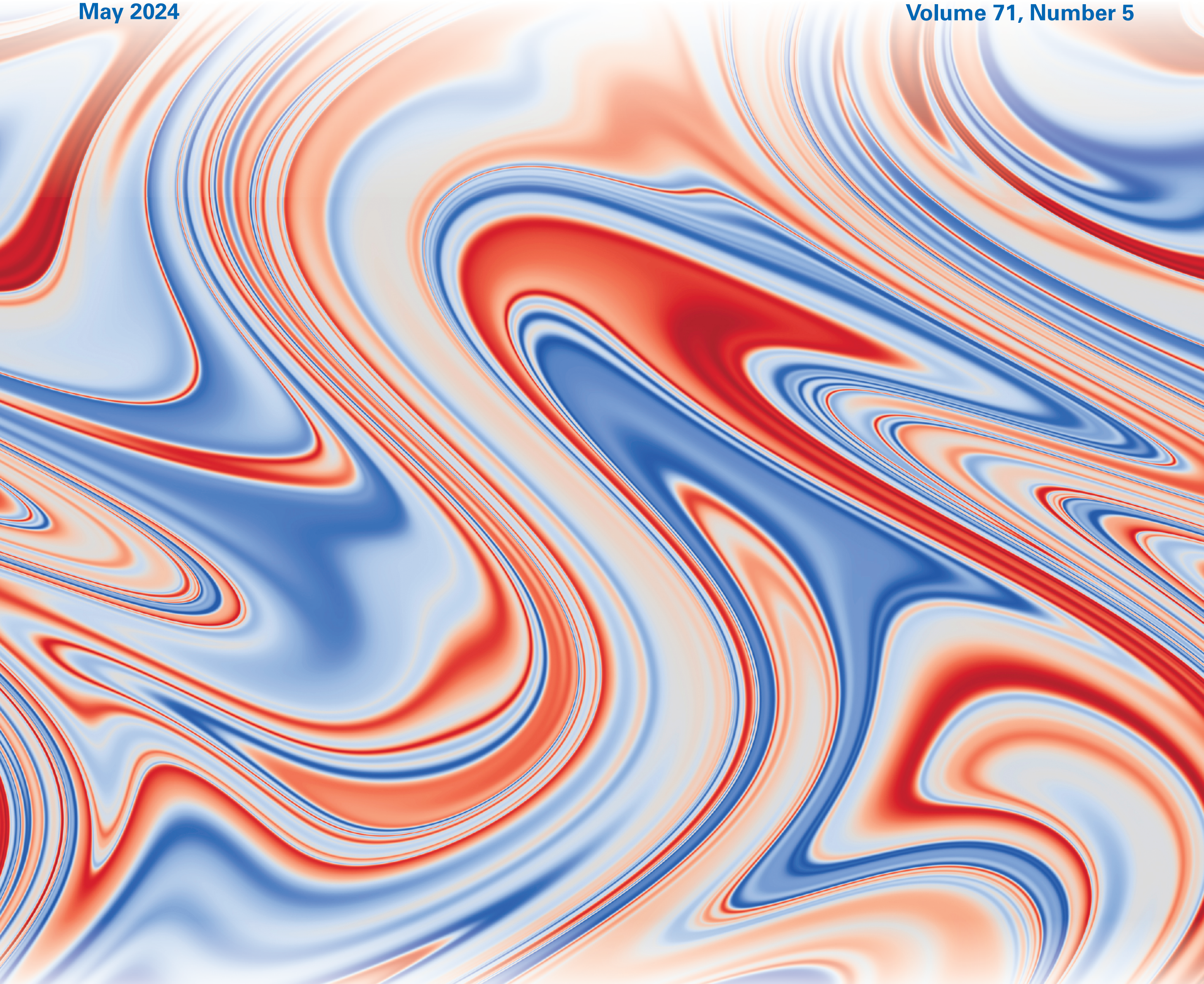
ISSN 0002-9920 (print)
ISSN 1088-9477 (online)

# Notices

## of the American Mathematical Society

May 2024 Volume 71, Number 5



AMS American Mathematical Society

Advancing research. Creating connections.

*The cover design is based on imagery from "Mixing in Incompressible Flows: Transport, Dissipation, and Their Interplay," page 593.*

Notices of the American Mathematical Society

May 2024

*Cover Credit: The image used in the cover design appears in "Mixing in Incompressible Flows: Transport, Dissipation, and Their Interplay," p. 593, and is courtesy of Anna L. Mazzucato, Michele Coti Zelati, Gianluca Crippa, and Gautam Iyer.*



## FROM THE AMS SECRETARY



## INVITATIONS FROM THE AMS



ERRATUM. On page 519 in the April 2024 issue, Ernst Chiandi should be Ernst Chladni.

# Notices
of the American Mathematical Society




### SUBSCRIPTION INFORMATION
Individual subscription prices for Volume 71 (2024) are as follows: nonmember, US$764, member, US$611.20. (The subscription price for members is included in the annual dues.) For information on institutional pricing, please visit https://www.ams.org/publications/journals/subscriberinfo. Subscription renewals are subject to late fees. Add US$6.50 for delivery within the United States; US$25 for surface delivery outside the United States. See www.ams.org/journal-faq for more journal subscription information.

### ADVERTISING
*Notices* publishes situations wanted and classified advertising, and display advertising for publishers and academic or scientific organizations. Advertising requests, materials, and/or questions should be sent to:
**classads@ams.org (classified ads)**
**notices-ads@ams.org (display ads)**

### PERMISSIONS
All requests to reprint *Notices* articles should be sent to:
**reprint-permission@ams.org**.

### SUBMISSIONS
The editor-in-chief should be contacted about articles for consideration after potential authors have reviewed the "For Authors" page at **www.ams.org/noticesauthors**.

The managing editor should be contacted for additions to our news sections and for any questions or corrections. Contact the managing editor at: **notices@ams.org**.

Letters to the editor should be sent to: **notices-letters@lists.ams.org**.

To make suggestions for additions to other sections, and for full contact information, see **www.ams.org/noticescontact**.

Supported by the AMS membership, this publication is freely available electronically through the AMS website, the Society's resource for delivering electronic products and services. Use the URL **www.ams.org/notices** to access the *Notices* on the website. The online version of the *Notices* is the version of record, so it may occasionally differ slightly from the print version.

The print version is a privilege of Membership. Graduate students at member institutions can opt to receive the print magazine by updating their individual member profiles at **www.ams.org/member-directory**. For questions regarding updating your profile, please call 800-321-4267.

For back issues see **www.ams.org/backvols**. Note: Single issues of the *Notices* are not available after one calendar year.

The American Mathematical Society is committed to promoting and facilitating equity, diversity and inclusion throughout the mathematical sciences. For its own long-term prosperity as well as that of the public at large, our discipline must connect with and appropriately incorporate all sectors of society. We reaffirm the pledge in the AMS Mission Statement to "advance the status of the profession of mathematics, encouraging and facilitating full participation of all individuals," and urge all members to conduct their professional activities with this goal in mind. (*as adopted by the April 2019 Council*)


[Notices of the American Mathematical Society (ISSN 0002-9920) is published monthly except bimonthly in June/July by the American Mathematical Society at 201 Charles Street, Providence, RI 02904-2213 USA, GST No. 12189 2046 RT****. Periodicals postage paid at Providence, RI, and additional mailing offices. POSTMASTER: Send address change notices to Notices of the American Mathematical Society, PO Box 6248, Providence, RI 02904-6248 USA.] Publication here of the Society's street address and the other bracketed information is a technical requirement of the US Postal Service.





Printed in the United States of America. The paper used in this journal is acid-free and falls within the guidelines established to ensure permanence and durability.


Opinions expressed in signed *Notices* articles are those of the authors and do not necessarily reflect opinions of the editors or policies of the American Mathematical Society.

# On Hamilton Cycles in Graphs Defined by Intersecting Set Systems

*Torsten Mütze*

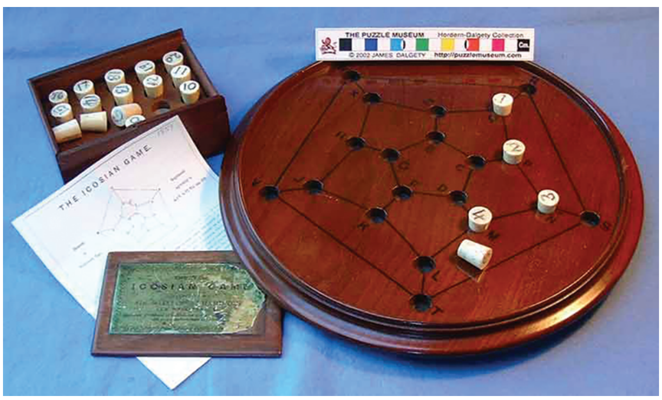



**Figure 1.** Hamilton's Icosian game.

## 1. Introduction

In 1857 the Irish mathematician William Rowan Hamilton invented a puzzle whose goal is to find a cycle in the graph of the dodecahedron that visits every vertex exactly once. He dubbed it the 'Icosian game', as the resulting cycle has exactly twenty ('icosa' in ancient Greek) edges and vertices. In honor of Hamilton, a cycle that visits every vertex of a graph exactly once is now called a *Hamilton cycle*. The dodecahedron has the interesting property that it looks the same from the point of view of any vertex. Formally, it is *vertex-transitive*, i.e., any two vertices can be mapped onto each other by an automorphism of the graph. In 1970 Lovász raised a conjecture which can be considered a highly advanced version of the Icosian game. Specifically, he conjectured that every connected vertex-transitive graph admits a Hamilton cycle, apart from five exceptional graphs, which are a single edge, the Petersen graph, the Coxeter graph, and the graphs obtained from the latter two by replacing every vertex by a triangle.

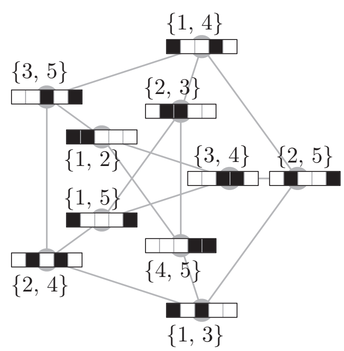



**Figure 2.** The Petersen graph as intersection graph of all 2-element subsets of $\{1,2,3,4,5\}$, with edges connecting disjoint sets. In the corresponding bitstring representations, 0s are drawn as white squares and 1s as black squares.

The Petersen graph, shown in Figure 2, is a 10-vertex graph that serves as a famous example and counterexample for many problems in graph theory, and we will encounter it again later in this article. A variant of Lovász' conjecture asserts that every connected vertex-transitive graph admits a *Hamilton path* (without exceptions), i.e., we still need to visit every vertex exactly once, but the first and last vertex of the tour need not be adjacent. While the Icosian game is about one particular graph, Lovász' conjecture talks about infinitely many graphs, which on the one hand are very constrained, but on the other hand very hard to get your hands on. One can easily construct numerous explicit families of vertex-transitive graphs for which it is not known whether they are Hamiltonian. An important example for this are *Cayley graphs*, which are defined for a group and a set of generators as the graph that has as vertices all group elements, and whose edges arise by multiplication with a generator. Even for particular groups like the symmetric group, and for generators with certain properties (like two generators, or generators that are involutions), only partial results are known.

In this expository article, we consider another rich family of vertex-transitive graphs, which features prominently

*Torsten Mütze is an assistant professor in the Department of Computer Science at the University of Warwick, United Kingdom, and in the Department of Theoretical Computer Science and Mathematical Logic at Charles University, Prague, Czech Republic. His email address is* torsten.mutze@warwick.ac.uk.

*This work was supported by Czech Science Foundation grant GA 22-15272S.*

*Communicated by* Notices *Associate Editor Emilie Purvine.*

*For permission to reprint this article, please contact:* reprint-permission@ams.org.

DOI: https://doi.org/10.1090/noti2907

throughout combinatorics, namely graphs defined by intersecting set systems. We give an overview of the many beautiful techniques and ingredients devised during the past 40 years that establish the existence of Hamilton cycles in those graphs, thus settling interesting special cases of Lovász' conjecture. Our discussion starts with the easy instances and ends with the hardest and most general ones, and it emphasizes the key obstacles on this journey.

**1.1. The hypercube.** The starting point is the Boolean lattice, i.e., the inclusion order on all subsets of $[n] := \{1, 2, \ldots, n\}$; see Figure 3. The cover relations of this poset are pairs of sets $X, Y$ which differ in a single element, i.e., $Y = X \cup \{i\}$ for some $i \in [n]$, and the corresponding cover graph $Q_n$ is the well-known *n-dimensional hypercube*. It is convenient to encode the vertices of $Q_n$ as bitstrings of length $n$, by considering the characteristic vector of each set, which has the $i$th bit equal to 1 if and only if the element $i$ is contained in the set. With this encoding, edges of $Q_n$ connect exactly pairs of vertices that differ in a single bit. This graph is vertex-transitive, and a Hamilton cycle can be found easily, by applying the following simple greedy rule discovered by Williams: Start at an arbitrary vertex, and repeatedly flip the rightmost bit that creates a previously unvisited vertex. The resulting cycle is known as *binary reflected Gray code*, named after Bell Labs researcher Frank Gray, and it is shown in Figure 3 for $n = 4$.

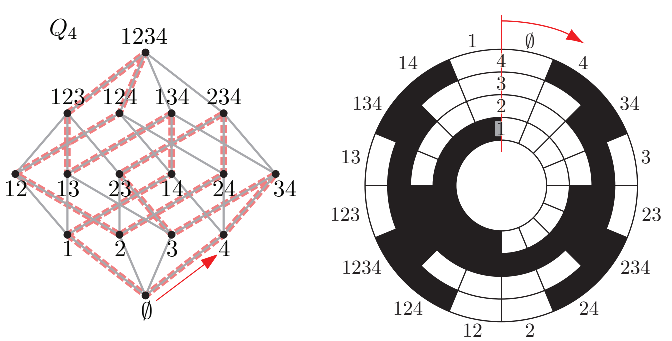



**Figure 3.** Left: The 4-dimensional hypercube $Q_4$ and one of its Hamilton cycles, the binary reflected Gray code (highlighted edges); Right: bitstring representation of the cycle (0 = white and 1 = black) with vertices in clockwise order starting at 12 o'clock. When printing the sets, curly brackets and commas are omitted for simplicity.

## 2. The Middle Levels Conjecture

The *kth level* of $Q_n$ is the set of all vertices with exactly $k$ many 1s. In terms of subsets in the Boolean lattice, these are the subsets of size exactly $k$. Now consider the hypercube of odd dimension $2k + 1$, and the subgraph induced by the middle two levels $k$ and $k + 1$; see Figure 4. We denote this subgraph by $M_k$, and we note that it is vertex-transitive. Havel [Hav83] in 1983, and independently Buck and Wiedemann [BW84] in 1984 conjectured that $M_k$ has a Hamilton cycle for all $k \geq 1$, and this problem became known as *middle levels conjecture*. It turns out that this problem is considerably harder than finding a Hamilton cycle in the entire cube. Thus, the conjecture is a prime example of an easy-to-state combinatorial proposition that one feels should be easy to prove at first sight, but that turns out to be surprisingly intricate upon further investigation. Also, it is an explicit instance of Lovász' conjecture, and failing to prove it for this particular family of graphs $M_k$ shows how little we understand about the general problem. Consequently, the middle levels conjecture has attracted a lot of attention in the literature (see e.g., [KT88, FT95, SW95], and it is mentioned in several popular books, in particular in Diaconis and Graham's book *Magical Mathematics*, and in Winkler's book *Mathematical Puzzles: A Connoisseur's Collection*. Furthermore, Knuth gave the middle levels conjecture the highest difficulty rating (49/50) among all open problems in his book *The Art of Computer Programming*, Volume 4A. In a recent survey, Gowers comments on the conjecture as follows:

> If one starts trying to build a Hamilton cycle in $M_k$, one runs into the problem of having too much choice and no obvious way of making it. (A natural thing to try to do is find some sort of inductive construction, but a lot of people have tried very hard to do this, with no success—a natural pattern just doesn't seem to emerge after the first few small cases.)

**2.1. Kierstead-Trotter matchings.** A *matching* in a graph is a set of pairwise disjoint edges, and a matching is *perfect* if it includes every vertex. Kierstead and Trotter [KT88] in 1988 suggested to tackle the middle levels conjecture by taking the union of two edge-disjoint perfect matchings in $M_k$, with the hope that their union forms the desired Hamilton cycle.

The two matchings they considered can be described explicitly as follows: We write $A_k$ and $B_k$ for the vertices in levels $k$ and $k + 1$ of $Q_{2k+1}$, respectively, i.e., these are the two partition classes of $M_k$. A *Dyck word* is a bitstring with the same number of 1s and 0s and the property that every prefix contains at least as many 1s as 0s. We write $D_k$ for the set of all Dyck words of length $2k$. Furthermore, for any string $x$ and any integer $i$, we write $\sigma^i(x)$ for the cyclic left shift of $x$ by $i$ steps. Note that every vertex $x \in A_k$ can be written uniquely as $x = \sigma^i(y0)$ for some $y \in D_k$ and $0 \leq i < 2k + 1$. Indeed, the counting works out correctly, as the number of Dyck words is the Catalan number $C_k = |D_k| = \frac{1}{k+1}\binom{2k}{k}$, the number of cyclic shifts is $2k + 1$, and $(2k + 1) \cdot C_k = \binom{2k+1}{k} = |A_k|$. We then define $f(x) := \sigma^i(y1) \in B_k$. Furthermore, we decompose $y$ uniquely as $y = 1u0v$ for $u, v \in \bigcup_{j \geq 0} D_j$, and we define $g(x) := \sigma^i(1u1v0) \in B_k$. It is easy to show

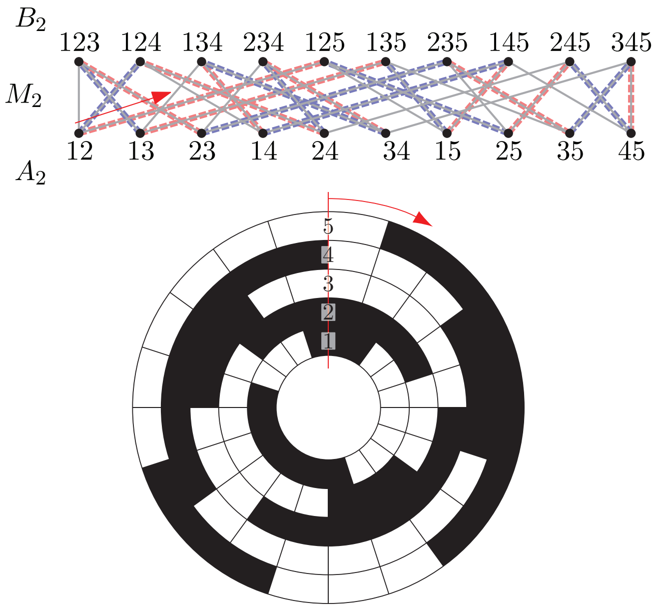



**Figure 4.** Top: The middle levels graph $M_2$ and the perfect matchings $X_f$ and $X_g$ (red and blue, respectively) whose union is a Hamilton cycle; Bottom: bitstring representation of the cycle.

that $f$ and $g$ are bijections between $A_k$ and $B_k$, and therefore $X_f := \{(x, f(x)) \mid x \in A_k\}$ and $X_g := \{(x, g(x)) \mid x \in A_k\}$ are two edge-disjoint perfect matchings between the two partition classes $A_k$ and $B_k$ of $M_k$. The union of $X_f$ and $X_g$ thus forms a *cycle factor* $F_k := X_f \cup X_g$ in the graph $M_k$, i.e., a collection of disjoint cycles that together visit all vertices. In fact, for $k = 2$ we are lucky and $F_k$ is a single cycle, i.e., a Hamilton cycle in $M_k$; see Figure 4. Unfortunately, our luck ends for larger values of $k$, as the number of cycles of $F_k$ for $k = 1, 2, \ldots$ is $1, 1, 2, 3, 6, 14, 34, 95, 280, 854, \ldots$.

**2.2. Interpretation of the cycles as plane trees.** The middle levels conjecture was finally solved by Mütze [Müt16] in 2016, who later provided a 2-page proof [Müt23a]. The short proof picks up on Kierstead and Trotter's argument as follows: The counting sequence for the number of cycles of $F_k$ is OEIS sequence A002995, which counts plane trees with $k$ edges, hinting at a bijection between plane trees with $k$ edges and cycles in $F_k$; see Figure 8.


**Figure 5.** Bijection between Dyck words (left), Dyck paths (middle) and ordered rooted trees (right). Our trees have the root (red) at the bottom and grow upward.

Indeed, this bijection follows from the definition of $f$ and $g$, and it also uses the following well-known bijection between Dyck paths with $2k$ steps and ordered rooted trees with $k$ edges; see Figure 5: Given a Dyck word $x$, we consider the corresponding Dyck path, which has an $\nearrow$-step for every 1-bit and a $\searrow$-step for every 0-bit of $x$. We then squeeze this Dyck path together, gluing every pair of an $\nearrow$-step and $\searrow$-step on the same height that 'see' each other (i.e., that have no other step at the same height in between them) together to form an edge of the ordered rooted tree.

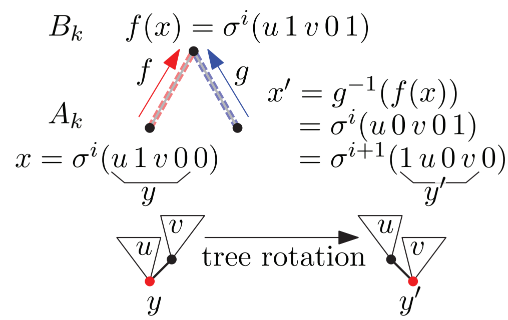



**Figure 6.** Two steps along a cycle of $F_k$ correspond to a tree rotation and cyclic left shift.

For a vertex $x \in A_k$, we consider the vertex $x' \in A_k$ that is two steps away along a cycle of $F_k$, i.e., $x'$ is obtained by traversing one edge of $X_f$ and one edge of $X_g$, so $x' = g^{-1}(f(x))$. Specifically, for $x = \sigma^i(u1v00) \in A_k$ with $u, v \in \bigcup_{j \geq 0} D_j$ and $0 \leq i < 2k+1$ we have $x' := g^{-1}(f(x)) = \sigma^i(u0v01) = \sigma^{i+1}(1u0v0)$. We now define $y := u1v0 \in D_k$ and $y' := 1u0v \in D_k$, and we consider the ordered rooted trees corresponding to the Dyck words $y$ and $y'$, which differ in a tree rotation. Consequently, walking two steps along a cycle of $F_k$ corresponds to a tree rotation and a cyclic left shift of the bitstring. As $2k+1$ (the number of shifts) and $k$ (the number of tree edges) are coprime, all cyclic shifts of every tree lie on the same cycle. As the equivalence classes of ordered rooted trees under tree rotation are precisely plane trees, obtained by 'forgetting' the root, we obtain the desired bijection. Now that we have a nice combinatorial interpretation of the cycles in the factor $F_k$, all that is left to do is to join the cycles to a single Hamilton cycle.

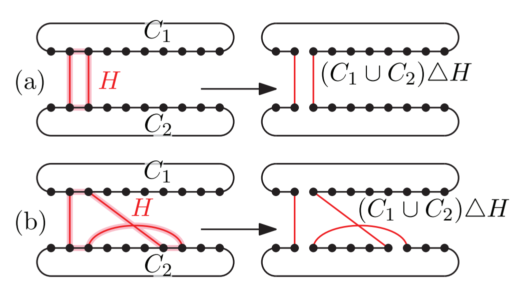



**Figure 7.** Gluing cycles (red) join cycles from the cycle factor (black).

**2.3. Gluing cycles.** The proof proceeds by gluing the cycles of the factor $F_k$ together via small gluing cycles. Given a cycle factor in a graph, a *gluing cycle* $H$ for a set $C_1, \dots, C_\ell$ of cycles from the factor has every second edge in common with one of the cycles $C_i$, in such a way that the symmetric difference of the edge sets of $(C_1 \cup \cdots \cup C_\ell) \triangle H$ is a single cycle on the same vertex set as $C_1, \dots, C_\ell$. In the simplest case $\ell = 2$ the cycle $H$ is a 4-cycle that joins two cycles $C_1, C_2$ as in Figure 7 (a). Unfortunately, the middle levels graph $M_k$ has no 4-cycles at all. However, we can use a 6-cycle $H$ that intersects $C_1, C_2$ as shown in Figure 7 (b), and obtain the same effect.

As it turns out, there is a large collection of such gluing 6-cycles for the factor $F_k$, and all these gluing cycles are edge-disjoint, i.e., the gluing operations do not interfere with each other. Furthermore, each gluing cycle that joins cycles $C_1$ and $C_2$ corresponds to a local modification of the plane trees associated with $C_1$ and $C_2$, which consists in removing a leaf from the plane tree, and reattaching it to a neighboring vertex. We can thus define an auxiliary graph $\mathcal{H}_k$ which has as nodes all plane trees with $k$ edges, corresponding to the cycles of the factor $F_k$, and which has edges between pairs of plane trees that differ in such a local modification, corresponding to gluing cycles; see Figure 8. It remains to show that the auxiliary graph $\mathcal{H}_k$ is connected, which is done by showing that every plane tree can be transformed into a star by a sequence of local modifications as described before.

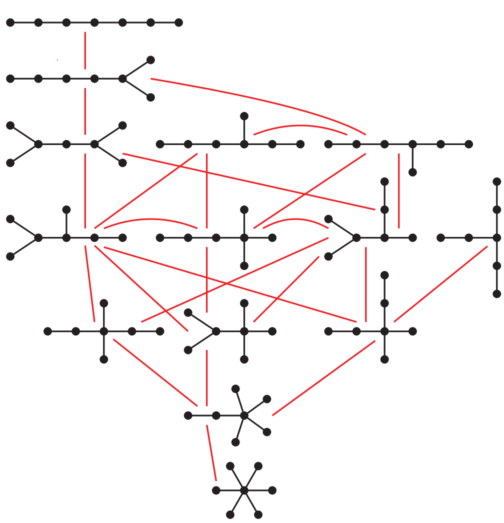


**Figure 8.** Auxiliary graph $\mathcal{H}_k$ on plane trees with $k = 6$ edges.

We thus reduced the problem of proving that $M_k$ has a Hamilton cycle to proving that the auxiliary graph $\mathcal{H}_k$ is connected, which is much easier, as nodes and edges of $\mathcal{H}_k$ have a nice combinatorial interpretation. This two-step approach of building a Hamilton cycle (cycle factor+gluing) and the corresponding reduction to a spanning tree problem is very powerful, and has been employed also in several of the proofs discussed later.

## 3. Bipartite Kneser Graphs

For integers $k \geq 1$ and $n \geq 2k+1$, the *bipartite Kneser graph* $H_{n,k}$ has as vertices all $k$-element and $(n-k)$-element subsets of $[n]$, and an edge between any two subsets $X$ and $Y$ with $X \subseteq Y$. The bipartite Kneser graph $H_{n,k}$ is the cover graph of the subposet of the Boolean lattice of $[n]$ induced by the levels $k$ and $n-k$. In particular, $H_{2k+1,k} = M_k$ is the middle levels graph, i.e., bipartite Kneser graphs generalize the middle levels graphs. Furthermore, bipartite Kneser graphs are vertex-transitive, which makes them interesting test cases for Lovász' conjecture. Simpson [Sim91] and independently Roth conjectured in 1991 that all bipartite Kneser graphs admit a Hamilton cycle. Note that the degrees of $H_{n,k}$ are large when $n$ is large w.r.t. $k$, i.e., intuitively, the middle levels case $n = 2k+1$ is the sparsest and hardest one, whereas the denser cases $n > 2k+1$ should be easier to prove. The densest graph $H_{n,1}$ is the cover graph of what poset theorists call the 'standard example', namely a complete bipartite graph minus a perfect matching. Indeed, there has been considerable work on establishing that sufficiently dense bipartite Kneser graphs $H_{n,k}$ have a Hamilton cycle. Once the sparsest case $n = 2k+1$ was established with the proof of the middle levels conjecture, the Hamiltonicity of all $H_{n,k}$ was shown shortly thereafter by Mütze and Su [MS17] in 2017. In fact, their proof is a 5-page inductive argument, which uses the sparsest case $n = 2k+1$ as a basis.

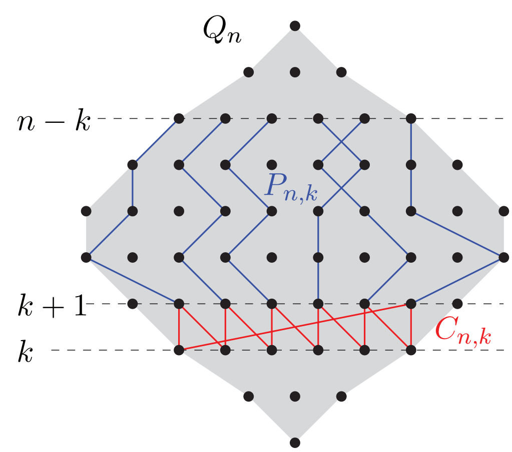



**Figure 9.** Structures in $Q_n$ used for the proof that $H_{n,k}$ has a Hamilton cycle.

### 3.1. Havel's construction and its subsequent refinement.

We write $Q_{n,k}$ for the subgraph of the hypercube $Q_n$ induced by levels $k$ and $k+1$. Already Havel [Hav83] in his 1983 paper considered the following strengthening of the middle levels conjecture: For any $k \geq 1$ and $n \geq 2k+1$, there is a cycle $C_{n,k}$ in $Q_{n,k}$ that visits all vertices in level $k$, i.e., in the smaller of the two partition classes, shown in red in Figure 9. For $n = 2k+1$ both partition classes have the same size and $C_{n,k}$ is a Hamilton cycle in $M_k$, i.e., this statement is the middle levels conjecture. For $n > 2k+1$ Havel proved this statement by an easy induction: Indeed, we can split $Q_{n,k}$ into two subgraphs $Q_{n-1,k}$ and $Q_{n-1,k-1}$, by partitioning all vertices according to the value of the last bit. Using induction, we can glue together the two cycles $C_{n-1,k}$ and $C_{n-1,k-1}$ to obtain $C_{n,k}$. Note that this approach fails for $n = 2k+1$, as in this case $Q_{n-1,k}$ lies above the middle and $Q_{n-1,k-1}$ lies below the middle, so the size difference between lower and upper level is opposite in both parts.

The method of Mütze and Su extends Havel's idea as follows: In addition to the cycle $C_{n,k}$ in $Q_{n,k}$, we maintain a set of vertex-disjoint paths $P_{n,k}$ in $Q_n$, shown in blue in Figure 9, each of which starts at a vertex of $C_{n,k}$ in level $k+1$ and ends at a vertex in level $n-k$, and visits only one vertex of each level in between. The number of such paths equals the number of vertices in levels $k$ or $n-k$, namely $\binom{n}{k} = \binom{n}{n-k}$, i.e., these paths visit all vertices in level $n-k$, but they do not visit all vertices in the levels below that. For $n = 2k+1$ the paths $P_{n,k}$ have no edges, so this strengthening is again the middle levels conjecture. For $n > 2k+1$ the cycles $C_{n,k}$ and paths $P_{n,k}$ can be constructed following a very similar inductive approach as before, by partitioning vertices according to the last bit and gluing together two copies of the structures obtained by induction.

From the cycle $C_{n,k}$ and the paths $P_{n,k}$ we construct a Hamilton cycle in $H_{n,k}$ as follows: We replace each of the vertices of the cycle $C_{n,k}$ in level $k+1$, which is the starting vertex of some path from $P_{n,k}$, by the other end vertex of this path in level $n-k$. This gives a cyclic sequence in which all vertices in level $k$ are interleaved with all vertices in level $n-k$, with the additional property that the predecessor and successor of any level-$k$ vertex are reachable from it by a path in $Q_n$ that moves up from level $k$ to level $n-k$. As moving up along a path in $Q_n$ corresponds to moving to a superset, this sequence is indeed a Hamilton cycle in $H_{n,k}$. This cycle has the remarkable additional closeness property that any two consecutive $k$-sets differ only in the exchange of a single element (as they have a common neighbor in level $k+1$). This construction is illustrated in Figure 10 (a1)–(a4) for the cases $n = 9$ and $k = 1, \ldots, 4$ ($k = 4$ is a solution to the middle levels conjecture).

## 4. Kneser Graphs

For integers $k \geq 1$ and $n \geq 2k+1$, the *Kneser graph* $K_{n,k}$ has as vertices all $k$-element subset of $[n]$, and edges between any two disjoint sets. Kneser graphs have many interesting properties, for example, their chromatic number was shown to be $n-2k+2$ by Lovász using topological methods, and their independence number is $\binom{n-1}{k-1}$ by the famous Erdős-Ko-Rado theorem. Kneser graphs are clearly vertex-

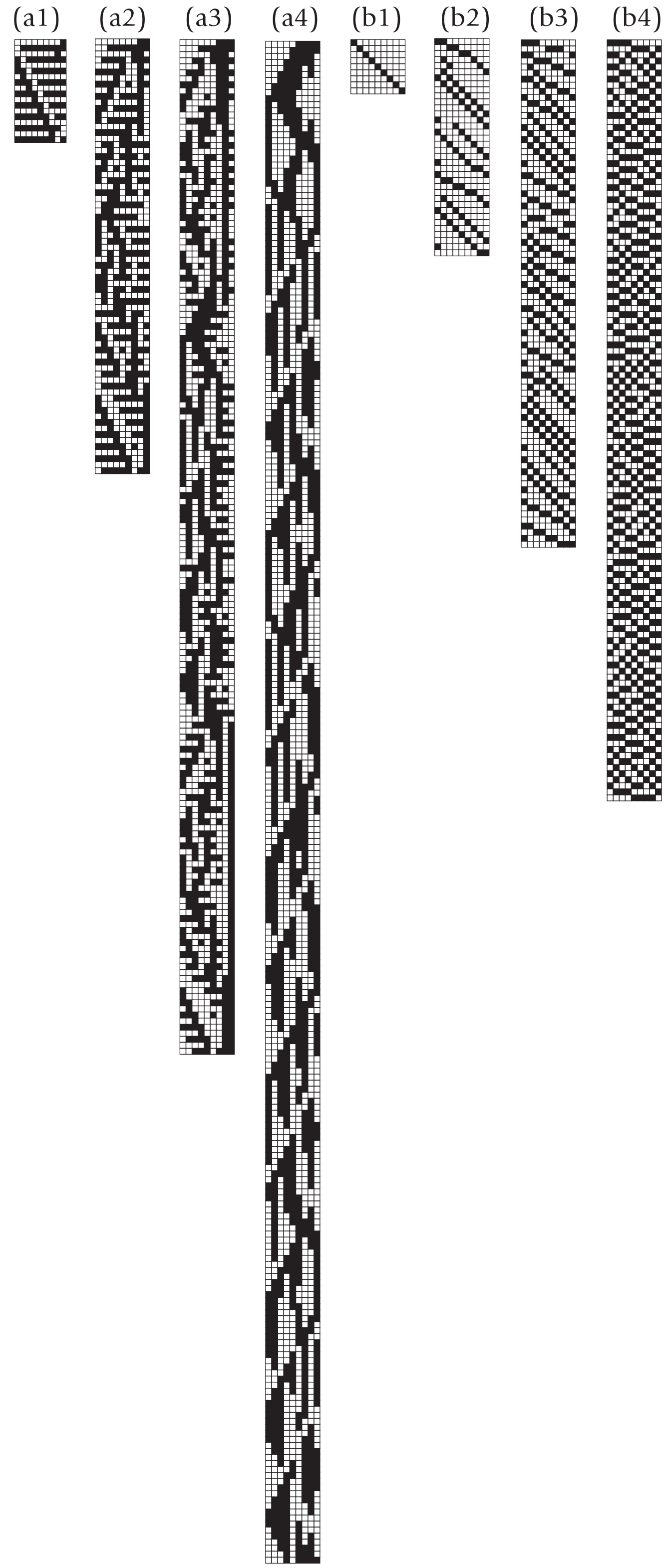
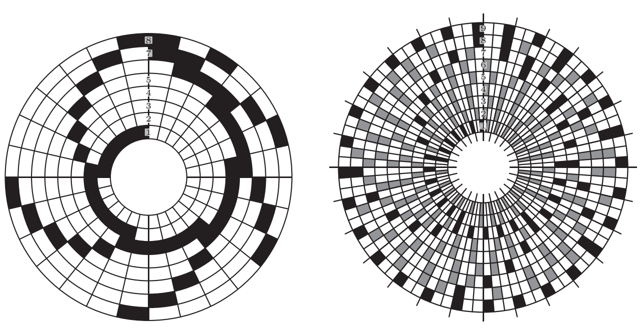



**Figure 10.** Hamilton cycles in (a1)–(a4) bipartite Kneser graphs $H_{9,k}$, for $k = 1, 2, 3, 4$; and (b1)–(b4) Kneser graphs $K_{9,k}$, for $k = 1, 2, 3, 4$. Vertices are in their bitstring representation ($0$ = white and $1$ = black).

transitive, and $H_{n,k}$ is the *bipartite double cover* of $K_{n,k}$, i.e., we take two copies of $K_{n,k}$ and replace every corresponding pair of edges inside the copies by the two 'diagonal' cross edges between them. Consequently, if $K_{n,k}$ admits a Hamilton cycle, then $H_{n,k}$ admits a Hamilton path or cycle. Indeed, given a Hamilton cycle $C = (X_1, \dots, X_\ell)$ in $K_{n,k}$, where $\ell = \binom{n}{k}$, we define $\overline{X_i} := [n] \setminus X_i$, and we consider the two sequences $P := (X_1, \overline{X_2}, X_3, \overline{X_4}, \dots)$ and $P' := (\overline{X_1}, X_2, \overline{X_3}, X_4, \dots)$, both of length $\ell$, obtained by complementing every even- or odd-indexed set in $C$, respectively. The last entries of $P$ and $P'$ are $X_\ell$ and $\overline{X_\ell}$, respectively, if $\ell$ is odd, and vice versa if $\ell$ is even. Furthermore, as $X_i \cap X_{i+1} = \emptyset$ we have $X_i \subseteq \overline{X_{i+1}}$ and $\overline{X_i} \supseteq X_{i+1}$, so $P$ and $P'$ are paths in $H_{n,k}$. Furthermore, if $\ell$ is odd, then their concatenation $PP'$ is a Hamilton cycle in $H_{n,k}$. On the other hand, if $\ell$ is even, then the two end vertices of $P$ are adjacent and the two end vertices of $P'$ are adjacent, so these two disjoint cycles in $H_{n,k}$ can be joined to a Hamilton path.

The sparsest Kneser graphs are obtained when $n = 2k + 1$, and they are also known as *odd graphs* $O_k := K_{2k+1,k}$. The odd graph $O_2 = K_{5,2}$ is the Petersen graph shown in Figure 2, which does not have a Hamilton cycle, but only a Hamilton path. The graph $O_3 = K_{7,3}$ is shown in Figure 13. The conjecture that $O_k$ for $k \geq 3$ has a Hamilton cycle was raised in the 1970s, even before the middle levels conjecture, in papers by Meredith and Lloyd [ML73], and by Biggs [Big79]. By our earlier observation, the Hamiltonicity of $K_{n,k}$ implies it for $H_{n,k}$. In particular, Hamiltonicity of the odd graphs implies the middle levels conjecture. Consequently, Kneser graphs attracted a lot of attention, and there was a long line of research on proving that sufficiently dense Kneser graphs $K_{n,k}$, i.e., those where $n$ is large w.r.t. $k$, admit a Hamilton cycle.



(a) (b)

**Figure 11.** (a) Exchange Gray code for 2-element subsets of [8] obtained by restricting the binary reflected Gray code in $Q_8$ to level 2; (b) Hamilton cycle in $K_{9,3}$ via the Chen-Füredi construction. Each indicated triple of vertices is a partition of [9], the first two are colored gray, and the last one is colored black and corresponds to the set from (a) by adding the last element (extra outermost black bit).

### 4.1. The Chen-Füredi construction via Baranyai's partition theorem.

In 2002, Chen and Füredi [CF02] found a particularly nice proof that $K_{n,k}$ has a Hamilton cycle when $n = pk$ for some integer $p \geq 3$. The first ingredient of their proof is *Baranyai's partition theorem*, which states that the $\binom{n}{k}$ vertices of the Kneser graph can be partitioned into $\binom{n}{k}/p$ groups of size $p$ such that the vertices in each group are a partition of $[n]$, i.e., they are pairwise disjoint and together cover $[n]$. The second ingredient is a method to list all $k$-element subset of $[n]$ in such a way that any two consecutive sets $X, Y$ differ in an element exchange, i.e., $Y = (X \setminus \{i\}) \cup \{j\}$ for some $i, j \in [n]$. It is well known that such a listing can be obtained from the binary reflected Gray code for $Q_n$ by restricting it to the vertices in level $k$, i.e., we simply delete from the full listing all vertices not in level $k$; see Figure 11 (a).

To prove that $K_{n,k}$ with $n = pk$ and $p \geq 3$ has a Hamilton cycle, Chen and Füredi first apply Baranyai's theorem, which partitions all vertices of $K_{n,k}$ into $\ell := \binom{n}{k}/p$ groups of size $p$, such that each group is a partition of $[n]$. Let $X_1^i, \dots, X_p^i$ be the sets in the $i$th group, for every $i \in [\ell]$. Each of those groups forms a clique in the Kneser graph, i.e., it can be traversed in any order. Without loss of generality we may assume that $n \in X_p^i$, i.e., the element $n$ is contained in the last set of each group. Furthermore, by the aforementioned Gray code result, we can assume that $X_p^1, X_p^2, \dots, X_p^\ell$ are ordered so that any two consecutive sets differ in an element exchange, i.e., $X_p^1 \setminus \{n\}, \dots, X_p^\ell \setminus \{n\}$ forms an exchange Gray code for all $(k-1)$-element subsets of $[n-1]$. Let $x^i \in [n]$ be the element in $X_p^i$ that is not contained in $X_p^{i+1}$ (the indices $i$ are considered modulo $\ell$). We may also assume that $X_1^{i+1}$ does not contain $x^i$, otherwise the sets $X_1^{i+1}, \dots, X_{p-1}^{i+1}$ can be reordered appropriately, which is possible as there are at least $p - 1 \geq 2$ of them. It follows that $X_p^i \cap X_1^{i+1} = \emptyset$ and consequently, $X_1^1, \dots, X_p^1, X_1^2, \dots, X_p^2, \dots, X_1^\ell, \dots, X_p^\ell$ is the desired Hamilton cycle in $K_{n,k}$; see Figure 11 (b). The method via Baranyai partitions was later refined by Chen [Che00] to establish Hamiltonicity of all $K_{n,k}$ with $n \geq 2.62k + 1$.

### 4.2. Settling the odd graphs via a Chung-Feller bijection.

In 2021, Mütze, Nummenpalo, and Walczak [MNW21] proved that the sparsest Kneser graphs, namely the odd graphs $O_k$ for all $k \geq 3$ have a Hamilton cycle. The starting point of their proof is the well-known *Chung-Feller theorem*. A *flaw* in a bitstring $x$ is a prefix of $x$ ending with 0 that has strictly less 1s than 0s; flaws are drawn as red steps in Figure 12. We write $L_k$ for the unique middle level $k$ of $Q_{2k}$, i.e., all bitstrings of length $2k$ with exactly $k$ many 1s, and we partition $L_k$ into sets $L_k^e$ for $e = 0, \dots, k$ according to the number $e$ of flaws. In particular, $L_k^0 = D_k$ are Dyck words, and $L_k^k$ are complemented Dyck words. The Chung-Feller theorem asserts

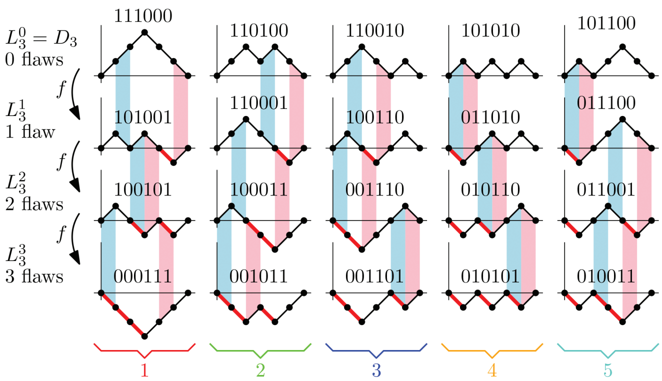



**Figure 12.** Proof of the Chung-Feller theorem for $k = 3$ via the minimum change bijection $f$. The two transposed bits are highlighted by vertical bars. Each column produces one of the five cycles of the factor in $O_3$ shown in Figure 13.

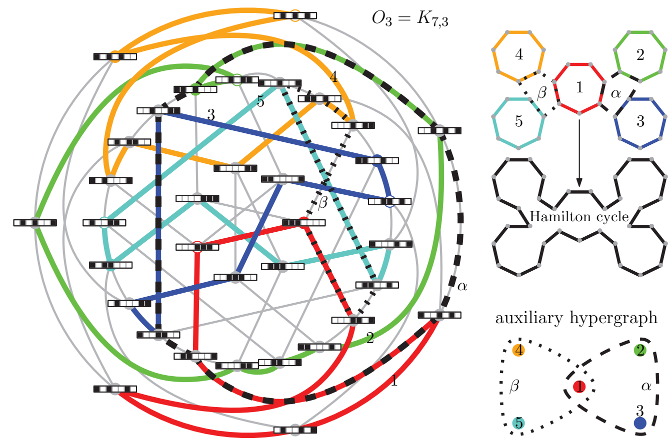



**Figure 13.** Cycle factor in $O_3$ obtained from the bijection $f$ in Figure 12, and corresponding gluing cycles to turn it into a Hamilton cycle.

that $|L_k^0| = |L_k^1| = \cdots = |L_k^k| = \frac{1}{k+1}\binom{2k}{k} = C_k$, i.e., the number of strings is the same independently of the number of flaws, and it is the $k$th Catalan number. It is not hard to prove this by establishing a bijection $f : L_k^e \to L_k^{e+1}$. In 2018, Mütze, Standke and Wiechert presented a new proof, using a bijection $f$ that has the following additional properties; see Figure 12: $f$ only transposes two bits (0 and 1), for any $x \in L_k$ we have that $f^k(x)$ is the complement of $x$, i.e., every bit is transposed (and thus complemented) exactly once when applying the bijection $k$ times, and the unique neighbors $\hat{x} := x \cup f(x)$ of $x$ and $f(x)$ in level $k+1$ of $Q_{2k}$ for $x \in L_k$ are all distinct and together cover precisely this level. We can thus build vertex-disjoint paths $(x, \hat{x}, f(x), \widehat{f(x)}, f^2(x), \widehat{f^2(x)}, \ldots, f^k(x))$, all of length $2k$, that together cover $Q_{2k,k}$ and that connect pairs of Dyck words and their complements (the length of each path is the number of its edges, which is one less than the number of vertices). Appending a 0-bit to all vertices and taking complements of the resulting vertices in level $k+1$ yields a cycle factor in $O_k$ which has $C_k$ many cycles of the same length $2k+1$; see Figure 13.

The proof of Mütze, Nummenpalo, and Walczak is completed by gluing the cycles of this factor together via 6-cycles and 8-cycles, which glue together 3 or 4 cycles, respectively, from the factor at a time. The main technical difficulty is that the corresponding auxiliary graph is now a hypergraph with hyperedges of cardinality 3 or 4, respectively; see the bottom right of Figure 13 (only 3-hyperedges are present in the figure). To obtain a Hamilton cycle in $O_k$, we seek a so-called *loose* spanning tree in the auxiliary hypergraph, i.e., a spanning tree in which any two hyperedges overlap at most in a singleton. For this it is not enough to prove that the hypergraph is connected, as there are connected hypergraphs that do not admit any loose spanning tree. Instead the paper constructs one particular loose spanning tree. The resulting Hamilton cycle in $O_4$ is shown in Figure 10 (b4).

### 4.3. Johnson's inductive construction.

In 2011, Johnson [Joh11] devised an inductive construction for Hamilton cycles in Kneser graphs. Specifically, he showed that $K_{n,k}$ for $n = 2k + s$ with even $s$ has a Hamilton cycle provided that the smaller Kneser graphs $K_{2\ell+s/2,\ell}$ have a Hamilton cycle for all $1 \le \ell \le \lfloor k/2 \rfloor$ (or they are the Petersen graph $K_{5,2}$).

His construction works for a ground set of even cardinality $n = 2k + s$ by partitioning it into fixed pairs $\{2i-1, 2i\}$ for $i = 1, \ldots, n/2$, and by considering the possible intersection patterns of subsets with those pairs. Specifically, we associate a set $X \subseteq [n]$ with a tuple $(X(1), \ldots, X(n/2)) \in \{-1, 0, 1, 2\}^{n/2}$ by defining $X(i) := 0, -1, 1$, or 2 if $X \cap \{2i-1, 2i\}$ equals $\emptyset, \{2i-1\}, \{2i\}$, or $\{2i-1, 2i\}$, respectively. In the simplest case, the subsets $X$ intersect the pairs in either 0 or 2 elements, i.e., $X(i) \in \{0, 2\}$ for all $i \in [n/2]$. In this case, for $\ell = k/2$ a Hamilton cycle in $K_{k+s/2,k/2}$ can be lifted to a cycle in $K_{n,k} = K_{2k+s,k}$ by replacing every element $i$ by the pair $\{2i-1, 2i\}$. More generally, consider subsets $X$ which intersect all but a fixed set of $t$ pairs in 0 or 2 elements, i.e., $X(i) \in \{-1, 1\}$ for a fixed $t$-set of indices $i \in [n/2]$. An edge in $K_{k+s/2-t,k/2-t/2}$ together with a $t$-set in $[n/2]$ lifts to a set of edges involving subsets which intersect all but this fixed set of $t$ pairs in 0 or 2 elements. For example, we have many edges (one for each pattern of $\pm 1$s) of the form $((0,0,2,0,2,2,0,\pm 1,\pm 1),(2,2,0,0,0,0,2,\mp 1,\mp 1))$ in $K_{18,8}$ that arise from the edge $(\{3,5,6\},\{1,2,7\})$ in $K_{7,3}$ with $\{8,9\}$ as the special $t$-set. Using this idea, a Hamilton cycle in $K_{k+s/2-t,k/2-t/2}$ lifts to a cycle (possibly of double the length) consisting of sets of this type in $K_{n,k}$. With some care, one can join together these cycles corresponding to different $t$-sets, and then for different values of $t$ to give a Hamilton cycle in $K_{n,k}$.

Combining Johnson's result with the solution for the sparsest case $n = 2k+1$ presented in the previous section, we obtain that $K_{2k+2^a,k}$ has a Hamilton cycle for all $k \ge 3$

and $a \geq 0$. This settles in particular the second-sparsest case $K_{2k+2,k}$.

**4.4. Settling the remaining cases via Greene-Kleitman parenthesis matching and gliders.** In 2022, Merino, Mütze, and Namrata [MMN22] proved that $K_{n,k}$ for $n \geq 2k+3$ has a Hamilton cycle, which combined with the results from the previous two sections completely settles the problem for Kneser graphs. Their proof starts with a new cycle factor in $K_{n,k}$, which is constructed using the following simple rule based on *parenthesis matching*, a technique that was pioneered by Greene and Kleitman [GK76] in the context of symmetric chain partitions of the Boolean lattice: We consider vertices of $K_{n,k}$ as bitstrings, and we interpret the 1s in $x$ as opening brackets and the 0s as closing brackets, and we match closest pairs of opening and closing brackets in the natural way, which will leave some 0s unmatched. This matching is done *cyclically* across the boundary of $x$, i.e., $x$ is considered as a cyclic string. We write $f(x)$ for the vertex obtained from $x$ by complementing all matched bits, leaving the unmatched bits unchanged. Note that $x$ and $f(x)$ have no 1s at the same positions, implying that $(x, f(x))$ is an edge in the Kneser graph. Furthermore, $f$ is invertible and and $f^2(x) \neq x$, so the union of all edges $(x, f(x))$ is a collection of disjoint cycles that together visit all vertices of $K_{n,k}$; see Figure 14.

The next step is to understand the structure of the cycles generated by $f$. Interestingly, the evolution of a bitstring $x$ under repeated applications of $f$ can be described by a kinetic system of multiple gliders that move at different speeds and that interact over time, somewhat reminiscent of the gliders in Conway's Game of Life. Specifically, each application of $f$ is viewed as one unit of time moving forward. Furthermore, we partition the matched bits of $x$ into groups, and each of these groups is called a *glider*. A glider has a *speed* associated to it, which is given by the number of 1s in its group. For example, in the cycle shown in Figure 14 (a), there is a single matched 1 and the corresponding matched 0, and together these two bits form a glider of speed 1 that moves one step to the right in every time step. Applying $f$ means going down to the next row in the picture, so the time axis points downward. Similarly, in Figure 14 (b), there are two matched 1s and the corresponding two matched 0s, and together these four bits form a glider of speed 2 that moves two steps to the right in every time step. As we see from these examples, a single glider of speed $v$ simply moves uniformly, following the basic physics law

$$s(t) = s(0) + v \cdot t,$$

where $t$ is the time (i.e., the number of applications of $f$) and $s(t)$ is the position of the glider in the bitstring as a function of time. The position $s(t)$ has to be considered modulo $n$, as bitstrings are considered as cyclic strings and the gliders hence wrap around the boundary. The situation gets more interesting and complicated when gliders of different speeds interact with each other. For example, in Figure 14 (c), there is one glider of speed 2 and one glider of speed 1. As long as these groups of bits are separated, each glider moves uniformly as before. However, when the speed 2 glider catches up with the speed 1 glider, an overtaking occurs. During an overtaking, the faster glider receives a boost, whereas the slower glider is delayed. This

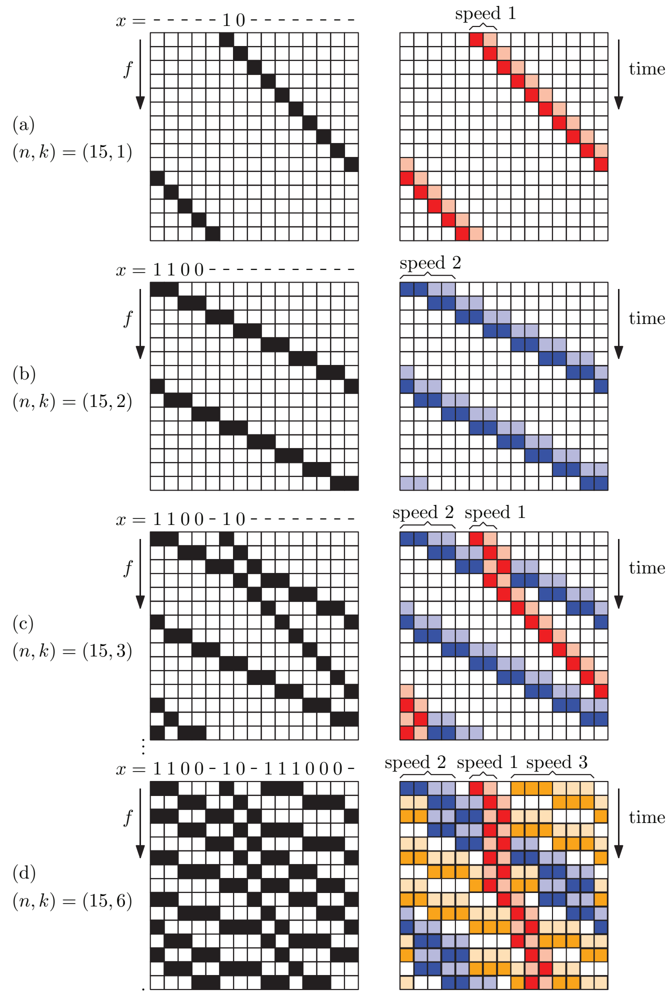



**Figure 14.** Cycles in different Kneser graphs $K_{n,k}$ constructed by parenthesis matching. The cycles in (a) and (b) are shown completely, whereas in (c) and (d) only the first 15 vertices are shown. When applying parenthesis matching to $x$, unmatched 0s are printed as -. The right-hand side shows the interpretation of certain groups of bits as gliders, and their movement over time. Matched bits belonging to the same glider are colored in the same color, with the opaque filling given to 1-bits, and the transparent filling given to 0-bits. (a) one glider of speed 1; (b) one glider of speed 2; (c) two gliders with speeds 1 and 2 that participate in an overtaking; (d) three gliders of speeds 1, 2, and 3 that participate in multiple overtakings. Animations of these examples are available at [Müt23b].

can be captured by augmenting the corresponding equations of motion by introducing additional terms, making them nonuniform. In the simplest case of two gliders of different speeds, the equations become

$$s_1(t) = s_1(0) + v_1 \cdot t - 2v_1 c_{1,2},$$
$$s_2(t) = s_2(0) + v_2 \cdot t + 2v_1 c_{1,2},$$

where the subscript 1 stands for the slower glider and the subscript 2 stands for the faster glider, and the additional variable $c_{1,2}$ counts the number of overtakings. Note that the terms $2v_1c_{1,2}$ occur with opposite signs in both equations, capturing the fact that the faster glider is boosted by the same amount that the slower glider is delayed. This can be seen as 'energy conservation' in the system of gliders. For more than two gliders, the equations of motion can be generalized accordingly, by introducing additional overtaking counters between any pair of gliders. From those equations of motions, important properties of the cycles can be extracted via combinatorial and algebraic arguments. One such property is that the number of gliders and their speeds are invariant along each cycle. For example, in Figure 14 (d), every bitstring along this cycle has three gliders of speeds 1, 2, and 3. For the reader's entertainment, we programmed an interactive animation of gliders over time, and we encourage experimentation with this code, which can be found at [Müt23b].

The last step of the proof joins the cycles of this factor via gluing 4-cycles (this is where the assumption $n \geq 2k+3$ is used). Specifically, the gluing cycles join pairs of cycles whose sets of glider speeds differ in a small modification, changing the speed of one glider by $-1$ and the speed of another by $+1$. We thus obtain a combinatorial interpretation of the gluings. To prove that all cycles of the factor can be joined to a single Hamilton cycle, it is argued that all cycles can be joined to one particular cycle in the factor, by considering the speed sets of gliders as number partitions, and by arguing that these partitions increase lexicographically along suitable gluings. The Hamilton cycles in $K_{9,k}$ for $k = 1, 2, 3$ resulting from this proof are shown in Figure 10 (b1)–(b3).

## 5. Generalized Johnson Graphs

The *generalized Johnson graph* $J_{n,k,s}$ has as vertices all $k$-element subsets of $[n]$, and an edge between any two sets whose intersection has size exactly $s$. It is defined for integers $k \geq 1$, $0 \leq s < k$ and $n \geq 2k - s + \mathbf{1}_{s=0}$, where $\mathbf{1}_{s=0}$ denotes the indicator function that equals 1 if $s = 0$ and 0 otherwise. For $s = 0$ we obtain Kneser graphs ($s = 0$), and for $s = k - 1$ we obtain Johnson graphs as special cases. By taking complements, we see that $J_{n,k,s}$ is isomorphic to $J_{n,n-k,n-2k+s}$. Chen and Lih [CL87] conjectured in 1987 that all graphs $J_{n,k,s}$ admit a Hamilton cycle except the Petersen graph $J_{5,2,0} = J_{5,3,1}$. This includes Hamiltonicity of the corresponding bipartite double covers, in particular a solution to the middle levels conjecture, which was the starting point of this article.

In fact, already Chen and Lih observed that $J_{n,k,s}$ can be partitioned into two subgraphs isomorphic to $J_{n-1,k,s}$ and $J_{n-1,k-1,s-1}$ (split vertices according to containment of some fixed element, say $n$), so if these two graphs have a Hamilton cycle, then we can glue them via a 4-cycle and obtain a Hamilton cycle in $J_{n,k,s}$. To complete the proof, it remains to observe that if $J_{n,k,s}$ is a generalized Johnson graph, then either it is a Kneser graph, or $J_{n-1,k,s}$ and $J_{n-1,k-1,s-1}$ are both generalized Johnson graphs. Using the results for Kneser graphs from the previous section we thus obtain Hamiltonicity for all generalized Johnson graphs by induction.

There is another closely related and heavily studied class of vertex-transitive graphs called *generalized Kneser graphs* $K_{n,k,s}$. This graph has as vertices all $k$-element subsets of $[n]$, and an edge between any two sets whose intersection has size *at most* $s$. Clearly, $J_{n,k,s}$ is a spanning subgraph of $K_{n,k,s}$, so the Hamiltonicity of $K_{n,k,s}$ is immediate.

## 6. What's Next?

Generalized Johnson graphs are the most general family of graphs defined by intersecting set systems and thus a natural end to our story. With regards to Lovász' conjecture, many other families of vertex-transitive graphs await to be tested for Hamiltonicity, in particular Cayley graphs. We may also ask how many distinct Hamilton cycles a vertex-transitive graph admits. In fact, the proofs via gluings of cycle factors for the middle levels graph $M_k$ and for the odd graph $O_k$ discussed in this article yield double-exponentially (in $k$) many distinct Hamilton cycles, and the trivial upper bound of $n!$ for the number of Hamilton cycles of an $n$-vertex graphs also yields a double-exponential function in $k$. A much harder problem is to find multiple *edge-disjoint* Hamilton cycles. Biggs [Big79] conjectured that the odd graph $O_k$ can be partitioned into $\lfloor (k+1)/2 \rfloor$ edge-disjoint Hamilton cycle for $k \geq 3$. We do not even know two edge-disjoint Hamilton cycles in the middle levels graph $M_k$. Katona conjectured that the Kneser graph $K_{n,k}$ contains the $r$th power of a Hamilton cycle, where $r := \lfloor n/k \rfloor - 2$, and this may even be true for $r := \lceil n/k \rceil - 2$. Also, for the graphs considered in this article, we may ask whether they are *Hamilton-connected*, i.e., they admit a Hamilton path between any two prescribed end vertices. For bipartite graphs, we may ask whether they are *Hamilton-laceable*, i.e., they admit a Hamilton path between any two prescribed end vertices, one from each partition class. In fact, the middle levels graph $M_k$ for $k \geq 2$ was shown to be Hamilton-laceable. Another generalization is to consider the *cycle spectrum*, which is the set of all possible cycle lengths in a graph. For example, does the middle levels graph $M_k$ admit cycles of all possible even lengths starting from 6 up to the number of vertices?

From an algorithmic point of view, one may ask which of the cycles described in this article can be computed efficiently? A satisfactory answer to this question is only known for the middle levels graph $M_k$, while all the other known constructions present fundamental obstacles to such algorithms. Furthermore, what about simple descriptions of Hamilton cycles, similar in flavor to Williams' greedy description of the binary reflected Gray code mentioned in Section 1.1? Even the simplest known solution of the middle levels conjecture is much more complicated than this.

There are many other intriguing problems about the interaction of different structures, such as matchings and cycles, in vertex-transitive graphs. For example, Ruskey and Savage asked whether every matching in the hypercube can be extended to a Hamilton cycle. For the case of perfect matchings this was answered affirmatively by Fink [Fin07]. Also, Kotzig's question on perfect 1-factorizations of the complete graph comes to mind naturally. A *perfect 1-factorization* is a decomposition of the edge set of a graph into perfect matchings, such that the union of any two of them forms a Hamilton cycle.

## References


[Big79] Norman Biggs, *Some odd graph theory*, Second International Conference on Combinatorial Mathematics (New York, 1978), Ann. New York Acad. Sci., vol. 319, New York Acad. Sci., New York, 1979, pp. 71–81. MR556008

[BW84] Marshall Buck and Doug Wiedemann, *Gray codes with restricted density*, Discrete Math. **48** (1984), no. 2-3, 163–171, DOI 10.1016/0012-365X(84)90179-1. MR737262

[CF02] Ya-Chen Chen and Zoltán Füredi, *Hamiltonian Kneser graphs*, Combinatorica **22** (2002), no. 1, 147–149, DOI 10.1007/s004930200007. MR1883565

[Che00] Ya-Chen Chen, *Kneser graphs are Hamiltonian for $n \geq 3k$*, J. Combin. Theory Ser. B **80** (2000), no. 1, 69–79, DOI 10.1006/jctb.2000.1969. MR1778200

[CL87] Bor Liang Chen and Ko-Wei Lih, *Hamiltonian uniform subset graphs*, J. Combin. Theory Ser. B **42** (1987), no. 3, 257–263, DOI 10.1016/0095-8956(87)90044-X. MR888679

[Fin07] Jiří Fink, *Perfect matchings extend to Hamilton cycles in hypercubes*, J. Combin. Theory Ser. B **97** (2007), no. 6, 1074–1076, DOI 10.1016/j.jctb.2007.02.007. MR2354719

[FT95] Stefan Felsner and William T. Trotter, *Colorings of diagrams of interval orders and $\alpha$-sequences of sets*, Discrete Math. **144** (1995), no. 1-3, 23–31, DOI 10.1016/0012-365X(94)00283-O. Combinatorics of ordered sets (Oberwolfach, 1991). MR1350586

[GK76] Curtis Greene and Daniel J. Kleitman, *Strong versions of Sperner's theorem*, J. Combinatorial Theory Ser. A **20** (1976), no. 1, 80–88, DOI 10.1016/0097-3165(76)90079-0. MR389608

[Hav83] Ivan Havel, *Semipaths in directed cubes*, Graphs and other combinatorial topics (Prague, 1982), Teubner-Texte Math., vol. 59, Teubner, Leipzig, 1983, pp. 101–108. MR737021

[Joh11] J. Robert Johnson, *An inductive construction for Hamilton cycles in Kneser graphs*, Electron. J. Combin. **18** (2011), no. 1, Paper 189, 12, DOI 10.37236/676. MR2836824

[KT88] Henry A. Kierstead and William T. Trotter, *Explicit matchings in the middle levels of the Boolean lattice*, Order **5** (1988), no. 2, 163–171, DOI 10.1007/BF00337621. MR962224

[ML73] Guy H. J. Meredith and E. Keith Lloyd, *The footballers of Croam*, J. Combinatorial Theory Ser. B **15** (1973), 161–166, DOI 10.1016/0095-8956(73)90016-6. MR321782

[MMN22] Arturo Merino, Torsten Mütze, and Namrata, *Kneser graphs are Hamiltonian*, STOC'23—Proceedings of the 55th Annual ACM Symposium on Theory of Computing, ACM, New York, [2023] ©2023, pp. 963–970, DOI 10.1145/3564246.3585137. MR4617441

[MNW21] Torsten Mütze, Jerri Nummenpalo, and Bartosz Walczak, *Sparse Kneser graphs are Hamiltonian*, J. Lond. Math. Soc. (2) **103** (2021), no. 4, 1253–1275, DOI 10.1112/jlms.12406. MR4273468

[MS17] Torsten Mütze and Pascal Su, *Bipartite Kneser graphs are Hamiltonian*, Combinatorica **37** (2017), no. 6, 1207–1219, DOI 10.1007/s00493-016-3434-6. MR3759914

[Müt16] Torsten Mütze, *Proof of the middle levels conjecture*, Proc. Lond. Math. Soc. (3) **112** (2016), no. 4, 677–713, DOI 10.1112/plms/pdw004. MR3483129

[Müt23a] Torsten Mütze. A book proof of the middle levels theorem. *arXiv:2306.13019*. To appear in *Combinatorica*, 2023.

[Müt23b] Torsten Mütze. Gliders in Kneser graphs, 2023. `http://tmuetze.de/gliders.html`.

[Sim91] James E. Simpson, *Hamiltonian bipartite graphs*, Proceedings of the Twenty-second Southeastern Conference on Combinatorics, Graph Theory, and Computing (Baton Rouge, LA, 1991), Congr. Numer. **85** (1991), 97–110. MR1152123

[SW95] Carla D. Savage and Peter Winkler, *Monotone Gray codes and the middle levels problem*, J. Combin. Theory Ser. A **70** (1995), no. 2, 230–248, DOI 10.1016/0097-3165(95)90091-8. MR1329390


Torsten Mütze

## Credits


Figure 1 is courtesy of The Puzzle Museum.
Figures 2–14 are courtesy of Torsten Mütze.
Photo of Torsten Mütze is courtesy of Fanny Mütze.

# Mixing in Incompressible Flows: Transport, Dissipation, and Their Interplay

*Michele Coti Zelati, Gianluca Crippa, Gautam Iyer, and Anna L. Mazzucato*

## 1. Introduction

Mixing in fluid flows is a ubiquitous phenomenon and arises in many situations ranging from everyday occurrences, such as mixing of cream in coffee, to fundamental physical processes, such as circulation in the oceans and the atmosphere. From a theoretical point of view, mixing has been studied since the late nineteenth century in different contexts, including dynamical systems, homogenization, control, hydrodynamic stability and turbulence theory. Although certain aspects of these theories still elude us, significant progress has allowed to provide a rigorous mathematical description of some fundamental mixing mechanisms. In this survey, we address mixing from the point of view of partial differential equations, motivated by applications that arise in fluid dynamics. A prototypical example is the movement of small, light tracer particles (in laboratory experiments, these are often tiny glass beads) in a liquid. One can visually see the particles "mix," and our interest is to quantify this phenomenon mathematically and formulate rigorous results in this context.

When the diffusive effects are negligible, the evolution of the density of tracer particles is governed by the transport equation

$$\partial_t \rho + u \cdot \nabla \rho = 0. \tag{1.1}$$

Here $\rho = \rho(t, x)$ is a scalar representing the density of tracer particles, and $u = u(t, x)$ denotes the velocity of the ambient fluid. We will always assume that the ambient fluid is *incompressible*, which mathematically translates to the requirement that $u$ is divergence free. Moreover, we will only study situations where $\rho$ is a *passive scalar* (or passively advected scalar)—that is, the effect of tracer particles on the flow is negligible and the evolution of $\rho$ does not influence the velocity field $u$. One example where passive advection arises in nature is when light, chemically nonreactant particles are carried by a large fluid body (e.g., plankton blooms in the ocean). Examples of *active scalars* (i.e., tracers for which their effect on the flow cannot be neglected) are quantities such as salinity and temperature in geophysical contexts.

Our interest is to study mixing away from boundaries, and hence we will study (1.1) with periodic boundary conditions. For simplicity, and clarity of exposition, we fix the dimension $d = 2$. The spatial domain is henceforth the torus $\mathbb{T}^2$, which we normalize to have unit sidelength. We mention, however, that most of the results we state can be extended to higher dimensions without too much difficulty. We supplement (1.1) with an initial condition $\rho_0$ at time $t = 0$. If the velocity $u$ is sufficiently regular (Lipschitz continuous in space uniformly in time, to be precise), solutions to equation (1.1) can be expressed explicitly in terms of the time-dependent flow map $\Phi(t, t_0, x)$ of the velocity field $u$, which is obtained by solving the system of

*Michele Coti Zelati is a reader in the Department of Mathematics at Imperial College, London, UK. His email address is* m.coti-zelati@imperial.ac.uk.

*Gianluca Crippa is a professor of analysis in the Department of Mathematics and Computer Science at University of Basel, Switzerland. His email address is* gianluca.crippa@unibas.ch.

*Gautam Iyer is a professor of mathematics at Carnegie Mellon University, Pittsburgh, USA. His email address is* gautam@math.cmu.edu.

*Anna L. Mazzucato is a professor of mathematics at Penn State University, University Park, USA. Her email address is* alm24@psu.edu.

*This work has been partially supported by the US National Science Foundation under grant DMS-2108080 and the CMU Center for Non Linear Analysis to GI, the US National Science Foundation under grants DMS-1615457, DMS-1909103, DMS-2206453 and Simons Foundation Grant 1036502 to ALM, the European Research Council under grant 676675 FLIRT and the Swiss National Science Foundation under grant 212573 FLUTURA to GC, the Royal Society URF\R1\191492 and the EPSRC Horizon Europe Guarantee EP/X020886/1 to MCZ.*

*Communicated by* Notices *Associate Editor Daniela De Silva.*

*For permission to reprint this article, please contact:* reprint-permission@ams.org.

DOI: https://doi.org/10.1090/noti2929

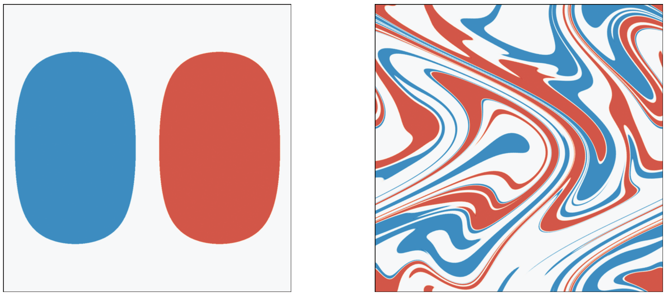


**Figure 1.** Example of mixing. The red and blue level sets in both figures have exactly the same area. For the left figure, averages on scales comparable to $1/8^{\text{th}}$ of the period are of order 1. On the right, however, the sets are stretched and interspersed in such a manner that averages at the same scale are much smaller.

ordinary differential equations:

$$\begin{cases}\partial_t \Phi(t,t_0;x) = u(t,\Phi(t,t_0;x)), & t_0,t\in\mathbb{R},\ x\in\mathbb{T}^2,\\ \Phi(t_0,t_0;x)=x, & x\in\mathbb{T}^2.\end{cases}$$

Now, a direct calculation shows that the unique solution to (1.1) with initial condition $\rho_0$ is given by the formula

$$\rho(t,x)=\rho_0(\Phi(0,t;x)).$$

This is known as the method of characteristics.

The condition $\operatorname{div} u = 0$ is equivalent to imposing that for each time $t$, $t_0$, the map $\Phi(t,t_0;\cdot)\colon \mathbb{T}^2\to\mathbb{T}^2$ is area preserving. Since $u$ is divergence free, integrating (1.1) in space shows that the total mass

$$\bar\rho = \int_{\mathbb{T}^2}\rho(t,x)\,\mathrm{d}\,x$$

is a conserved quantity (i.e., remains constant in time). Thus replacing $\rho_0$ with $\rho_0-\bar\rho$ if necessary, there is no loss of generality in assuming $\rho_0$ (and hence $\rho(t,\cdot)$) has zero mean. *We stipulate from now on that* $\bar\rho = 0$. Physically, $\rho$ now represents the deviation from the mean of the density of tracer particles.

Mixing, informally speaking, is the process by which uneven initial configurations transform into a spatially uniform one. In our setting (since the flow is area preserving), the area of regions of relatively higher (or lower) concentration is preserved. That is, for any $c\in\mathbb{R}$, the area of the sublevel sets $\{x \mid \rho(t,x)<c\}$ and the superlevel sets $\{x\mid \rho(t,x)>c\}$ are both constant in time. Thus if initially the set where $\rho_0$ is positive occupies half the torus, then for all time the set where $\rho(t)$ is positive must also occupy half the torus. The process of mixing will transform $\rho$ in such a way that the set $\{\rho(t)>0\}$ will be stretched into many long thin filaments (that still occupy a total area of half), and are interspersed with filaments of the set $\{\rho(t)<0\}$ in such a manner that averages at any fixed scale become small (see Figure 1, below). Mathematically, this process is known as *weak convergence*. More precisely, we say $\rho(t)$ becomes mixed as $t\to\infty$ if $\rho(t)$ converges weakly to $\bar\rho=0$ in $L^2$. That is, for every $L^2$ test function $f$ we have

$$\lim_{t\to\infty}\int_{\mathbb{T}^2}\rho(t,x)f(x)\,\mathrm{d}\,x = \bar\rho\int_{\mathbb{T}^2} f(x)\,\mathrm{d}\,x = 0. \tag{1.2}$$

The standard notation for this convergence is to write

$$\rho(t)\xrightharpoonup[L^2]{t\to\infty} 0,$$

while we denote *strong $L^2$ convergence* by

$$\rho(t)\xrightarrow[L^2]{t\to\infty} 0,$$

which means $\|\rho(t)\|_{L^2}\to 0$ as $t\to\infty$. As the name suggests, strong convergence implies weak convergence, but the converse is generally false. In fact, in our situation where $u$ is divergence free, for all $t\in\mathbb{R}$ we have $\|\rho(t)\|_{L^2} = \|\rho_0\|_{L^2}$. Thus while many of our examples exhibit mixing (i.e., weak convergence of $\rho(t)$ to 0), they will not have strong convergence of $\rho(t)$ to 0, unless $\rho_0$ is identically 0. We remark that all $L^p$ norms are conserved by a volume-preserving map, but $L^2$ is a natural choice. Indeed, readers familiar with ergodic theory will recognize that weak convergence in $L^2$ is equivalent to the concept of strong mixing in dynamical systems.

Even though weak convergence is a natural way to study mixing, the disadvantage is that it does not *a priori* give a quantifiable rate. To explain further, if $\rho(t)$ converges to 0 strongly in $L^2$, then at time $t$ the quantity $\|\rho(t)\|_{L^2}$ is a measure of how close $\rho(t)$ is to its (strong) limit. If $\rho(t)$ converges to 0 *weakly*, then $\|\rho(t)\|_{L^2}$ may not contain any useful information about the convergence. (Indeed, for solutions to (1.1), $\|\rho(t)\|_{L^2}$ is independent of $t$.) It turns out, however, that in our situation, weak $L^2$ convergence to 0 is *equivalent* to strong convergence to 0 in any *negative Sobolev space*. Now the norm in these negative Sobolev spaces (which we will define shortly) can be used as a measure of how "well mixed" the distribution is (see for instance [Thi12]).

It is easiest to define the negative Sobolev norms using the Fourier series. Given an (integrable) function $\theta$ on the torus, we define its Fourier coefficients by

$$\hat\theta_k = \int_{\mathbb{T}^2}\theta(x)\,\mathrm{e}^{-2\pi i\langle x,k\rangle}\,\mathrm{d}\,x,\quad \text{where}\quad k\in\mathbb{Z}^2.$$

For mean-zero functions the 0-th Fourier coefficient, $\hat\theta_0$, vanishes. Now, for any $s\in\mathbb{R}$ we define the homogeneous Sobolev norm of index $s$ by

$$\|\theta\|_{\dot H^s}^2 \stackrel{\text{def}}{=} \sum_{k\in\mathbb{Z}^2,k\neq 0}|k|^{2s}|\hat\theta_k|^2.$$

Note that for $s>0$ the norm puts more weight on higher frequencies. Thus functions that have a smaller fraction of their Fourier mass in the high frequencies will be "less oscillatory" and have a smaller $\dot H^s$ norm. For $s<0$, however, the norm puts less weight on higher frequencies. Thus functions that have a larger fraction of their Fourier mass in the high frequencies will be "very oscillatory," and have a smaller $\dot{H}^s$ norm. This is consistent with what we expect from "mixed" distributions. Moreover, the result mentioned previously guarantees that mixing of $\rho$ is equivalent to

$$\|\rho(t)\|_{\dot{H}^s} \xrightarrow{t\to\infty} 0, \quad \text{for every } s < 0.$$

Thus, for any $s < 0$, the quantity $\|\rho(t)\|_{\dot{H}^s}$ can be used as a measure of how "mixed" the distribution is at time $t$.

For this reason negative Sobolev norms are often referred to as "mix-norms." Choosing $s = -1$ is particularly convenient, as the ratio of the $\dot{H}^{-1}$ norm to the $L^2$ norm scales like a length, and can be interpreted as a characteristic scale of the tracer field. Since the $L^2$ norm is preserved by equation (1.1), we can identify the $\dot{H}^{-1}$ norm with the *mixing scale* of the scalar field $\rho$. We mention that there is also a related notion of mixing scale, which is more geometric in nature (see for instance [Bre03]), but when studying evolution equations the mix-norms described above are easier to work with.

Practically, in order to mix a given initial configuration to a certain degree, one has to expend energy by "stirring the fluid." A natural, physically meaningful question is to bound the mixing efficiency [LTD11]. That is, given a certain "cost" associated with stirring the ambient fluid, what is the most efficient way to mix a given initial configuration? Two cost functions that are particularly interesting are the energy $\|u\|_{L^2}^2$ (which is proportional to the actual *kinetic energy* of the fluid, assuming the fluid is homogeneous), or the enstrophy $\|u\|_{\dot{H}^1}^2$ (which is proportional to the fluid's viscous energy dissipation rate). Hence our question about mixing efficiency can now be formulated as follows: *what are optimal bounds on $\|\rho(t)\|_{\dot{H}^{-1}}$ in terms of the fluid's energy or enstrophy?*

Before answering this question, we note that solutions to (1.1) can be given a classical meaning if the velocity field is Lipschitz. However, the natural constraints on the velocity $u$ described above do not require $u$ to be Lipschitz. Moreover, one intuitively expects efficient mixing flows to be "turbulent," and many established turbulence models have velocity fields that are only Hölder continuous at every point. Thus, in many natural situations arising in the study of fluids, one has to study (1.1) when the advecting velocity field is not Lipschitz. Seminal work of DiPerna and Lions in '89, and the important extension of Ambrosio in '04, addresses this situation, and shows that certain "renormalized" solutions to (1.1) are unique, provided $u, \nabla u \in L^1$.

Returning to the question of mixing efficiency formulated above, one can use direct energy estimates to show that if the fluid is *energy constrained* (i.e., if $\|u(t)\|_{L^2}^2 \leq E$, for some constant $E$), then $\|\rho(t)\|_{H^{-1}}$ can decrease at most linearly as a function of time. An elegant slice-and-dice construction of Bressan [Bre03], using piecewise-constant shear flows in orthogonal directions in a self-similar fashion, provides an example where this bound is indeed attained (see [LLN+12]). In particular, this provides an example where for some $T < \infty$ and an incompressible, finite-energy velocity field $u$, we have $\rho(t) \to 0$ weakly in $L^2$ as $t \to T$. That is, the fluid mixes the initial configuration "perfectly" in finite time.

On the other hand, if one imposes an *enstrophy constraint* (i.e., a restriction on the growth of $\|u(t)\|_{H^1}^2$), or more generally a restriction on the growth of $\|u\|_{L^p} + \|\nabla u(t)\|_{L^p}$, then the DiPerna-Lions theory guarantees finite-time perfect mixing cannot occur. Indeed, equation (1.1) is time reversible, and so finite-time perfect mixing would provide one nontrivial solution to (1.1) with initial data 0. Since $\rho \equiv 0$ is clearly another solution, we have nonuniqueness for weak solutions to (1.1), which is not allowed by the DiPerna-Lions theory when $u, \nabla u \in L^1$. So one can not have finite-time perfect mixing in this case.

Quantitatively, one can use the regularity of DiPerna-Lions flows [CDL08] to obtain explicit exponential lower bounds on the mix norm. Namely, one can prove [IKX14]

$$\|\rho(t)\|_{H^{-1}} \geq C_0 \exp\Big(-C_1 \int_0^t \|\nabla u(\tau)\|_{L^p} \, \mathrm{d}\tau\Big), \qquad (1.3)$$

for every $p \in (1, \infty]$, and some constants $C_0, C_1$ that depend on $\rho_0$ and $p$. (We remark that, for $p = \infty$, an elementary proof of the lower bound (1.3) follows from Gronwall's inequality and the method of characteristics. For $p \in (1, \infty)$, however, the proof is more involved and requires some tools from geometric measure theory.)

Interestingly, whether or not (1.3) holds for $p = 1$ is an open question. Indeed, the proof of the needed regularity estimates for the flow in [CDL08] relies on boundedness of a maximal function, which fails for $p = 1$. The bound (1.3) for $p = 1$ is related to a conjecture of Bressan [Bre03] on the cost of rearranging a set, which is still an open question.

For optimality, there are now several constructions of velocity fields that show (1.3) is sharp. These constructions produce enstrophy-constrained velocity fields for which $\|\rho(t)\|_{\dot{H}^{-1}}$ decays exponentially in time. A construction in [ACM19] does this by starting with initial data which is supported in a strip and finds a Lipschitz velocity field that pushes it along a space filling curve. Constructions in [BCZG23, BBPS21, MHSW22] produce regular velocity fields for which

$$\|\rho(t)\|_{\dot{H}^{-1}} \leq D\, \mathrm{e}^{-\gamma t} \|\rho_0\|_{\dot{H}^1}, \qquad (1.4)$$

for every initial data $\rho_0 \in \dot{H}^1$. Such flows are called *exponentially mixing*, and we revisit this in more detail in Section 2.

In addition to optimal mixing, there are three other themes discussed in this article. We briefly introduce these themes here, and elaborate on them in subsequent sections.

(A.) *Loss of regularity.* When $u$ is regular, classical theory guarantees regularity of the initial data $\rho_0$ is propagated by the equation (1.1). However, when $u$ is irregular (e.g., when $\nabla u \in L^p$ with $p < \infty$), it may happen that all regularity of the initial data $\rho_0$ is immediately lost. Not surprisingly, this loss is intrinsically related to mixing. Indeed, the process of mixing generates high frequencies, making the solution more irregular. When $u$ is not Lipschitz one can arrange rapid enough growth of high frequencies to ensure that *all* Sobolev regularity of the initial data is immediately lost. We describe this construction in detail in Section 3.

(B.) *Enhanced dissipation.* In several physically relevant situations, both diffusion and transport are simultaneously present. The nature of diffusion is to rapidly dampen high frequencies. Since mixing generates high frequencies, the combined effect of mixing and diffusion will lead to energy decay of solutions that is an order of magnitude faster than when diffusion acts alone. This phenomenon is known as *enhanced dissipation* and is described in Section 4.

(C.) *Anomalous dissipation.* Even under the enhanced dissipation mentioned above, for which the energy decay is much faster due to the combined effect of diffusion and transport, the energy decay rate vanishes with the diffusivity. In some sense, this outcome is expected, as solutions to (1.1) (formally) conserve energy. For certain (irregular) flows, however, it is possible for the energy decay rate in the presence of small diffusivity to stay uniformly positive, a phenomenon known as *anomalous dissipation*. It implies, in particular, that the vanishing-diffusivity limit can produce *dissipative solutions* of (1.1), that is, weak solutions for which the energy decreases with time. We discuss anomalous dissipation in Section 5.

## 2. Optimally Mixing Flows

In this section, our primary focus centers around understanding the concept of *shearing* as one of the central mechanisms of mixing, and how this mechanism gives rise to flows that mix optimally.

**2.1. Shear flows.** Shear flows are the simplest example of incompressible flows on $\mathbb{T}^2$. Their streamlines (lines tangent to the direction of the velocity vector) are parallel to each other and the velocity takes the form $u = (v(x_2), 0)$. The corresponding transport equation is

$$\partial_t \rho + v(x_2)\partial_{x_1}\rho = 0, \qquad \rho(0,x) = \rho_0(x), \tag{2.1}$$

the solution $\rho(t, x_1, x_2) = \rho_0(x_1 - v(x_2)t, x_2)$ of which can be computed explicitly via the method of characteristics.

If the initial datum only depends on $x_2$, then the solution remains constant for all times. Otherwise, a hint of creation of small scales is given by the growth of $\|\partial_{x_2}\rho\|_{L^2}$ linearly in time. To deduce a quantitative mixing estimate, one can take a partial Fourier transform in $x_1$ of (2.1): denoting $\hat\rho(t,k,x_2)$, with $k \in \mathbb{Z}$, the Fourier coefficients of $\rho$, (2.1) becomes

$$\partial_t \hat\rho + 2\pi i k v(x_2)\hat\rho = 0, \quad \hat\rho(0,k,x_2) = \hat\rho_0(k,x_2). \tag{2.2}$$

Since $\hat\rho(t,k,x_2) = \mathrm{e}^{-2\pi i k v(x_2)t}\,\hat\rho_0(k,x_2)$, mixing follows by estimating oscillatory integrals of the form

$$\int_{\mathbb{T}} \mathrm{e}^{-2\pi i k v(x_2)t}\,\hat\rho_0(k,x_2)\hat\phi(k,x_2)\,\mathrm{d}\,x_2, \qquad \phi \in \dot H^1(\mathbb{T}^2).$$

A duality argument and an application of the stationary-phase lemma entails a (sharp) mixing estimate.

**Theorem 2.1.** *Assume that $v \in C^m(\mathbb{T})$, for some integer $m \geq 2$, and its derivatives up to order $m$ do not vanish simultaneously: $|v'(x_2)| + |v''(x_2)| + \cdots + |v^{(m)}(x_2)| > 0$, for all $x_2 \in \mathbb{T}$. Then there exists a positive constant $C = C(v)$ such that*

$$\|\rho(t)\|_{\dot H^{-1}} \leqslant \frac{C}{t^{1/m}}\|\rho_0\|_{\dot H^1}, \qquad \forall t \geq 1,$$

*for all initial data $\rho_0 \in \dot H^1(\mathbb{T}^2)$ with vanishing $x_1$-average.*

The mixing rate is solely determined by how degenerate the critical points of $v$ are, and Theorem 2.1 tells us that the flatter the critical points, the slower the (universal) mixing rate. In the case of simple or nondegenerate critical points, for which the second derivative does not vanish, such as for the Kolmogorov flow $v(x_2) = \sin(2\pi x_2)$, we have $m = 2$.

We observe that if $k = 0$ in (2.2) then $\hat\rho(t,0,x_2)$ is constant in time; hence unless $\hat\rho(0,0,x_2) = 0$, no mixing can occur. Therefore, it is essential to assume that the initial condition $\rho_0$ has vanishing average in the $x_1$ variable, which is equivalent to imposing that $\hat\rho_0(0,x_2) = 0$. This requirement excludes functions that are constant on streamlines, i.e., eigenfunctions (with eigenvalue 0) of the transport operator, which do not enjoy any mixing.

**2.2. General two-dimensional flows.** Although regular shear flows can achieve algebraic mixing rates, we could be inclined to think that their simple structure constitutes an obstruction to faster mixing. It turns out that if $H$ is an autonomous, nonconstant Hamiltonian function on $\mathbb{T}^2$, of class $C^2$, generating an incompressible velocity field $u = \nabla^\perp H = (-\partial_{x_2}H, \partial_{x_1}H)$, then the mixing rate of $u$ is at best $1/t$, and can be even slower depending on the structure of $H$, see [BCZM22]. This result can be interpreted as follows: despite the fact that $H$ could have hyperbolic points, at which the flow map displays exponential stretching and compression, shearing is the main mixing mechanism in 2d. This can be deduced from the existence of an invariant domain for the Lagrangian flow on which $u$ is bounded away from zero, which in turn implies that there

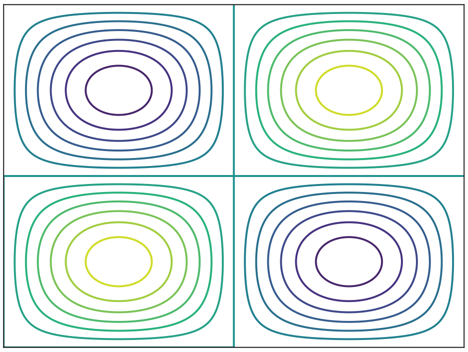


**Figure 2.** Level sets of the Hamiltonian $H(x_1, x_2) = \sin(2\pi x_1)\sin(2\pi x_2)$.

exists a well-defined, regular and invertible change of coordinates $(x_1, x_2) \mapsto (h, \theta) \in \mathbb{T} \times (h_0, h_1)$, where the interval $(h_0, h_1) \subset \mathrm{range}(H)$ is determined by the invariant set. An example is a simple *cellular flow*, where $H(x_1, x_2) = \sin(2\pi x_1)\sin(2\pi x_2)$. The level sets of the Hamiltonian $H$ are given in Figure 2, and clearly show shearing away from the separatrices.

**Proposition 2.2.** *Let $H \in C^2(\mathbb{T}^2)$. There exists an invariant open set $\mathcal{I} \subset \mathbb{T}^2$ such that for any $\rho_0 \in C^1(\mathbb{T}^2)$ with* $\mathrm{supp}(\rho_0) \subset \mathcal{I}$, *the corresponding solution $\rho$ of* (1.1) *satisfies*

$$\|\rho(t)\|_{\dot{H}^1} \leqslant C(1+t)\|\nabla \rho_0\|_{L^\infty}, \tag{2.3}$$

*for some $C = C(\mathcal{I}, H)$ and all $t \geq 0$.*

The set of coordinates $(h, \theta)$ are the so-called action-angle coordinates, and reduce the transport operator $u \cdot \nabla$ to the much simpler from $\frac{1}{T(h)}\partial_\theta$, where $T(h)$, which is a $C^1$ function, is the period of the closed orbit $\{H(x_1, x_2) = h\}$. The analogy with (2.1) is then apparent, and estimate (2.3) is derived from the explicit solution, obtained via the method of characteristics. Thanks to interpolation, the growth (2.3) is a lower bound on the mix-norm of $\rho$, hence proving that $1/t$ is a lower bound on the mixing rate of $u$.

**2.3. Exponentially mixing flows.** Obtaining a faster mixing rate necessarily involves nonautonomous velocity fields. A widely used exponential mixer, especially in numerical simulations, is due to Pierrehumbert [Pie94], and consists of randomly alternating shear flows on $\mathbb{T}^2$. The beauty of this example is its simplicity: at discrete time steps $t_n$, it alternates the horizontal shear $(\sin(2\pi(x_1 - \omega_{1,n})), 0)$ and the vertical shear $(0, \sin(2\pi(x_2 - \omega_{2,n})))$. Here, $\omega = \{\omega_{1,n}, \omega_{2,n}\}$ is a sequence of independent uniformly distributed random variables so the phases are randomly shifted. While widely believed to be exponentially mixing, the first proof of this fact appeared only recently in [BCZG23].

**Theorem 2.3.** *There exists a random constant $D$ (with good bounds on its moments) and $\gamma > 0$ such that we have* (1.4), *almost surely.*

By taking a realization of the above velocity field, this result settles the question of the existence of a smooth exponential mixer on $\mathbb{T}^2$, although it does not produce a time-periodic velocity field.

The proof of Theorem 2.3 relies on tools from random dynamical system theory and adopts a Lagrangian approach to the problem: this approach involves proving the positivity of the top Lyapunov exponent of the flow map via Furstenberg's criterion and a Harris theorem.

A related example has been produced in [MHSW22, ELM23], constructed by alternating two piecewise linear shear flows. This example is fully deterministic and produces a time-periodic, Lipschitz velocity field. The important feature of this flow is that it generates a uniformly hyperbolic map on $\mathbb{T}^2$.

In general, constructing exponentially mixing flows on $\mathbb{T}^2$ has proven to be quite a challenge, and only recently there have been tremendous developments. Besides the two works described above, that constitute the latest works in the field, we mention the deterministic constructions of [ACM19, EZ19], the latter building upon prior results of Yao and Zlatos, and the beautiful work on velocity fields generated by stochastically forced Navier-Stokes equations of [BBPS21].

## 3. Loss of Regularity

One of the effects of mixing is the creation of striation in the scalar field. Quantitatively, this effect corresponds to growth of derivatives of $\rho$, which can also be seen from the well-known interpolation inequality:

$$\|\rho(t)\|_{L^2}^2 \leq \|\rho(t)\|_{\dot{H}^{-s}}\|\rho(t)\|_{\dot{H}^s}. \tag{3.1}$$

In fact, recalling that the $L^2$ norm of $\rho$ is conserved by the flow of $u$, if the negative Sobolev norms of $\rho(t)$ decay to zero at some time $T \leq \infty$, at that same time the positive Sobolev norms must blow up. However, we note that growth of derivatives is a local phenomenon that can occur in the absence of mixing, which is a global phenomenon.

One can ask whether the growth of Sobolev norms can lead to loss of regularity for solutions of the transport equation (1.1), when the velocity field is not sufficiently smooth. The Cauchy-Lipschitz theory implies that, if $u$ is Lipschitz uniformly in time, then the flow of $u$ is also Lipschitz continuous, although its Lipschitz constant can grow exponentially fast in time. Hence, at least some regularity of the initial data $\rho_0$ is preserved in time. When the gradient of $u$ is not bounded, but it is still in some $L^p$ space with $p < \infty$, the direct estimates from the Cauchy-Lipschitz theory do not apply. Therefore, it is natural to investigate what, if any, regularity of the initial data $\rho_0$ is preserved under advection by $u$.

We present two examples to show that no Sobolev regularity is preserved in general: the first where the flow is mixing and all Sobolev regularity, including fractional regularity, is lost instantaneously; the second where the flow is not mixing and we are able to show at least that the $H^1$ (and any higher) norm blows up instantaneously. The second construction applies to (almost) all initial conditions in $H^1$, though the resulting velocity field $u$ still depends on the initial condition $\rho_0$. In both examples, the simple key idea is to utilize the linearity of the transport equation to construct a weak solutions by adding infinitely many suitably rescaled copies of a base flow and a base solution. The rescaling pushes energy to higher and higher frequencies or small scales, leading to an accelerated growth of the derivatives, which ultimately results in an instantaneous blow-up (see [CEIM22] and references therein).

The first result is the following:

**Theorem 3.1.** *There exists a bounded velocity field $u$ such that $\nabla u(t) \in L^p(\mathbb{R}^2)$, $1 \leq p < \infty$, uniformly in time and a smooth, compactly supported function $\rho_0 \in C_c^\infty(\mathbb{R}^2)$, such that both $u$ and the unique bounded weak solution $\rho$ with initial data $\rho_0$ are compactly supported in space and smooth outside a point in $\mathbb{R}^2$, but $\rho(t)$ does not belong to $\dot{H}^s(\mathbb{R}^2)$ for any $s > 0$ and $t > 0$.*

This result implies lack of continuity of the flow map in Sobolev spaces and can be shown to be a generic phenomenon in the sense of Baire's Category Theorem.

We sketch the proof of Theorem 3.1. Utilizing a suitable exponentially mixing flow on the torus, it is possible to construct a smooth, bounded, divergence-free vector field $u^0$ with $\nabla u^0(t) \in L^p(\mathbb{R}^2)$, $1 \leq p < \infty$, uniformly in time and a smooth solution $\rho^0$ of the transport equation with velocity field $u^{(0)}$, both supported on the unit square $\mathcal{Q}_0$ in the plane, such that all positive Sobolev norms $\|\rho^0(t)\|_{\dot{H}^s}$, $s > 0$, grow exponentially fast in time. For each $n \in \mathbb{N}$, we define velocity fields $u^{(n)}$ and functions $\rho^{(n)}$ on squares $\mathcal{Q}_n$ of sidelength $\lambda_n$ by rescaling:

$$
\begin{aligned}
u^{(n)}(t,x) &= \frac{\lambda_n}{\tau_n} u^{(0)}\Big(\frac{t}{\tau_n}, \frac{x}{\lambda_n}\Big), \\
\rho^{(n)}(t,x) &= \gamma_n \rho^{(0)}\Big(\frac{t}{\tau_n}, \frac{x}{\lambda_n}\Big),
\end{aligned} \tag{3.2}
$$

for some sequences $\lambda_n$, $\tau_n$, and $\gamma_n$ to be chosen, up to some rigid motions, which do not change the norms and which we suppress for ease of notation. The squares $\mathcal{Q}_n$ can be taken pairwise disjoint. Then, by setting

$$
u \stackrel{\text{def}}{=} \sum_n u^{(n)}, \qquad \rho \stackrel{\text{def}}{=} \sum_n \rho^{(n)},
$$

we have that $\rho$ is a weak solution of (1.1) with velocity field $u$. Lastly, we pick $\lambda_n$, $\tau_n$, and $\gamma_n$ in such a way that the squares $\mathcal{Q}_n$ converge to a point, the only point where $u$ and $\rho$ are not smooth, the norms $\|u(t)\|_{\dot{W}^{1,p}}$ and $\|\rho(0)\|_{\dot{H}^s}$ are controlled, while the norm $\|\rho(t)\|_{\dot{H}^s} = \infty$.

The second result is the following:

**Theorem 3.2.** *Given any nonconstant function $\rho_0 \in H^1(\mathbb{R}^2)$, there exists a bounded, compactly supported, divergence-free velocity field $u$, with $\nabla u(t) \in L^p(\mathbb{R}^2)$ for any $1 \leq p < \infty$ uniformly in time, and smooth outside a point in $\mathbb{R}^2$, such that the unique weak solution $\rho(t)$ of (1.1) in $L^2(\mathbb{R}^2)$ with initial data $\rho_0$ does not belong to $\dot{H}^1(\mathbb{R}^2)$ (even locally) for any $t > 0$.*

In fact, a stronger statement is true. The velocity field $u$ is in all Sobolev spaces that are not embedded in the Lipschitz space (namely, $u \in W^{r,p}$ for all $r < 2/p + 1$, $1 \leq p < \infty$).

The main steps in the proof of Theorem 3.2 are as follows. The first step follows by a direct calculation on functions on the torus $\mathbb{T}^2$. Given any nonconstant periodic function $\bar{\phi}$, applying either a sine or cosine shear flow parallel to one of the coordinate axes must increase the $\dot{H}^1$ norm of $\bar{\phi}$ by a constant factor $A > 1$ at time $t = 1$. Hence, at time $t = n$, where $n$ is a positive integer, the $\dot{H}^1$-norm of $\bar{\phi}$ has grown by a factor $A^n$, which implies that this norm grows exponentially in time. By unfolding the action of the shear flows on the torus, this observation can be adapted to showing exponential growth of the $\dot{H}^1$-norm of functions supported on a square (let's say again the unit square) in $\mathbb{R}^2$ by a combination of sine and cosines shear flows. Next, rescaling the flow alone in a manner similar to (3.2) gives a sequence of well-separated squares, shrinking to a point, such that the rescaled flow grows the $\dot{H}^1$-norm of functions supported on each square by a larger and larger factor at time 1. The final step consists in choosing the location of the squares and the rescaling factor in such a way that the $\dot{H}^1$ norm of the solution diverges at any positive time, but the velocity field remains sufficiently regular.

In view of the results above, one can ask if any regularity of $\rho_0$, measured by a norm that is not comparable with the Sobolev norm $\dot{H}^s$, is preserved under the advection by $u$. It was shown by Bruè and Nguyen that essentially only the "logarithm" of a derivative is preserved. In this sense, the loss of regularity in Theorem 3.1 can be viewed as optimal.

## 4. Enhanced Dissipation

We now turn our attention to problems where both diffusion and convection are present, and study the combined effect of both. A prototypical example is the evolution of the concentration of a solute in an ambient fluid (e.g., cream in coffee). The evolution of the (normalized) concentration is modelled by the advection-diffusion equation

$$
\partial_t \rho^\kappa + u \cdot \nabla \rho^\kappa - \kappa \Delta \rho^\kappa = 0. \tag{4.1}
$$

Here $\rho^\kappa$ denotes the deviation of the concentration of the solute from its spatial average, the quantity $\kappa > 0$ is the molecular diffusivity, and $u$ denotes the velocity field of the ambient fluid. As in the previous sections, we will impose periodic boundary conditions on (4.1) and restrict our attention to the incompressible setting where $u$ is divergence free. In this case the spatial average is still preserved by (4.1), so we may, without loss of generality, uphold our convention that $\rho^\kappa$ is spatially mean zero.

Multiplying equation (4.1) by $\rho^\kappa$ and integrating by parts gives the following identity:

$$\|\rho^\kappa(t)\|_{L^2}^2 - \|\rho^\kappa(0)\|_{L^2}^2 = -2\kappa \int_0^t \|\nabla\rho^\kappa(\tau)\|_{L^2}^2 \, \mathrm{d}\tau \quad \text{for any } t \in [0,T] \tag{4.2}$$

which expresses the fact that any loss in the $L^2$ norm of the scalar $\rho^\kappa$, called the energy by analogy with the kinetic energy in fluids, is balanced by the dissipation due to diffusion. Furthermore, using incompressibility and Poincaré's inequality, one also has that

$$\|\rho^\kappa(t)\|_{L^2} \le \mathrm{e}^{-4\pi^2\kappa t}\|\rho_0\|_{L^2}. \tag{4.3}$$

This estimate, however, is completely blind to the effect of the fluid advection, and in practice one expects $\|\rho^\kappa\|_{L^2}$ to decay much faster than the rate predicted by (4.3). The reason for this is a phenomenon most people have likely observed themselves when stirring cream into coffee: the fluid flow initially spreads the cream into fine filaments; diffusion acts faster on fine filaments, and so these uniformize very quickly.

### 4.1. Quantifying dissipation enhancement.

One way to mathematically quantify and study this phenomenon is through the *dissipation time*, denoted by $t_{\mathrm{dis}}$. Explicitly define $t_{\mathrm{dis}} = t_{\mathrm{dis}}(u,\kappa)$ to be the smallest time $t \ge 0$ so that

$$\|\rho^\kappa(t_0+t)\|_{L^2} \le \frac{1}{2}\|\rho^\kappa(t_0)\|_{L^2},$$

for every time $t_0 \ge 0$ and mean-zero initial data $\rho^\kappa(s) \in L^2$. Clearly (4.3) shows $t_{\mathrm{dis}} \le 1/(4\pi^2\kappa)$, and one can precisely define *enhanced dissipation* as situations where $t_{\mathrm{dis}} \ll 1/(4\pi^2\kappa)$. We now list several situations where enhanced dissipation is exhibited.

*Shear flows.* If $u$ is a shear flow with a $C^2$ profile that has nondegenerate critical points, then classical work of Kelvin shows $t_{\mathrm{dis}} \le C/\sqrt{\kappa}$. A matching lower bound was also proved by Coti Zelati and Drivas.

*Cellular flows.* A cellular flow models the movement of a 2D fluid in the presence of a strong array of opposing vortices. The simplest example was already introduced in Section 2. Here, the flow is rescaled to generate stronger vortices at smaller scales, and it is given by the formulas:

$$u = \begin{pmatrix} -\partial_{x_2}H \\ \partial_{x_1}H \end{pmatrix}, \quad H = A\epsilon \sin\Big(\frac{2\pi x_1}{\epsilon}\Big)\sin\Big(\frac{2\pi x_2}{\epsilon}\Big),$$

where $A \gg 1$ is the flow amplitude and $\epsilon \ll 1$ is the cell size. The enhancement of diffusion by the presence of a periodic flow has been extensively studied. We mention in particular the early work of Childress and then the work of Fannjiang and Papanicolau. Standard homogenization results show that as $\epsilon \to 0$, the operator $-u\cdot\nabla + \kappa\Delta$ behaves like $D_{\mathrm{eff}}\Delta$, where

$$D_{\mathrm{eff}} \approx C\sqrt{\frac{\kappa A}{\epsilon}}$$

is the effective diffusivity. As a result one would expect

$$t_{\mathrm{dis}} = O\Big(\frac{1}{D_{\mathrm{eff}}}\Big) = O\Big(\sqrt{\frac{\epsilon}{\kappa A}}\Big)$$

as $\kappa, \epsilon \to 0$, $A \to \infty$. This is indeed the case (provided $\kappa/\epsilon \ll A \ll \kappa/\epsilon^3$), and was proved recently by Iyer and Zhou. A matching lower bound for $t_{\mathrm{dis}}$ was recently proved in [BCZM22].

*Relaxation enhancing flows.* The seminal work of Constantin et al. [CKRZ08] shows that for time independent velocity fields, $t_{\mathrm{dis}} = o(1/\kappa)$ if and only if the operator $u\cdot\nabla$ has no eigenfunctions in $H^1$. Such flows are called *relaxation enhancing*. It is known that *weakly mixing* flows are relaxation enhancing, but relaxation enhancing flows need not be weakly mixing.[1]

*Mixing flows.* Thus far, the examples provided only reduce the dissipation time to an algebraic power $1/\kappa^\alpha$ for some $\alpha < 1$. Using *exponentially mixing* flows, it is possible to reduce the dissipation time further to $|\ln\kappa|^2$ (see [FI19, CZDE20]). In fact, we recall that a velocity field $u$ is *exponentially mixing* if, for any $\rho_0 \in \dot{H}^1$, the mix norm of solutions to (1.1) decays exponentially as in (1.4). Informally, then any solution to (1.1) with initial data that is localized to a ball of radius $\epsilon$ becomes essentially uniformly spread in time $O(|\ln\epsilon|)$. If $u$ is exponentially mixing, one can show that

$$t_{\mathrm{dis}} \le C|\ln\kappa|^2.$$

An elementary heuristic argument, however, suggests we should have the stronger bound

$$t_{\mathrm{dis}} \le C|\ln\kappa|. \tag{4.4}$$

Indeed, if the solute is initially concentrated at one point $x$, then after time $O(1)$ it will spread to a ball of radius $O(\sqrt{\kappa})$. Now, since $u$ is exponentially mixing, it will get spread almost uniformly on the entire domain in time $O(|\ln\kappa|)$, if the effect of diffusion is negligible. Unfortunately, the effects of diffusion may not be negligible on the time scales of order $O(|\ln\kappa|)$, and so this argument cannot be easily made rigorous.

Even though proving (4.4) for general exponentially mixing flows is still an open question, there are several examples of flows for which (4.4) is known: for instance, when $u$ is the velocity field from the stochastically

[1] Weakly mixing *is a notion from dynamical systems which requires* (1.2) *to hold in a time-averaged sense.*

forced Navier-Stokes equations (see [BBPS21] and references therein), or when $u$ consists of alternating horizontal/vertical shears with a tent profile and a sufficiently large amplitude [ELM23]. It is also known that $t_{\text{dis}}$ cannot be smaller than $O(|\ln \kappa|)$ for velocity fields that are Lipschitz in space uniformly in time.

4.2. **Blow up suppression.** One application of enhanced dissipation is to control certain nonlinear phenomena. For concreteness and simplicity, we focus our attention on a simplified version of the Keller-Segel system of equations, which is used to model the evolution of the population density of micro-organisms when chemotactic effects are present. We again impose periodic boundary conditions and study this system on two-dimensional tori. If $n = n(t, x) \geq 0$ represents the bacterial population density, $c \geq 0$ represents the concentration of a chemoattractant produced by the bacteria, and $\chi > 0$ is a sensitivity parameter, then a simplified version of the Keller-Segel model is the following system:

$$\begin{aligned} \partial_t n - \Delta n &= -\nabla \cdot (n\chi \nabla c), \qquad (4.5) \\ -\Delta c &= n - \bar{n}, \\ \bar{n} &= \int_{\mathbb{T}^2} n \, \mathrm{d}x. \end{aligned}$$

This model stipulates that bacterial diffusion is biased in the direction of the gradient of the concentration of a chemoattractant that is emitted by the bacteria themselves, and that the chemoattractant diffuses much faster than the bacteria do.

From the equation we see that there is a competition between two effects: The diffusive term $\Delta n$ drives bacteria away from regions of high population, and the chemotactic term $\nabla \cdot (n\chi\nabla c)$ drives the bacteria towards it. If the chemotactic effects dominate, they will lead to a population explosion. This has been well studied and it is now known that the diffusive effects dominate (and there is no population explosion) if and only if the total initial population is below a certain threshold.

One natural question is to study the effect of movement of the ambient fluid on this system. If the ambient fluid has velocity field $u$, equation (4.5) becomes

$$\partial_t n + u \cdot \nabla n - \Delta n = -\nabla \cdot (n\chi\nabla c). \qquad (4.5')$$

Intuitively, we expect that regions of high concentration of bacteria can be dispersed by vigorous stirring. This result can be established rigorously.

**Theorem 4.1.** *There exists $t_* = t_*(\|n_0\|_{L^2})$ such that if*

$$t_{\text{dis}}(u, 1) < t_*,$$

*then there is no population explosion in* (4.5′).

The main idea being the proof can be explained in an elementary fashion and we now provide a quick sketch of the proof of Theorem 4.1.

First, we rewrite (4.5′) as

$$\partial_t \theta + u \cdot \nabla\theta - \Delta\theta = \mathcal{N}(\theta), \qquad (4.6)$$

where $\theta = n - \bar{n}$ and

$$\mathcal{N}(\theta) = -\nabla \cdot ((\theta + \bar{n})\chi\nabla c).$$

Multiplying (4.6) by $\theta$ and integrating in space show that

$$\frac{1}{2}\partial_t \|\theta(t)\|_{L^2}^2 + \|\nabla\theta\|_{L^2}^2 \leq \int_{\mathbb{T}^d} \theta\mathcal{N}(\theta) \, \mathrm{d}x.$$

Now using standard energy estimates one can show that when the diffusive term $\|\nabla\theta\|_{L^2}^2$ is large, then $\partial_t \|\theta\|_{L^2} \leq 0$. More precisely, one can show there exists a constant $C_1 = C_1(\chi, d, \|\theta_0\|_{L^2})$ such that if

$$\frac{1}{t_*}\int_0^{t_*} \|\nabla\theta\|_{L^2}^2 \, \mathrm{d}t \geq C_1, \qquad (4.7)$$

then we must also have $\|\theta(t_*)\|_{L^2} \leq \|\theta_0\|_{L^2}$.

Suppose now (4.7) does not hold. In this case we use Duhamel's formula to write

$$\theta(t_*) = \mathcal{S}_{0,t_*}\theta_0 + \int_0^{t_*} S_{\tau,t_*}\mathcal{N}(\theta(\tau)) \, \mathrm{d}\tau,$$

where $\mathcal{S}_{s,t}$ is the solution operator to (4.1). This implies

$$\|\theta(t_*)\|_{L^2}^2 \leq \|\mathcal{S}_{0,t_*}\theta_0\|_{L^2}^2 + \int_0^{t_*} \|\mathcal{N}(\theta(\tau))\|_{L^2} \, \mathrm{d}\tau.$$

We notice that, since $t_{\text{dis}} \leq t_*$, the first term on the right is at most $\|\theta_0\|_{L^2}/2$. For the second term we use standard energy estimates to control $\|\mathcal{N}(\theta)\|_{L^2}$ by $\|\nabla\theta\|_{L^2}^2$ and $\|\theta\|_{L^2}$. Combined with the assumption that (4.7) does not hold, we obtain an inequality of the form

$$\|\theta(t_*)\|_{L^2}^2 \leq \Big(\frac{1}{2} + t_* F(C_1)\Big)\|\theta_0\|_{L^2}^2,$$

for some explicit function $F$ that arises from the bound on the nonlinearity. If $t_* \leq 1/(2F(C_1))$ then the right-hand side is at most $\|\theta_0\|_{L^2}$. Iterating this step, one immediately sees that $\sup_{t<\infty}\|\theta(t)\|_{L^2} < \infty$. This bound is enough to show that there is no population explosion in (4.5′), concluding the proof. A more complete version of this proof can be found in [IXZ21]. It has also be extended to fourth order equations, such as the Kuramoto-Sivashinsky equation, a model of flame-front propagation, by Feng and Mazzucato.

## 5. Anomalous Dissipation

In the previous section we saw several examples of *enhanced dissipation*, where the solution to (4.1) loses a constant fraction of its $L^2$ energy in time scales much smaller than the dissipative time scale $1/\kappa$. The examples outlined exhibited the energy loss on time scales $1/\kappa^\alpha$, $|\ln \kappa|^2$ or $|\ln \kappa|$, which diverge to infinity in the vanishing diffusivity limit $\kappa \to 0$. A natural question to ask is whether there are situations where solutions to (4.1) lose a constant fraction of their $L^2$ energy on a time scale that is $O(1)$ as $\kappa \to 0$. This phenomenon is called *anomalous dissipation*. More precisely, anomalous dissipation is the existence of solutions of (4.1) with progressively smaller diffusivity $\kappa$ that converge to a *dissipative* solution of the transport equation in the limit $\kappa \to 0$. That is, as $\kappa \to 0$, we can find solutions $\rho^\kappa$ to (4.1), which converge (possibly along a subsequence) to a weak solution $\rho^0$ of (1.1), where

$$\|\rho^0(T)\|_{L^2}^2 < \|\rho^0(0)\|_{L^2}^2,$$

for some time $T < \infty$.

The dissipation of the $L^2$ energy of the solution $\rho^\kappa$ to (4.1) for time $t \in [0, T]$ is encoded in the energy identity (4.2). In the limit case $\kappa = 0$, (4.2) expresses (formally) the conservation of the $L^2$ norm for solutions of the transport equation (1.1). The velocity field does not appear explicitly in (4.2). However, the action of a mixing velocity field results in filamentation of the scalar and consequently in the creation of large gradients, thus conceivably allowing for scenarios in which the right-hand side of (4.2) remains bounded away from zero even in the limit $\kappa \to 0$ of vanishing diffusivity.

This phenomenon is the analogue in the linear case of the so-called 0-th law of turbulence of the Onsager-Kolmogorov theory of turbulence for the Euler/Navier-Stokes equations. The 0-th law predicts uniform-in-viscosity dissipation of the kinetic energy, due to the nonlinear transfer of energy to high frequencies and to the corresponding enhanced effect of the diffusion. The 0-th law usually applies to stochastically forced flows and the dissipation is the average over many realizations of the flow. Our setting is akin to so-called decaying turbulence instead, in the absence of forcing.

In order to identify the critical regularity for anomalous dissipation in solutions to (4.1) as $\kappa$ vanishes, heuristically at least, we can formally rewrite the contribution of the advection term in the energy estimate as

$$\int_{\mathbb{T}^2} (u \cdot \nabla \rho)\rho \, \mathrm{d}\, x \sim \int_{\mathbb{T}^2} \nabla^\alpha u \left(\nabla^{\frac{1-\alpha}{2}} \rho\right)^2 \, \mathrm{d}\, x,$$

for any $0 \leq \alpha \leq 1$, where $\rho$ denotes both $\rho^\kappa$ and $\rho^0$. The fractional derivatives $\nabla^\alpha$ can be estimated via norms in Hölder's spaces $C^\alpha$. Criticality is therefore expressed by the so-called Yaglom's relation: focusing for simplicity on regularity in space only, for $u \in C^\alpha$ and $\rho \in C^\beta$, the combined Hölder's regularity of the velocity field and the solution is:

- subcritical, if $\alpha + 2\beta > 1$,
- critical, if $\alpha + 2\beta = 1$,
- supercritical, if $\alpha + 2\beta < 1$.

The critical threshold for the exponents $\alpha$ and $\beta$ above is the analogue for linear advection-diffusion equations of the critical $\frac{1}{3}$-Hölder's regularity in the case of anomalous energy dissipation for solutions of the Navier-Stokes equations as viscosity vanishes, according to the Onsager-Kolmogorov theory of turbulence. In fact, setting $u = \rho$ and so $\alpha = \beta$ gives precisely $\alpha = 1/3$.

The Obukhov-Corrsin theory of scalar turbulence (1949–1951) predicts that:

- in the subcritical regime, for a given $u \in C^\alpha$ there exists a unique solution $\rho^0 \in C^\beta$ of (1.1) and such a solution conserves the $L^2$ norm,
- in the supercritical regime, there exist velocity fields $u \in C^\alpha$ such that nonuniqueness and dissipation of the $L^2$ norm are possible for solutions $\rho^0 \in C^\beta$ of (1.1); moreover, anomalous dissipation is possible, in the sense that

$$\limsup_{\kappa \to 0} \, \kappa \int_0^T \|\nabla \rho^\kappa(\tau)\|_{L^2}^2 \, \mathrm{d}\, \tau > 0 \tag{5.1}$$

for solutions $\rho^\kappa$ of (4.1) uniformly bounded in $C^\beta$.

The statement in the subcritical case can be proven using a commutator argument similar to that of Constantin, E, and Titi for the Euler equations, see for instance [DEIJ22, Theorem 4]. We stress that the uniqueness statement strongly relies on the linearity of the equation. Addressing the supercritical case is more challenging and has been done only very recently from a rigorous mathematical perspective. We briefly discuss these recent results in the next subsection.

**5.1. Anomalous dissipation for bounded solutions.** The endpoint case $\alpha < 1$ and $\beta = 0$ has been addressed in [DEIJ22]. In this paper, for any $\alpha < 1$, the authors provide an example of a bounded velocity field, which belongs to $L^1([0, T]; C^\alpha(\mathbb{T}^2))$ and is smooth except at the singular time $t = T$, that exhibits anomalous dissipation for all initial data sufficiently close to a (nontrivial) harmonics. They are also able to construct velocity fields that exhibit anomalous dissipation for any (regular enough) initial datum, although the velocity field depends on the chosen initial datum. The strategy for both examples is to construct a velocity field which develops smaller and smaller scales when the time approaches the singular time $t = T$. This construction can be interpreted as mimicking the development in time of a turbulent cascade. However, it also causes the anomalous dissipation to be concentrated at the singular time $t = T$, in the sense that for any $\varepsilon > 0$ it holds

$$\lim_{\kappa \to 0} \, \kappa \int_0^{T-\varepsilon} \|\nabla \rho^\kappa(\tau)\|_{L^2}^2 \, \mathrm{d}\, \tau = 0.$$

Several criteria that imply anomalous dissipation are given in [DEIJ22]. The criterion that provides the most intuition on the mechanism for anomalous dissipation (even though such a criterion is not the one effectively exploited in the proof in [DEIJ22]) establishes a link to mixing in solutions of the transport equation (1.1) and asserts that, if the solution $\rho$ of (1.1) satisfies

$$\int_0^T \|\nabla\rho(\tau)\|_{L^2}^2 \,\mathrm{d}\tau = \infty$$
$$\text{and} \quad \|\rho(t)\|_{\dot H^{-1}}\|\rho(t)\|_{\dot H^1} \le C\|\rho(t)\|_{L^2}^2, \quad (5.2)$$

then anomalous dissipation holds. In view of the interpolation inequality (3.1) with $s = 1$, the second condition in (5.2) in particular implies that, in the absence of diffusion, the velocity field mixes essentially at the optimal rate. However, it is not easy to produce velocity fields with such strong mixing properties as in (5.2), and this is the reason why the proof of [DEIJ22] needs to rely on weaker criteria. In fact, the velocity field is a self-similar version of the alternating shear flows example by Pierrehumbert [Pie94]. Although it enjoys weaker mixing properties, the velocity field constructed in [DEIJ22] exhibits anomalous dissipation due to the following heuristic reason: mixing requires all energy to be sent to high frequencies, while anomalous dissipation just requires a given fraction of the energy to be sent to high frequencies.

As a corollary, the construction in [DEIJ22] provides a new example of nonuniqueness for the transport equation (1.1) with a velocity field in $L^1([0,T];C^\alpha(\mathbb{T}^2))$ (for $\alpha < 1$), but outside the DiPerna-Lions class. In order to see this fact, we extend the velocity field for time $t \in [T,2T]$ by reflecting it oddly across $t = T$, that is, setting $u(t) = -u(2T-t)$. We can then see that the following are two distinct weak solutions:

- the vanishing-diffusivity solution $\rho^{\mathrm{vd}}(t)$ for time $t \in [0,2T]$, and
- the solution given by $\rho^{\mathrm{refl}}(t) = \rho^{\mathrm{vd}}(t)$ for time $t \in [0,T]$, and by the odd reflection $\rho^{\mathrm{refl}}(t) = \rho^{\mathrm{vd}}(2T-t)$ for time $t \in [T,2T]$.

Indeed, they are distinct for $t \in [T,2T]$, since $\rho^{\mathrm{vd}}$ dissipates the $L^2$ norm while $\rho^{\mathrm{refl}}$ conserves it. It must be noted that, based on the approach in [DEIJ22], it is unclear whether $\rho^{\mathrm{refl}}$ can be constructed in the limit of vanishing diffusivity along a suitable subsequence $\kappa_n \to 0$.

**5.2. Anomalous dissipation and lack of selection.** The possibility of having two distinct solutions, both arising in the limit of vanishing diffusivity, puts into question the validity of the zero-diffusivity limit as a selection principle for weak solutions of the transport equation outside of the DiPerna-Lions theory. Solutions arising as zero-diffusivity limit may be considered "more physical" than general weak solutions and one may wonder whether uniqueness could be restored in the sense of the possibility of a selection mechanism among the many weak solutions. This was the leading question behind the analysis in [CCS23], in which it was shown that neither vanishing diffusivity nor regularization of the velocity field provide such a selection mechanism. The lack of a selection principle for weak solutions of the transport equation was also explored in the recent work of Huysmans and Titi, where the authors construct two different renormalized solutions of (1.1) that are strong limits of solutions to (4.1) along two different subsequences as $\kappa \to 0$. They also use perfect mixing and unmixing in time to construct another sequence of solutions to (4.1) that has, as unique limit as $\kappa \to 0$, an entropy-inadmissible solution to (1.1).

In terms of anomalous dissipation, in [CCS23] the authors construct, for any $\alpha$ and $\beta$ in the supercritical regime $\alpha + 2\beta < 1$, a velocity field $u \in L^\infty([0,T];C^\alpha(\mathbb{T}^2))$ and a smooth initial datum $\rho_0$ such that the solutions $\rho^\kappa$ to the advection-diffusion equation (4.1) with initial data $\rho_0$ are uniformly bounded in $L^2([0,T];C^\beta(\mathbb{T}^2))$ and exhibit anomalous dissipation (more general exponents for the integrability in time can be considered).

The basic mechanism is based on the same slice-and-dice strategy that leads to the nonuniqueness results of Depauw and Bressan for the transport equation and is intrinsically related to mixing. In such an example, the solution takes opposite constant values on alternating tiles of a checkerboard, the size of which gets refined as time increases, resulting in perfect mixing (i.e., weak convergence to zero, the average of the solution) at the critical time $t = T$. Compared to the previous literature on mixing, a few important twists are necessary for the analysis in [CCS23]. The refinement of the size of the checkerboards does not follow the classical dyadic rescaling, but rather obeys a superexponential scaling law, which allows (upon suitable choices of the many parameters in the construction) to achieve optimal regularity and to separate the relevant scales at each step of the evolution. The transition from a checkerboard to the next one is realized by shear flows, which are also concentrated at suitable spatial scales. The approach of [CCS23] is fairly explicit and relies on the fact that solutions of the advection-diffusion equation (4.1) have a stochastic Lagrangian representation via the Feynman-Kac formula

$$\rho^\kappa(t,x) = \mathbb{E}\big[\rho_0\big((X^\kappa)^{-1}(t,x)\big)\big].$$

Here $X^\kappa$ satisfies the stochastic differential equation

$$dX^\kappa(t,x) = u(t,X^\kappa(t,x))\,\mathrm{d}t + \sqrt{2\kappa}\,dW,$$

where $W$ is a Brownian motion.

The principle behind the lack of selection in the limit of zero diffusivity can be best understood by first considering the related question of whether regularizing the velocity via convolution with a suitable smoothing kernel can act as a selection principle for the solution, as the regularization parameter vanishes. That is, we pose the question whether limit points of solutions $\rho^\varepsilon$ of

$$\partial_t \rho^\varepsilon + (u * \eta_\varepsilon) \cdot \nabla \rho^\varepsilon = 0$$

are unique, where $\eta_\varepsilon$ is a standard mollifier. To this extent, consider the slice-and-dice construction for $t \in [0, T]$ sketched above, and reflect it oddly across $t = T$, therefore "reconstructing large scales" for $t \in [T, 2T]$. At each step in the reconstruction of the large scales, add a new "move" which has the effect of changing the parity of the checkerboard, that is, it swaps the black and the white tiles. When considering the convolution of the velocity field with $\eta_\varepsilon$, all scales below $\varepsilon$ are "filtered" and therefore the solution does not get fully mixed at time $t = T$, but rather stays at a finite scale when it crosses the singular time. After the singular time, large scales are reconstructed, but, as $\varepsilon \to 0$, the solution will converge either to an even or odd checkerboard for $t \sim 2T$, depending on the "parity of $\varepsilon$." This construction provides two subsequences that converge to distinct limit solutions.

We next focus on the case when diffusion is present. Consider again the evolution of the checkerboards for $t \in [0, T]$. Now, add at each step a time interval of suitable length on which the velocity vanishes and, therefore, the solution obeys the heat equation. Separation of scales (due to the choice of a superexponential sequence) provides the existence of a critical time $t_{\text{crit}}(\kappa) \to T$ such that for $0 < t < t_{\text{crit}}(\kappa)$ diffusion is a perturbation, while it is the dominant effect for times $t \sim t_{\text{crit}}(\kappa)$ due to the intervals where velocity vanishes. Although acting only for a short time, diffusion is enhanced by the high frequencies in the solution, eventually reaching a balance which leads to dissipate a fixed fraction, independent of $\kappa > 0$, of the $L^2$ norm of the solution. Hence, anomalous dissipation occurs.

In the construction sketched above, the possibility of anomalous dissipation relies on a specific choice of the subsequence $\kappa \to 0$ that depends on all other parameters. In fact, another result of [CCS23] is the possibility of choosing another (distinct) subsequence $\kappa \to 0$ with the following property. Exploiting the isotropy of the Brownian motion, the corresponding subsequence of solutions $\rho^\kappa$ converges to a solution of the transport equation which conserves the $L^2$ norm. This example shows that the limit of zero diffusion cannot be used as a selection principle for weak solutions of the transport equation outside of the DiPerna-Lions class.

Building on the results of [CCS23], in collaboration with Bruè and De Lellis the authors have obtained analogous results for the forced Navier-Stokes and Euler equations with full Onsager-supercritical regularity, i.e., for velocity fields in Hölder spaces $C^\alpha$ with any $\alpha < 1/3$.

**5.3. Anomalous dissipation via fractal homogenization.** Both in [DEIJ22] and in [CCS23], anomalous dissipation only occurs at the singular time $t = T$ (see (5.1)), due to the nature of the construction based on mixing and on the development in time of small scales. Such a situation is somewhat inconsistent with the theory of homogeneous isotropic turbulence, which postulates (statistical) stationarity and, therefore, the fact that there should be no "preferred" time in turbulent phenomena: anomalous dissipation should happen continuously in time, and for any randomly chosen times $t_1$ and $t_2$ the corresponding values of the velocity field $u(t_1, \cdot)$ and $u(t_2, \cdot)$ should be macroscopically indistinguishable.

In [AV23], the authors address this issue by relying on an approach based on homogenization theory. For any $\alpha < \frac{1}{3}$, they construct a time-periodic velocity field on the torus that belongs to $C^\alpha$ uniformly in time and exhibits continuous-in-time anomalous dissipation for bounded solutions with arbitrary $H^1$ initial data.

In contrast to the examples in [CCS23] and [DEIJ22], for the velocity field constructed in [AV23] all scales are active at all times. Homogenization theory allows to understand and quantify the enhancement of diffusivity due to the creation of small scales in the solution at all times. This result goes under the name of renormalization of effective diffusivities: each homogenization step along the cascade of scales enhances the effective diffusivity, which after an iteration over all scales remains of order one even as $\kappa \to 0$ and, therefore, gives anomalous dissipation. Homogenization is well understood for a finite number of scales, but the authors of [AV23] need to deal with infinitely many scales at once.

## References

[ACM19] Giovanni Alberti, Gianluca Crippa, and Anna L. Mazzucato, *Exponential self-similar mixing by incompressible flows*, J. Amer. Math. Soc. **32** (2019), no. 2, 445–490, DOI 10.1090/jams/913. MR3904158

[AV23] Scott Armstrong and Vlad Vicol, *Anomalous diffusion by fractal homogenization*, Preprint, `arXiv:2305.05048`, May 2023.

[BBPS21] Jacob Bedrossian, Alex Blumenthal, and Sam Punshon-Smith, *Almost-sure enhanced dissipation and uniform-in-diffusivity exponential mixing for advection-diffusion by stochastic Navier-Stokes*, Probab. Theory Related Fields **179** (2021), no. 3-4, 777–834, DOI 10.1007/s00440-020-01010-8. MR4242626

[BCZG23] Alex Blumenthal, Michele Coti Zelati, and Rishabh S. Gvalani, *Exponential mixing for random dynamical systems and an example of Pierrehumbert*, Ann. Probab. **51** (2023), no. 4, 1559–1601, DOI 10.1214/23-aop1627. MR4597327

[BCZM22] Elia Bruè, Michele Coti Zelati, and Elio Marconi, *Enhanced dissipation for two-dimensional Hamiltonian flows*, Preprint, `arXiv:2211.14057`, November 2022.

[Bre03] Alberto Bressan, *A lemma and a conjecture on the cost of rearrangements*, Rend. Sem. Mat. Univ. Padova **110** (2003), 97–102. MR2033002

[CCS23] Maria Colombo, Gianluca Crippa, and Massimo Sorella, *Anomalous Dissipation and Lack of Selection in the Obukhov–Corrsin Theory of Scalar Turbulence*, Ann. PDE **9** (2023), no. 2, Paper No. 21, https://doi.org/10.1007/s40818-023-00162-9. MR4662772

[CDL08] Gianluca Crippa and Camillo De Lellis, *Estimates and regularity results for the DiPerna-Lions flow*, J. Reine Angew. Math. **616** (2008), 15–46, DOI 10.1515/CRELLE.2008.016. MR2369485

[CEIM22] Gianluca Crippa, Tarek Elgindi, Gautam Iyer, and Anna L. Mazzucato, *Growth of Sobolev norms and loss of regularity in transport equations*, Philos. Trans. Roy. Soc. A **380** (2022), no. 2225, Paper No. 24, 12, DOI 10.1098/rsta.2021.0024. MR4430388

[CKRZ08] P. Constantin, A. Kiselev, L. Ryzhik, and A. Zlatoš, *Diffusion and mixing in fluid flow*, Ann. of Math. (2) **168** (2008), no. 2, 643–674, DOI 10.4007/annals.2008.168.643. MR2434887

[CZDE20] Michele Coti Zelati, Matias G. Delgadino, and Tarek M. Elgindi, *On the relation between enhanced dissipation timescales and mixing rates*, Comm. Pure Appl. Math. **73** (2020), no. 6, 1205–1244, DOI 10.1002/cpa.21831. MR4156602

[DEIJ22] Theodore D. Drivas, Tarek M. Elgindi, Gautam Iyer, and In-Jee Jeong, *Anomalous dissipation in passive scalar transport*, Arch. Ration. Mech. Anal. **243** (2022), no. 3, 1151–1180, DOI 10.1007/s00205-021-01736-2. MR4381138

[ELM23] Tarek M. Elgindi, Kyle Liss, and Jonathan C. Mattingly, *Optimal enhanced dissipation and mixing for a time-periodic, Lipschitz velocity field on* $\mathbb{T}^2$, Preprint, arXiv:2304.05374, April 2023.

[EZ19] Tarek M. Elgindi and Andrej Zlatoš, *Universal mixers in all dimensions*, Adv. Math. **356** (2019), 106807, 33, DOI 10.1016/j.aim.2019.106807. MR4008523

[FI19] Yuanyuan Feng and Gautam Iyer, *Dissipation enhancement by mixing*, Nonlinearity **32** (2019), no. 5, 1810–1851, DOI 10.1088/1361-6544/ab0e56. MR3942601

[IKX14] Gautam Iyer, Alexander Kiselev, and Xiaoqian Xu, *Lower bounds on the mix norm of passive scalars advected by incompressible enstrophy-constrained flows*, Nonlinearity **27** (2014), no. 5, 973–985, DOI 10.1088/0951-7715/27/5/973. MR3207161

[IXZ21] Gautam Iyer, Xiaoqian Xu, and Andrej Zlatoš, *Convection-induced singularity suppression in the Keller-Segel and other non-linear PDEs*, Trans. Amer. Math. Soc. **374** (2021), no. 9, 6039–6058, DOI 10.1090/tran/8195. MR4302154

[LLN+12] Evelyn Lunasin, Zhi Lin, Alexei Novikov, Anna Mazzucato, and Charles R. Doering, *Optimal mixing and optimal stirring for fixed energy, fixed power, or fixed palenstrophy flows*, J. Math. Phys. **53** (2012), no. 11, 115611, 15, DOI 10.1063/1.4752098. MR3026556

[LTD11] Zhi Lin, Jean-Luc Thiffeault, and Charles R. Doering, *Optimal stirring strategies for passive scalar mixing*, J. Fluid Mech. **675** (2011), 465–476, DOI 10.1017/S0022112011000292. MR2801050

[MHSW22] J. Myers Hill, R. Sturman, and M. C. T. Wilson, *Exponential mixing by orthogonal non-monotonic shears*, Phys. D **434** (2022), Paper No. 133224, 14, DOI 10.1016/j.physd.2022.133224. MR4399244

[Pie94] RT Pierrehumbert, *Tracer microstructure in the large-eddy dominated regime*, Chaos, Solitons & Fractals **4** (1994), no. 6, 1091–1110.

[Thi12] Jean-Luc Thiffeault, *Using multiscale norms to quantify mixing and transport*, Nonlinearity **25** (2012), no. 2, R1–R44, DOI 10.1088/0951-7715/25/2/R1. MR2876867


Michele Coti Zelati


Gianluca Crippa


Gautam Iyer


Anna L. Mazzucato

### Credits

Figures 1 and 2 are courtesy of Anna L. Mazzucato, Michele Coti Zelati, Gianluca Crippa, and Gautam Iyer.

Author photo of Michele Coti Zelati is courtesy of Michele Coti Zelati.

Author photo of Gianluca Crippa is courtesy of Gianluca Crippa.

Author photo of Gautam Iyer is courtesy of Carnegie Mellon University.

Author photo of Anna L. Mazzucato is courtesy of Chuck Fong.

# Root Systems in Lie Theory: From the Classic Definition to Nowadays

*Iván Angiono*

## 1. Root Systems: The Origin

The purpose of this article is to discuss the role played by root systems in the theory of Lie algebras and related objects in representation theory, with focus on the combinatorial description and properties.

**1.1. Semisimple Lie algebras.** The study of Lie algebras began toward the end of the 19th century. They emerged as the algebraic counterpart of a purely geometric object: *Lie groups*, which we can briefly define as groups that admit a differentiable structure such that multiplication and the function that computes inverses are differentiable. Lie algebras appeared as some algebraic structure attached to the tangent space of the unit of this group.

Initially Lie algebras were only considered over complex or real numbers, but the abstraction of the definition led to Lie algebras over arbitrary fields.

**Definition 1.1.** Let $\Bbbk$ be a field. A *Lie algebra* is a pair $(\mathfrak{g}, [,])$, where $\mathfrak{g}$ is a $\Bbbk$-vector space and $[,] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ is a bilinear map (called the *bracket*) such that the following equalities hold for all $x, y, z \in \mathfrak{g}$:

$$\begin{aligned} [x,y] &= -[y,x], && \textit{antisymmetry}, \\ [x,[y,z]] &= [[x,y],z] + [y,[x,z]], && \textit{Jacobi identity}. \end{aligned}$$

There is a subtle difference when the field is of characteristic two: the antisymmetry is replaced by $[x,x] = 0$ for all $x \in \mathfrak{g}$ (which implies the former one). From now on all Lie algebras considered here are assumed to be finite-dimensional.

An easy example is to pick a vector space $\mathfrak{g}$ together with trivial bracket $[x,y] = 0$ for all $x, y \in \mathfrak{g}$; these Lie algebras are called *abelian*.

There is a general way to move from an associative algebra $A$ to a Lie algebra: take $\mathfrak{g} = A$ as vector space and set $[a,b] \coloneqq ab - ba$ for each pair $a, b \in A$. A prominent example of this construction is the general linear algebra $\mathfrak{gl}(V)$, which is the set of linear endomorphisms of a finite-dimensional vector space $V$. Other classical examples appear as Lie subalgebras (that is, subspaces closed under the bracket) of $\mathfrak{gl}(V)$:

- $\mathfrak{sl}(V)$, those endomorphisms whose trace is zero; if $V = \Bbbk^n$, then we simply denote $\mathfrak{sl}(V)$ by $\mathfrak{sl}(n, \Bbbk)$, or $\mathfrak{sl}(n)$ when the field $\Bbbk$ is clear from the context.
- The orthogonal and symplectic Lie subalgebras $\mathfrak{so}(V,b)$, respectively $\mathfrak{sp}(V,b)$, of those endomorphisms $T$ such that

$$b\,(T(v), w) + b\,(v, T(w)) = 0 \qquad \text{for all } v, w \in V,$$

where $b$ is a symmetric, respectively antisymmetric, nondegenerate bilinear form on $V$.

Analogously, we may start with $A = \Bbbk^{n\times n}$, the algebra of $n\times n$ matrices, and take some subalgebras, as the subspaces of upper triangular matrices, those of trace 0, the orthogonal matrices, between others.

Once we have a notion of *algebra*, it is natural to ask for ideals: in the case of Lie algebras, these are subspaces $\mathfrak{I} \subseteq \mathfrak{g}$ such that $[\mathfrak{I}, \mathfrak{g}] \subseteq \mathfrak{I}$. This leads to consider *simple* Lie algebras, those Lie algebras $\mathfrak{g}$ such that $\dim \mathfrak{g} > 1$ and the unique ideals are the trivial ones: 0 and $\mathfrak{g}$. In addition, we say that a Lie algebra $\mathfrak{g}$ is *semisimple* if $\mathfrak{g}$ is isomorphic to the direct sum of simple Lie algebras.

For the rest of this section we fix $\Bbbk = \mathbb{C}$. We know that a Lie algebra $\mathfrak{g}$ is simple if and only if $\mathfrak{g}$ is (isomorphic to) $\mathfrak{sl}(V)$, $\mathfrak{so}(V,b)$, $\mathfrak{sp}(V,b)$, and a few exceptional examples $E_k$, $k = 6,7,8$, $F_4$, $G_2$. That is, up to 5 exceptions, all the complex simple Lie algebras are subalgebras of matrices. Thus one may wonder if some properties of the algebras of matrices still hold for simple Lie algebras. We will recall some of them by the end of this section, following [Hum78].

As for associative algebras, we can study modules over Lie algebras. A $\mathfrak{g}$-*module* is a pair $(V, \cdot)$, where $V$ is a

*Iván Angiono is a professor of mathematics at Universidad Nacional de Córdoba, Argentina. His email address is* ivan.angiono@unc.edu.ar.

*Communicated by* Notices *Associate Editor Han-Bom Moon.*



DOI: https://doi.org/10.1090/noti2906

$\Bbbk$-vector space and $\cdot : \mathfrak{g} \otimes V \to V$ is a linear map such that

$$[x, y] \cdot v = x \cdot (y \cdot v) - y \cdot (x \cdot v), \qquad x, y \in \mathfrak{g}, v \in V.$$

For example, the bracket gives an action of $\mathfrak{g}$ over itself, called the *adjoint action*.

For each $x \in \mathfrak{g}$ we look at the *inner derivation*

$$\operatorname{ad} x : \mathfrak{g} \to \mathfrak{g}, \qquad \operatorname{ad} x(y) = [x, y], \qquad y \in \mathfrak{g},$$

associated to the adjoint action. These endomorphisms induce a symmetric bilinear form on $\mathfrak{g}$, called the *Killing form*:

$$\kappa : \mathfrak{g} \times \mathfrak{g} \to \Bbbk, \quad \kappa(x, y) := \operatorname{Tr}(\operatorname{ad} x \operatorname{ad} y), \quad x, y \in \mathfrak{g}.$$

The Killing form and the $\mathfrak{g}$-modules give other characterizations of semisimplicity: $\mathfrak{g}$ is semisimple if and only if $\kappa$ is nondegenerate if and only if every module is semisimple, i.e., every $\mathfrak{g}$-submodule admits a complement which is a $\mathfrak{g}$-submodule.

When $\mathfrak{g}$ is one of the Lie algebras of matrices above, the action of diagonal matrices is, in fact, diagonalizable. Mimicking this fact we look for subalgebras such that the action of their elements is diagonalizable, called *toral* subalgebras.

From now on assume that $\mathfrak{g}$ is also semisimple. It can be shown that toral algebras are abelian, and we pick a maximal one $\mathfrak{h}$. Thus $\mathfrak{g}$ decomposes as the direct sum of the $\mathfrak{h}$-eigenspaces:

$$\mathfrak{g} = \oplus_{\alpha \in \mathfrak{h}^*} \mathfrak{g}_\alpha, \quad \text{where } \mathfrak{g}_\alpha := \{x \in \mathfrak{g} | [h, x] = \alpha(h) x\}.$$

As $\mathfrak{h}$ is abelian, we have that $\mathfrak{h} \subseteq \mathfrak{g}_0$: one can show that we have an equality, $\mathfrak{h} = \mathfrak{g}_0$. Thus, if we set $\Delta := \{\alpha \in \mathfrak{h}^* | \alpha \neq 0, \mathfrak{g}_\alpha \neq 0\}$, then $\Delta$, a finite set called the *root system* of $\mathfrak{g}$, gives a decomposition of $\mathfrak{g}$ into $\mathfrak{h}$-eigenspaces as follows:

$$\mathfrak{g} = \mathfrak{h} \oplus (\oplus_{\alpha \in \Delta} \mathfrak{g}_\alpha).$$

This decomposition is compatible with the bracket,

$$[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] \subseteq \mathfrak{g}_{\alpha+\beta} \qquad \text{for all } \alpha, \beta \in \mathfrak{h}^*,$$

and the Killing form

$$\kappa_{\mathfrak{g}_\alpha \times \mathfrak{g}_\beta} = 0 \qquad \text{if } \alpha + \beta \neq 0.$$

We can derive that $\kappa_{\mathfrak{h} \times \mathfrak{h}}$ is nondegenerate, thus it induces a symmetric nondegenerate bilinear form $(\cdot, \cdot) : \mathfrak{h}^* \times \mathfrak{h}^* \to \mathbb{C}$.

**Example 1.2.** If $\mathfrak{g} = \mathfrak{sl}(4)$, the Lie algebra of $4 \times 4$ matrices with trace 0, then $\mathfrak{h}$ is the subspace of diagonal matrices, with basis $h_i := E_{ii} - E_{i+1,i+1}$, $i = 1, 2, 3$. Here, $E_{ij}$ is the matrix with 1 in the $(i, j)$-entry and 0 otherwise. Let $A := \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$, and set $\alpha_j \in \mathfrak{h}^*$ as the element such that $\alpha_j(h_i) = a_{ij}$. Then

- $e_i := E_{i,i+1} \in \mathfrak{g}_{\alpha_i}$, $i = 1, 2, 3$;
- $E_{13} \in \mathfrak{g}_{\alpha_1+\alpha_2}$, $E_{24} \in \mathfrak{g}_{\alpha_2+\alpha_3}$, $E_{14} \in \mathfrak{g}_{\alpha_1+\alpha_2+\alpha_3}$;
- for all $i < j$, if $E_{ij} \in \mathfrak{g}_\alpha$, then $E_{ji} \in \mathfrak{g}_{-\alpha}$. In particular, $f_i := E_{i+1,i} \in \mathfrak{g}_{-\alpha_i}$, $i = 1, 2, 3$.

Thus, if we set $\alpha_{ij} := \sum_{k=i}^{j} \alpha_k$, $i \leq j$, then

$$\Delta = \{\pm \alpha_{ij} | 1 \leq i \leq j \leq 3\}.$$

This example has a straightforward generalization to $\mathfrak{sl}(n)$ for any $n \geq 2$.

### 1.2. Root systems for Lie algebras.

We may derive strong properties of the root system $\Delta$ using the representation theory of $\mathfrak{sl}(2)$, we refer to [Bou02, Hum78] for more details.

(i) $\mathfrak{h}^*$ is spanned by $\Delta$.

(ii) If $\alpha \in \Delta$, then $-\alpha \in \Delta$. Moreover, for each $\alpha \in \Delta$, $\Delta \cap \mathbb{C}\alpha = \{\pm \alpha\}$.

(iii) For each $\alpha \in \Delta$, the eigenspace $\mathfrak{g}_\alpha$ is one-dimensional. Moreover, $S_\alpha := \mathfrak{g}_\alpha \oplus \mathfrak{g}_{-\alpha} \oplus [\mathfrak{g}_\alpha, \mathfrak{g}_{-\alpha}]$ is a subalgebra isomorphic to $\mathfrak{sl}(2)$. Notice that $[\mathfrak{g}_\alpha, \mathfrak{g}_{-\alpha}] \subset \mathfrak{h}$.

(iv) If $\alpha, \beta, \alpha + \beta \in \Delta$, then $[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] = \mathfrak{g}_{\alpha+\beta}$.

(v) Let $\alpha, \beta \in \Delta$ be such that $\alpha \neq \pm\beta$. Then there exist $q, r \in \mathbb{N}_0$ such that

$$\{i \in \mathbb{Z} | \beta + i\alpha \in \Delta\} = \{-r \leq i \leq q\}.$$

Moreover, $r - q = \frac{2(\beta,\alpha)}{(\alpha,\alpha)}$. That is, the *root string* over $\beta$ in the direction of $\alpha$ has no holes.

By (i) there exists a basis $B$ of $\mathfrak{h}^*$ contained in $\Delta$. We can check that all the coefficients of any $\beta \in \Delta$, written in terms of $B$, are rational numbers, so we may consider the $\mathbb{Q}$-linear subspace $\mathfrak{h}^*_{\mathbb{Q}}$ generated by $\Delta$ and take the extension to $\mathbb{R}$: we get a finite-dimensional $\mathbb{R}$-vector space $V$ which *contains* all the information and the geometry of $\Delta$.

*Remark* 1.3. $V$ becomes an Euclidean vector space with the scalar product induced by the Killing form.

For each $\alpha \in \Delta$ set $s_\alpha : V \to V$,

$$s_\alpha(\beta) = \beta - \frac{2(\beta,\alpha)}{(\alpha,\alpha)} \alpha, \qquad \beta \in V.$$

Then $s_\alpha$ is a linear automorphism of the Euclidean space $V$ such that $s_\alpha^2 = \mathrm{id}$, and by (v), $s_\alpha(\Delta) = \Delta$.

## 2. Classical Root Systems

From the information above one may wonder if there exists an abstract notion of root system. The answer is *yes*, and we will recall it following [Bou02], see also [Hum78]. We can classify all finite root systems in terms of so-called finite Cartan matrices. We will also recall a way to come back from (abstract) root systems to complex Lie algebras.

### 2.1. Abstract definition.

**Definition 2.1** ([Bou02]). Let $V$ be a finite-dimensional $\mathbb{R}$-vector space. A finite subset $\Delta \subset V$ is a *root system* in $V$ if

(RS1) $0 \notin \Delta$ and $V$ is spanned by $\Delta$.

(RS2) For each $\alpha \in \Delta$, there exists $\alpha^\vee \in V^*$ such that $\alpha^\vee(\alpha) = 2$ and the reflection

$$s_\alpha : V \to V, \qquad s_\alpha(\beta) = \beta - \alpha^\vee(\beta)\alpha, \qquad \beta \in V,$$

satisfies that $s_\alpha(\Delta) = \Delta$.

(RS3) For all $\alpha, \beta \in \Delta$, $\alpha^\vee(\beta) \in \mathbb{Z}$.

The elements of $\Delta$ are called *roots*, and $\dim V$ is the *rank* of $\Delta$. The (finite) subgroup $\mathcal{W}$ of $\mathrm{Aut}(V)$ generated by $s_\alpha$, $\alpha \in \Delta$, is the *Weyl group* of $\Delta$.

In [Hum78] one also requires that for each $\alpha \in \Delta$, $\Delta \cap \mathbb{R}\alpha = \{\pm\alpha\}$. In other references, root systems with this extra property are called *reduced*.

The reflections $s_\alpha$, $\alpha \in \Delta$ are univocally determined and there exists a symmetric invariant nondegenerate bilinear form $(\cdot|\cdot) : V \times V \to \mathbb{R}$, which is moreover invariant by $\mathcal{W}$ and positive definite. Now, the elements $\alpha^\vee$ are recovered using this form:

$$\alpha^\vee(\beta) = \tfrac{2(\beta,\alpha)}{(\alpha,\alpha)} \text{ for all } \alpha, \beta \in \Delta.$$

Also, the set $\Delta^\vee := \{\alpha^\vee : \alpha \in \Delta\}$ is a root system of $V^*$, with $(\alpha^\vee)^\vee = \alpha$. There are four examples of reduced root systems in rank 2: $A_1 \times A_1$, $A_2$, $B_2$ and $G_2$, with 2, 6, 8, and 12 roots, respectively. The third one is depicted in Figure 1.

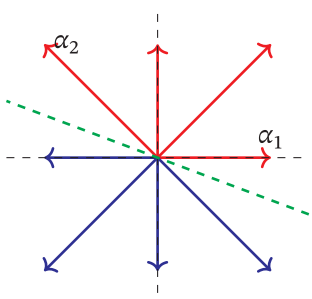



**Figure 1.** Root system of type $B_2$.

Let $\alpha, \beta \in \Delta$ be such that $\alpha \neq \pm\beta$. One may check that $\alpha + \beta \in \Delta$ (respectively, $\alpha - \beta \in \Delta$) if $(\alpha, \beta) < 0$ (respectively, $(\alpha, \beta) > 0$). This is the starting point, together with (RS2) and (RS3), to check that an analogue of (v) holds for (abstract) root systems.

Another key point is the existence of a *base* of a root system. It means a subset $B \subset \Delta$ such that $B$ is a basis of $V$ (as a vector space), and every $\beta \in \Delta$ is written, in terms of $B$, as a linear combination whose coefficients are all nonnegative integers, or all nonpositive integers.

The proof of existence of bases gives the geometric flavor behind root systems. We take a vector $\gamma$ such that the orthogonal hyperplane $P$ to $\gamma$ does not contain any root. Indeed $\gamma$ belongs to $V - \cup_{\alpha\in\Delta} H_\alpha$, where $H_\alpha$ is the kernel of $\alpha^\vee$, i.e., the hyperplane orthogonal to $\alpha$: the connected components of $V - \cup_{\alpha\in\Delta} H_\alpha$ are called the *Weyl chambers*. Thus $\Delta = \Delta^+(\gamma) \cup \Delta^-(\gamma)$, where

$$\Delta^\pm(\gamma) = \{\beta \in \Delta | \pm(\beta, \gamma) > 0\}.$$

A base is made by those *indecomposable* roots in $\Delta^+(\gamma)$: those $\beta \in \Delta^+(\gamma)$ which cannot be written as a sum $\beta = \beta_1 + \beta_2$, with $\beta_i \in \Delta^+(\gamma)$. Moreover every base can be constructed in this way.

For example, in Figure 1 we take the green hyperplane: the positive roots are the red ones, the negative are the blue ones, and $B = \{\alpha_1, \alpha_2\}$ is a base.

The Weyl group $\mathcal{W}$ permutes bases (and Weyl chambers as well), and the action is simply transitive. We check then that any root $\alpha \in \Delta$ belongs to a base, and for each base $B$, $\mathcal{W}$ is generated by $s_\alpha$, $\alpha \in B$ (we reduce the number of generators of $\mathcal{W}$ to the rank of the root system). This leads to the study of *groups generated by reflections* and *Coxeter groups* considered in [Bou02], which became an important subject of research on its own, and remains active until now.

## 2.2. The classification.

As for algebraic objects, we may ask for *irreducible* root systems: those which cannot split into two orthogonal subsets (otherwise each subset is itself a root system). Every root system $\Delta$ of $V$ decomposes uniquely as a union of irreducible root systems $\Delta_i$ corresponding to the subspaces $V_i$ of $V$ spanned by $\Delta_i$. Thus, in order to classify root systems, we can restrict to the irreducible ones.

Assume now that $\Delta$ is an irreducible root system of rank $\theta$. Set $A^\Delta \in \mathbb{Z}^{\theta\times\theta}$ as the matrix with entries

$$a_{ij} := \alpha_i^\vee(\alpha_j) = \tfrac{2(\alpha_j,\alpha_i)}{(\alpha_i,\alpha_i)}, \qquad 1 \le i, j \le \theta,$$

where $B = \{\alpha_i\}_{1\le i\le\theta}$ is a base. One can check that $A^\Delta$ is well-defined; i.e., it does not depend on the chosen base. In addition, $A$ is indecomposable: for all $i < j$ there exist $i_k \in \{1, \cdots, \theta\}$ such that $a_{ii_1}a_{i_1i_2}\cdots a_{i_tj} \neq 0$. Moreover,

(GCM1) $a_{ii} = 2$ for all $1 \le i \le \theta$,

(GCM2) $a_{ij} = 0$ if and only if $a_{ji} = 0$,

(GCM3) for all $i \neq j$, $a_{ij} \le 0$.

Any $A \in \mathbb{Z}^{\theta\times\theta}$ satisfying (GCM1)–(GCM3) is called a *generalized Cartan matrix* (GCM) [Kac90]. The information of GCM is encoded in a graph called the *Dynkin diagram*: it has $\theta$ vertices, labelled with $1, 2, \dots, \theta$, and for each pair $1 \le i < j \le \theta$,

- if $a_{ij}a_{ji} \le 4$, then we add $\max\{|a_{ij}|, |a_{ji}|\}$ edges between vertices $i$ and $j$, with an arrow from $j$ to $i$ (respectively $i$ to $j$) if $|a_{ij}| > 1$ (respectively, $|a_{ji}| > 1$); in particular, if $a_{ij} = 0$ (so $a_{ji} = 0$ as well) then we draw no edges between $i$ and $j$, and if $a_{ij} = a_{ji} = -1$, then we draw just a line;
- if $a_{ij}a_{ji} > 4$, then we draw a thick line between $i$ and $j$ labelled with $(|a_{ij}|, |a_{ji}|)$ .

For example, the Dynkin diagrams of $\left[\begin{smallmatrix} 2 & -2 \\ -2 & 2 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 & -3 \\ -2 & 2 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{smallmatrix}\right]$ are, respectively

$$\circ \Longleftrightarrow \circ \qquad \circ \overset{(3,2)}{\text{———}} \circ \qquad \circ \text{———} \circ \text{———} \circ \,.$$

One reason to differentiate between $a_{ij}a_{ji} \le 4$ and $a_{ij}a_{ji} > 4$ is all finite and affine Dynkin diagrams satisfy the first condition, and these are probably the most studied cases. We refer to [Bou02, Hum78] for the definition of affine

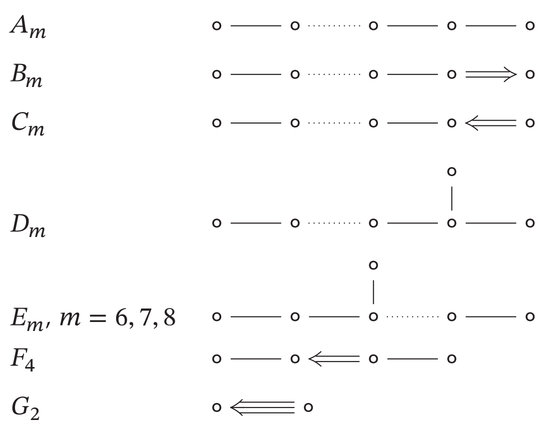



**Figure 2.** Finite connected Dynkin diagrams.

Dynkin diagrams while finite ones are depicted in Figure 2, in connection with finite-dimensional complex Lie algebras.

One may define the Weyl group $\mathcal{W}^A$ of a GCM $A$ as the subgroup of $\mathrm{Aut}(V)$ generated by reflections $s_i : V \to V$, $s_i(\alpha_j) = \alpha_j - a_{ij}\alpha_i$, where $(\alpha_i)_{1\le i\le\theta}$ is the canonical basis of $V = \mathbb{R}^\theta$: if $A$ is the Cartan matrix of a Lie algebra as above, then the Weyl group of $\Delta$ is generated by these $s_i$'s. Analogously, we can define

$$\Delta^A = \{w(\alpha_i) | w \in \mathcal{W}^A, 1 \le i \le \theta\}.$$

Then one can prove that $\mathcal{W}^A$ is finite if and only if $\Delta^A$ is finite, which is equivalent to the notion of *finite GCM*. Finite GCM are parametrized by *finite* Dynkin diagrams, i.e., those in Figure 2.

**Theorem 2.2.** *Reduced irreducible root systems are parametrized by Dynkin diagrams in Figure 2.*

Up to now we deal with three notions:

(i) Simple Lie algebras over $\mathbb{C}$,
(ii) Irreducible root systems,
(iii) Finite Cartan matrices, or the corresponding Dynkin diagrams.

We moved first from (i) to (ii), and then state a correspondence (ii)$\longleftrightarrow$(iii). Now we need to come back to (i). We can check that $\mathfrak{sl}(n+1)$ has Cartan matrix of type $A_n$ (see Example 1.2), while matrices of types $B_n$, $C_n$ and $D_n$ appear for orthogonal and symplectic Lie algebras. For each one of the exceptional finite Cartan matrices $A$ in Figure 2 we can construct *by hand* a simple Lie algebra with Cartan matrix $A$. The natural question is if there exists a *systematic* way to build these Lie algebras. We will recall it in the next subsection, i.e., a correspondence (iii)$\longrightarrow$(i).

### 2.3. Back to Lie algebras: Kac-Moody construction.

Looking at Example 1.2, the Cartan matrix of $\mathfrak{sl}(4)$ can be recovered from the action of the Cartan subalgebra $\mathfrak{h}$ on eigenvectors of a base of the root system $\Delta$. In addition the decomposition $\Delta = \Delta^+ \cup \Delta^-$ into positive and negative roots for the chosen base corresponds in this case to the upper and lower triangular matrices $\mathfrak{n}_\pm$ of $\mathfrak{sl}(4)$ (recall that $\mathfrak{h}$ is spanned by the set of all the diagonal matrices in $\mathfrak{sl}(4)$).

As for associative algebras, we have a notion of a Lie algebra *presented by generators and relations* as the appropiate quotient of a *free* Lie algebra. We will attach a Lie algebra $\mathfrak{g} := \mathfrak{g}(A)$ to each matrix $A \in \mathbb{C}^{\theta\times\theta}$; these algebras were introduced by Serre in 1966 for finite matrices $A$, and by Kac and Moody in two independent and simultaneous works in the late sixties, see [Kac90] and the references therein. For the sake of simplicity of the exposition we assume that $\det A \neq 0$.

Let $\tilde{\mathfrak{g}} := \tilde{\mathfrak{g}}(A)$ be the Lie algebra presented by generators $e_i, h_i, f_i$, $1 \le i \le \theta$, and relations

$$\begin{aligned} [h_i, h_j] &= 0, & [h_i, e_j] &= a_{ij} e_j, \\ [e_i, f_j] &= \delta_{ij} h_i, & [h_i, f_j] &= -a_{ij} f_j. \end{aligned} \tag{1}$$

Let $\mathfrak{h}$ be the subspace spanned by $(h_i)_{1\le i\le\theta}$, $\tilde{\mathfrak{n}}_\pm$ the subalgebra generated by $(e_i)_{1\le i\le\theta}$, respectively $(f_i)_{1\le i\le\theta}$. We have the following facts:

(a) $\tilde{\mathfrak{n}}_\pm$ is a free Lie algebra in $\theta$ generators.
(b) As a vector space, $\tilde{\mathfrak{g}} = \tilde{\mathfrak{n}}_+ \oplus \mathfrak{h} \oplus \tilde{\mathfrak{n}}_-$.
(c) The adjoint action of $\mathfrak{h}$ on $\tilde{\mathfrak{n}}_\pm$ is diagonalizable.
(d) Among all the ideals of $\tilde{\mathfrak{g}}$ intersecting trivially $\mathfrak{h}$, there exists a maximal one $\mathfrak{r}$, which satisfies

$$\mathfrak{r} = (\mathfrak{r} \cap \tilde{\mathfrak{n}}_+) \oplus (\mathfrak{r} \cap \tilde{\mathfrak{n}}_-).$$

**Definition 2.3.** The *contragredient* Lie algebra $\mathfrak{g}(A)$ associated to $A$ (sometimes called the *Kac-Moody algebra*) is the quotient $\mathfrak{g}(A) := \tilde{\mathfrak{g}}(A)/\mathfrak{r}$.

Because of the definition of $\mathfrak{r}$, $\mathfrak{g}(A)$ is generated by $e_i, f_i, h_i$, $1 \le i \le \theta$, has a triangular decomposition

$$\mathfrak{g}(A) = \mathfrak{n}_+ \oplus \mathfrak{h} \oplus \mathfrak{n}_-,$$

where $\mathfrak{n}_\pm$ is the image of $\tilde{\mathfrak{n}}_\pm$ under the projection $\pi : \tilde{\mathfrak{g}}(A) \twoheadrightarrow \mathfrak{g}(A)$, i.e., the subalgebra generated by (the image of) $(e_i)_{1\le i\le\theta}$, respectively $(f_i)_{1\le i\le\theta}$, and any other Lie algebra with a triangular decomposition as above, generated by the same set of generators satisfying (1), projects onto $\mathfrak{g}(A)$.

**Theorem 2.4.** (A) *Let $A$ be a finite Cartan matrix. Then $\mathfrak{g}(A)$ is a finite-dimensional simple Lie algebra, with Cartan matrix $A$.*
(B) *The list of Dynkin diagrams in Figure 2 provides a classification of all finite-dimensional simple Lie algebras over $\mathbb{C}$.*

When the generalized Cartan matrix $A$ is not of finite type, the associated Kac-Moody Lie algebra $\mathfrak{g}(A)$ is infinite-dimensional. Although for the purposes of this exposition we are interested in the finite-dimensional examples, the infinite-dimensional Lie algebras $\mathfrak{g}(A)$ (or at least some of them, mainly the affine ones) are quite important since they have appeared in connection either with other areas of mathematics, especially representation theory, or theoretical physics, for example in conformal field theory.

## 3. Root Systems for Other Kinds of Lie Algebras

Next we deal with contragredient Lie algebras over fields of positive characteristic and later with Lie superalgebras over any field. We will recall the main differences with the picture of Lie algebras over $\mathbb{C}$ which leads to a more general notion of root system. This root system still captures the combinatorics of these Lie theoretic objects.

### 3.1. Lie algebras over fields of positive characteristic.

Let $\Bbbk$ be an algebraically closed field of characteristic $p > 0$. The study of simple Lie algebras becomes more and more complicated as far as $p$ is smaller, see, e.g., [Str04]. A main difference with the case of complex numbers is that not all simple Lie algebras have a triangular decomposition as above, and the Cartan subalgebra plays a weaker role in the structure of the whole Lie algebra.

On the other hand, Definition 2.3 still holds over $\Bbbk$, so we may ask about the classification of finite-dimensional contragredient Lie algebras. A subtle difference is that we restrict to $\mathbb{Z}$-homogeneous ideals intersecting trivially $\mathfrak{h}$, where each $e_i$ has degree 1, each $f_i$ has degree -1 and each $h_i$ has degree 0. Thus, there exists a finer grading of the Lie algebra $\mathfrak{g}(A)$ by $\mathbb{Z}^\theta$, where $\deg e_i = \alpha_i$ (the $i$-th element of the canonical basis), $\deg f_i = -\alpha_i$ and $\deg h_i = 0$ as well. Let $\Delta^A \subset \mathbb{Z}^\theta$ be the subset of all nonzero degrees whose homogeneous components are nontrivial.

*Remark* 3.1. From the triangular decomposition,

$$\Delta^A \subset \mathbb{N}_0^\theta \cup (-\mathbb{N}_0^\theta),$$

that is, the coefficients of each $\alpha \in \Delta$ are all nonnegative, or else all nonpositive. Also, there exists an involution $\omega$ of $\mathfrak{g}(A)$ (called the Chevalley involution) such that $e_i \mapsto f_i$, $f_i \mapsto e_i$, $h_i \mapsto -h_i$. As $\omega\left(\mathfrak{g}(A)\right)_\beta = \mathfrak{g}(A)_{-\beta}$ for all $\beta \in \mathbb{Z}^\theta$, we have that

$$\Delta^A = -\Delta^A.$$

For example we can consider the finite Cartan matrices over $\Bbbk$, since the entries of these Cartan matrices are integer numbers, and show that the associated Lie algebras are finite-dimensional. But, even for contragredient Lie algebras, there are significant differences with the case of complex numbers. As shown in [VK71], there are examples of finite-dimensional Lie algebras with diagonal entries $a_{ii} = 0$, and two different matrices can give place to isomorphic contragredient Lie algebras. The classification shown in [VK71] was incomplete: there was a missing example for $p = 3$, the 29-dimensional Brown algebra $\mathfrak{br}(3)$, discovered by Brown in the eighties, whose realization as contragredient Lie algebra with two different matrices was shown in [Skr93]:

**Theorem 3.2.** *Fix $p = 3$. Let*

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & 0 \\ -2 & 2 & -1 \\ 0 & 1 & 0 \end{bmatrix}.$$

*Then there exists an isomorphism $\Phi : \mathfrak{g}(A) \to \mathfrak{g}(B)$ such that*

$$\Phi(e_1) = \underline{e}_1, \qquad \Phi(e_2) = (\operatorname{ad} \underline{e}_3)^2 \underline{e}_2, \qquad \Phi(e_3) = \underline{f}_3,$$
$$\Phi(f_1) = \underline{f}_1, \qquad \Phi(f_2) = (\operatorname{ad} \underline{f}_3)^2 \underline{f}_2, \qquad \Phi(f_3) = \underline{e}_3.$$

The expression of $\Phi$ is close to that for the action of reflections of the Weyl group on complex Lie algebras, but here $\Phi$ relates two "different" contragredient data.

*Remark* 3.3. We fix the following GCM

$$C^A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -2 & 2 \end{bmatrix}, \quad C^B = \begin{bmatrix} 2 & -1 & 0 \\ -2 & 2 & -1 \\ 0 & -2 & 2 \end{bmatrix},$$

and set $s_i^A, s_i^B : \mathbb{Z}^3 \to \mathbb{Z}^3$ as the corresponding reflections defined by $C^A$, $C^B$, respectively, $i = 1, 2, 3$. Notice that $s_3^A = s_3^B$ and $\Phi\left(\mathfrak{g}(A)_\beta\right) = \mathfrak{g}(B)_{s_3^A(\beta)}$ for all $\beta \in \mathbb{Z}^3$. Thus $\Delta^B = s_3^A(\Delta^A)$.

In addition, there exist automorphisms

$$\Phi_i^A : \mathfrak{g}(A) \to \mathfrak{g}(A), \quad \Phi_i^B : \mathfrak{g}(B) \to \mathfrak{g}(B), \quad i = 1, 2,$$

such that $\Phi_i^A\left(\mathfrak{g}(A)_\beta\right) = \mathfrak{g}(A)_{s_i^A(\beta)}$, $\Phi_i^B\left(\mathfrak{g}(B)_\beta\right) = \mathfrak{g}(B)_{s_i^B(\beta)}$ for all $\beta \in \mathbb{Z}^3$. This implies that

$$\Delta^A = s_i^A(\Delta^A), \qquad \Delta^B = s_i^B(\Delta^B), \qquad i = 1, 2.$$

### 3.2. Lie superalgebras.

Recall that a *Lie superalgebra* is a $\mathbb{Z}_2$-graded vector space $\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1$ ($\mathfrak{g}_0$ is the *even* part and $\mathfrak{g}_1$ is the *odd* part) together with a linear $\mathbb{Z}_2$-graded map $[,] : \mathfrak{g} \otimes \mathfrak{g} \to \mathfrak{g}$ satisfying analogous versions of antisymmetry and Jacobi identity:

$$[x, y] = -(-1)^{|x||y|}[y, x],$$
$$[x, [y, z]] = [[x, y], z] + (-1)^{|x||y|}[y, [x, z]],$$

for all homogeneous elements $x, y, z \in \mathfrak{g}$, see, e.g., [Kac77]. Here, $|x| \in \{0, 1\}$ denotes the degree of $x$. We have examples from associative algebras, analogous to those of Lie algebras: given $\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1$ a $\mathbb{Z}_2$-graded associative algebra, set

$$[x, y] = xy - (-1)^{|x||y|}yx.$$

In particular we have, for $V = V_0 \oplus V_1$, the Lie superalgebra $\mathfrak{gl}(V) = \operatorname{End}(V)$, with

$$\mathfrak{gl}(V)_i = \{T \in \mathfrak{gl}(V) : T(V_j) \subseteq V_{i+j}\}.$$

For each $T \in \mathfrak{gl}(V)$ set $\operatorname{str}(T) := \operatorname{tr}\left(T_{|V_0}\right) - \operatorname{tr}\left(T_{|V_1}\right)$, the *super trace* of $T$. We can consider the subalgebra

$$\mathfrak{sl}(V) = \{T \in \mathfrak{gl}(V) : \operatorname{str}(T) = 0\}.$$

Here we consider contragredient data $(A, \mathbf{p})$, where $A = (a_{ij})_{1 \le i, j \le \theta} \in \Bbbk^{\theta \times \theta}$ is still the matrix of scalars determining the action of the generators $h_i$ on the remaining generators, and $\mathbf{p} = (p_i)_{1 \le i \le \theta} \in \mathbb{Z}_2^\theta$ gives the $\mathbb{Z}_2$-grading: $e_i, f_i \in \mathfrak{g}(A, \mathbf{p})_{p_i}$ for all $i$. Notice that all $h_i$ are necessarily even.

As observed in [Kac77] when $\Bbbk = \mathbb{C}$, different contragredient data can give isomorphic Lie superalgebras. But, similar to Lie algebras in positive characteristic, we can give isomorphisms between some pairs $\mathfrak{g}(A, \mathbf{p})$, $\mathfrak{g}(A', \mathbf{p}')$ with formulas close to the action of the Weyl group for complex simple Lie algebras. In this direction, Serganova [Ser96] introduced the notion of *odd reflection* relating two different pairs by a kind of reflection but on a simple odd root $e_i$ such that $a_{ii} = 0$ (called *isotropic*). This is consistent with one of the differences with Lie algebras: there exists a symmetric bilinear form on $\mathfrak{h}$, but either the bilinear form can have isotropic roots $\alpha$ (i.e., $(\alpha, \alpha) = 0$) or else the matrix $A \in \mathbb{C}^{\theta \times \theta}$ can take nonintegral values. We have to distinguish the matrix $A$ from the GCM $C^A \in \mathbb{Z}^{\theta \times \theta}$ responsible for the odd reflections.

**Example 3.4.** Let $V = (\mathbb{C}^2)_0 \oplus \mathbb{C}_1$, that is a $\mathbb{Z}_2$-graded vector space with even component of dimension 2, and odd component of dimension 1. Here $\mathfrak{sl}(V)$ is denoted simply by $\mathfrak{sl}(2|1)$: as for Lie algebras we identify $\mathfrak{sl}(2|1)$ with matrices $\mathtt{A} \in \mathbb{C}^{3\times 3}$, here with zero supertrace, i.e., $\mathtt{a}_{11} + \mathtt{a}_{22} - \mathtt{a}_{33} = 0$. The Lie superalgebra $\mathfrak{sl}(2|1)$ is $\mathbb{Z}^2$-graded, with $\mathfrak{h}$ (the diagonal matrices) in degree 0, and one-dimensional components of degrees $\pm\alpha_1$ (even roots, since the corresponding spaces are $E_{12}$ and $E_{21}$), $\pm\alpha_2$, $\pm(\alpha_1 + \alpha_2)$ (these four roots are odd). The contragredient datum is $A = \left[\begin{smallmatrix} 2 & -1 \\ 1 & 0 \end{smallmatrix}\right]$, $\mathbf{p}_2 = (0, 1)$. The odd reflection in $\alpha_2$ moves to the pair $B = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, $\mathbf{p}_{12} = (1, 1)$. Thus we may apply the odd reflection in $\alpha_1$ to $(B, \mathbf{p}_{12})$ and obtain $(C, \mathbf{p}_1)$, where $C = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 2 \end{smallmatrix}\right]$, $\mathbf{p}_1 = (1, 0)$. These are all the possible movements between pairs whose associated Lie superalgebra is isomorphic to $\mathfrak{sl}(2|1)$.

We see that the set of $\mathbb{Z}^2$-degrees of the nontrivial components is the *same*, the Cartan matrix is $\left[\begin{smallmatrix} 2 & -1 \\ -1 & 2 \end{smallmatrix}\right]$ for the three pairs, but the parity of the elements is not the same. For example, for $(B, \mathbf{p}_{12})$, $\pm(\alpha_1 + \alpha_2)$ are even roots while $\pm\alpha_1$, $\pm\alpha_2$ are odd.

If we study Lie superalgebras over fields of positive characteristic, then we can have more and more exceptional examples. Finite-dimensional contragredient Lie algebras over fields of prime characteristic were classified in [BGL09]. The picture is the same: several pairs of contragredient data $(A, \mathbf{p})$ give isomorphic Lie superalgebras.

The question is then how to handle uniformly all possible pairs giving isomorphic Lie superalgebras, and their corresponding roots (i.e., the $\mathbb{Z}^\theta$-degrees of the nontrivial components). This will be done with a groupoid, i.e., a category where all the morphisms are invertible. As we look for a generalization of the Weyl group, we will consider a groupoid generated by *reflections*.

### 3.3. Generalized root systems.

There exist different notions of *generalized* root systems in the literature. They try to capture different situations, as for example the one by Serganova in [Ser96] for complex finite-dimensional Lie superalgebras. A nice axiomatic version was given in [HY08], see also [HS20] for a refined version of these ideas.

Fix $\theta \in \mathbb{N}$, $\mathbb{I} = \{1, \cdots, \theta\}$. Let $\mathcal{X} \neq \emptyset$ a set (which will correspond to the different contragredient data). A *semi-Cartan graph* $\mathcal{G}(\mathbb{I}, \mathcal{X}, (C^X)_{X\in\mathcal{X}}, (\rho_i)_{i\in\mathbb{I}})$ ($\mathcal{G}$ for short) of *rank* $\theta$ over $\mathcal{X}$ consist of

- functions $\rho_i : \mathcal{X} \to \mathcal{X}$, $i \in \mathbb{I}$, such that $\rho_i^2 = \mathrm{id}_\mathcal{X}$,
- GCM $C^X = (c_{ij}^X)_{i,j\in\mathbb{I}} \in \mathbb{Z}^{\theta\times\theta}$, $X \in \mathcal{X}$,

such that $c_{ij}^X = c_{ij}^{\rho_i(X)}$ for all $X \in \mathcal{X}$, $i, j \in \mathbb{I}$.

As for Lie algebras, we set $s_i^X \in \mathrm{GL}(\mathbb{Z}^\theta)$ as the reflection $s_i^X(\alpha_j) = \alpha_j - c_{ij}^X \alpha_i$.

Let $\mathcal{M}$ be a monoid. There exists a small category $\mathcal{D}(\mathcal{X}, \mathcal{M})$ whose set of objects is $\mathcal{X}$ and the set of morphisms between any two objects is $\mathcal{M}$. Given $f \in \mathcal{M}$ and $X, Y \in \mathcal{X}$ we write $(Y, f, X)$ for $f$ viewed as an element of $\mathrm{Hom}(X, Y)$, so the composition becomes

$$(Z, f, Y) \circ (Y, g, X) = (Z, fg, X),$$

for any $X, Y, Z \in \mathcal{X}$, $f, g \in \mathcal{M}$.

We are interested in the case $\mathcal{M} = \mathrm{GL}(\mathbb{Z}^\theta)$, the group of automorphisms of $\mathbb{Z}^\theta$.

**Definition 3.5.** The *Weyl groupoid* of $\mathcal{G}$ is the full subcategory $\mathcal{W}(\mathbb{I}, \mathcal{X}, (A_x)_{x\in\mathcal{X}}, (\rho_i)_{i\in\mathbb{I}})$ of $\mathcal{D}(\mathcal{X}, \mathrm{GL}(\mathbb{Z}^\theta))$ generated by

$$\sigma_i^X := (\rho_i(X), s_i^X, X), \qquad i \in \mathbb{I}, X \in \mathcal{X}.$$

Notice that $\mathcal{W}$ is indeed a groupoid, since

$$\sigma_i^{\rho_i(X)} \sigma_i^X = (X, \mathrm{id}_X, X) \text{ for all } i \in \mathbb{I}, X \in \mathcal{X}.$$

Fix $X \in \mathcal{X}$. Then $\Delta^{X,\mathrm{re}}$ is the set of all elements of the form $w(\alpha_i) \in \mathbb{Z}^\theta$, where $i \in \mathbb{I}$, $Y \in \mathcal{X}$, $(X, w, Y) \in \mathrm{Hom}_\mathcal{W}(X, Y)$. This is the set of *real roots* of $\mathcal{G}$. As for roots of Lie (super)algebras, we consider the subsets

$$\Delta_+^{X,\mathrm{re}} := \Delta^{X,\mathrm{re}} \cap \mathbb{N}_0^\theta, \qquad \Delta_-^{X,\mathrm{re}} := \Delta^{X,\mathrm{re}} \cap (-\mathbb{N}_0^\theta),$$

of positive and negative real roots, and set

$$m_{ij}^X := |\Delta^{X,\mathrm{re}} \cap (\mathbb{N}_0\alpha_i + \mathbb{N}_0\alpha_j)| \in \mathbb{N} \cup \{\infty\}.$$

If $\Delta^{X,\mathrm{re}}$ is finite for all $X \in \mathcal{X}$ (equivalently, for some $X \in \mathcal{X}$), then we say that $\mathcal{G}$ is *finite*.

A semi-Cartan graph $\mathcal{G}$ is a *Cartan graph* if the following conditions hold for all $X \in \mathcal{X}$:

- $\Delta^{X,\mathrm{re}} = \Delta_+^{X,\mathrm{re}} \cup \Delta_-^{X,\mathrm{re}}$;
- for all $i \neq j$ such that $m_{ij}^X < \infty$, $(\rho_i\rho_j)^{m_{ij}^X}(X) = X$.

Mimicking what happens for Lie algebras, see Remarks 3.1 and 3.3, we introduce the following notion:

**Definition 3.6.** A *root system* over $\mathcal{G}$ is a family $\mathcal{R} = (\Delta^X)_{X\in\mathcal{X}}$ of subsets $\Delta^X \subset \mathbb{Z}^\theta$ such that

$$0 \notin \Delta^X, \qquad \Delta^X \subset \mathbb{N}_0^\theta \cup (-\mathbb{N}_0^\theta),$$
$$\alpha_i \in \Delta^X, \qquad s_i^X(\Delta^X) = \Delta^{\rho_i(X)},$$

for all $i \in \mathbb{I}$ and all $X \in \mathcal{X}$.

We say that $\mathcal{R}$ is *reduced* if $\mathbb{Z}\alpha \cap \Delta^X = \{\pm\alpha\}$ for all $\alpha \in \Delta^X$, $X \in \mathcal{X}$. $\mathcal{R}$ is *finite* if every $\Delta^X$ is so.

Cuntz and Heckenberger obtained the classification of finite root systems [CH15]. The proof involves the bijection between root systems of rank $\theta$ and crystallographic arrangements (certain subsets of hyperplanes) in $\mathbb{R}^\theta$. About the list of finite root systems, in rank $\theta = 2$ there are infinitely many examples, in bijection with triangulations of $n$-gons for any $n \geq 3$. For $\theta \geq 9$ we only have families corresponding to Lie superalgebras and Lie algebras of types $A, B, C, D$, while for $3 \leq \theta \leq 8$ we have members of the families of Lie (super)algebras and several exceptions.

The definition of a root system seems to carry the possibility to have several examples attached to the same Cartan graph $\mathcal{G}$. But this is not the case when $\mathcal{G}$ is finite. Indeed, by [HS20, 10.4.7], if $\mathcal{G}$ is a finite Cartan graph, then $\mathcal{R} = (\Delta^{X,\mathrm{re}})_{X\in\mathcal{X}}$ is the only reduced root system over $\mathcal{G}$.

*Remark* 3.7. In [HY08] the authors state the existence of a Weyl groupoid for finite-dimensional complex Lie superalgebras, the ones coming from the $\mathbb{Z}^\theta$-grading as above. Moreover, Andruskiewitsch and Angiono proved that the same holds for Lie superalgebras over fields of arbitrary characteristic, in a work in progress. In the same work they derived the classification of finite-dimensional Lie superalgebras from the classification of finite root system in [CH15].

It should be noted that not all finite root systems come from a Lie superalgebra.

Once we show the existence of a finite root system for a Lie superalgebra, there are many strong properties derived from the combinatorics of the Weyl groupoid. For example:

- $\dim \mathfrak{g}(A, \mathbf{p})_\alpha = 1$ for all $\alpha \in \Delta^{(A,\mathbf{p})}$.
- There might exist roots $\alpha \in \Delta^{(A,\mathbf{p})}$ such that $2\alpha \in \Delta^{(A,\mathbf{p})}$, which are the odd nonisotropic roots. All of them are the image of simple odd nonisotropic roots of some pair $(A', \mathbf{p}')$ obtained up to odd reflections, and $\dim \mathfrak{g}(A, \mathbf{p})_{2\alpha} = 1$, as shown by Andruskiewitsch-Angiono.
- The whole set $\Delta^{(A,\mathbf{p})}$ is obtained up to reflections of the simple roots, attaching $2\alpha$ for each odd nonisotropic root.

In the same line we may wonder if there exists a *geometric-combinatoric* side on these Lie superalgebras (Weyl chambers and so on) coming from the associated crystallographic arrangements.

**Example 3.8.** We continue with the study of $\mathfrak{br}(3)$. Here $\mathcal{X} = \{A, B\}$, with $\rho_3(A) = B$ and $\rho_1 = \rho_2 = \mathrm{id}$. The associated GCM are those in Remark 3.3. Thus we get all the roots applying repeatedly the reflections. For example, $s_2^B(\alpha_1) = \alpha_1 + 2\alpha_2 \in \Delta^B$, so

$$s_3^B(\alpha_1 + 2\alpha_2) = \alpha_1 + 2\alpha_2 + 4\alpha_3 \in \Delta^A.$$

Using the notation $1^a2^b3^c := a\alpha_1 + b\alpha_2 + c\alpha_3$, we can check that

$$\begin{aligned}\Delta_+^A =&\{1, 12, 123, 1^22^33^4, 12^23^2, 12^23^3, 12^23^4,\\ &12^33^4, 123^2, 2, 23^2, 23, 3\},\\ \Delta_+^B =&\{1, 12^2, 12, 123^2, 12^33^2, 1^22^33^2, 12^23^2,\\ &123, 12^23, 2, 23^2, 23, 3\}.\end{aligned}$$

Thus, $\dim \mathfrak{n}_\pm = 13$, so $\dim \mathfrak{br}(3) = 29$. In addition one can show that $[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] = \mathfrak{g}_{\alpha+\beta}$ for every pair $\alpha, \beta \in \Delta_+^A$ such that $\alpha + \beta \in \Delta_+^A$. Thus we can obtain recursively a nonzero element $e_\alpha \in \mathfrak{g}_\alpha$:

$$e_{12} := [e_1, e_2], \quad e_{123} := [e_{12}, e_3], \quad e_{123^2} := [e_{123}, e_3],$$

and so on.

## 4. Other Contexts and Problems

We finish by recalling other algebraic structures where these generalized root systems appear, as Nichols algebras, and posing some related problems where they could play a key role: Lie algebras in a broad sense and their representations. We are not going to introduce all the involved concepts, we refer to the corresponding papers for more information.

**4.1. Nichols algebras.** Quantized enveloping algebras are certain deformations of enveloping algebras of semisimple Lie algebras introduced in the eighties by Drinfeld and Jimbo, depending on a parameter $q$. Later on, Lusztig considered Hopf algebras obtained by evaluation of $q$ at a root of unity, which lead to some finite-dimensional examples, usually called *Frobenius-Lusztig kernels*. These examples have a triangular decomposition whose zero part is a group algebra of copies of finite cyclic groups, and the positive (also, the negative) part is a kind of Hopf algebra.

In the denomination currently used, these positive parts are examples of Nichols algebras. Nichols algebras are Hopf algebras in the category of Yetter-Drinfeld modules over a group algebra (or more precisely, over a Hopf algebra), which play a fundamental role in the classification of finite-dimensional Hopf algebras. Following the line of work of Andruskiewitsch and Schneider, joined by Heckenberger, one can define Hopf algebras with triangular decomposition, whose positive part is a Nichols algebra and which have generalised root systems, see the book [HS20], and also [AHS10].

The list of all finite-dimensional Nichols algebras is known when the group is finite and abelian, thanks to the work of Heckenberger, and almost complete when the group is finite but nonabelian, by Heckenberger-Vendramin. Both works explode the existence of the generalized root system.

We can see that the list of all generalized root systems appearing for some Nichols algebras contains properly the list of all those appearing for Lie superalgebras, but there are some root systems not attached to any Nichols algebras.

**4.2. Lie algebras in symmetric tensor categories and representations.** One can extend the definition of Lie algebra to symmetric tensor categories. Indeed Lie superalgebras are essentially Lie algebras in the category of super vector spaces. When $\Bbbk$ is of characteristic zero, Deligne proved that any symmetric tensor category (under a mild condition) fibers over the category of supervector spaces, so any Lie algebra over these symmetric tensor categories can be considered as a Lie superalgebra. When $\Bbbk$ is of characteristic $p > 0$, Coulembier-Etingof-Ostrik proved recently [CEO23] that any symmetric tensor category (under a mild condition) fibers over the Verlinde category $\mathrm{Ver}_p$. This category is the semisimplification of the category of representations of $\mathbb{Z}_p$ over $\Bbbk$ and contains properly the category of super vector spaces. Thus, in this case, the consideration of Lie algebras in symmetric tensor categories essentially reduces to Lie algebras in $\mathrm{Ver}_p$. One may ask about the existence of contragredient Lie algebras in $\mathrm{Ver}_p$, and root systems.

In the classical case (that is, over $\mathbb{C}$), the root system controls the representation theory of simple Lie algebras, or more precisely a quite interesting subcategory called the *category $\mathcal{O}$*. For example, finite-dimensional modules are parametrized by nonnegative *weights* associated to the root system, and the Weyl group describes a *character formula* for these simple modules. The situation is a bit more complicated for Lie superalgebras, and a character formula exists for certain weights. Recently Sergeev and Veselov used what they called a Weyl groupoid (which is not clearly related to the one considered here) to describe strong properties on the representations. Also, Yamane described very recently character formulas for the so-called atypical weights of quantized enveloping Lie superalgebras by means of the Weyl groupoid. So one may wonder if the Weyl groupoid plays a key role in the description of the representations of Lie algebras in a broad sense.

## References


[AHS10] Nicolás Andruskiewitsch, István Heckenberger, and Hans-Jürgen Schneider, *The Nichols algebra of a semisimple Yetter-Drinfeld module*, Amer. J. Math. **132** (2010), no. 6, 1493–1547. MR2766176

[BGL09] Sofiane Bouarroudj, Pavel Grozman, and Dimitry Leites, *Classification of finite dimensional modular Lie superalgebras with indecomposable Cartan matrix*, SIGMA Symmetry Integrability Geom. Methods Appl. **5** (2009), Paper 060, 63, DOI 10.3842/SIGMA.2009.060. MR2529187

[Bou02] Nicolas Bourbaki, *Lie groups and Lie algebras. Chapters 4–6*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 2002. Translated from the 1968 French original by Andrew Pressley, DOI 10.1007/978-3-540-89394-3. MR1890629

[CEO23] Kevin Coulembier, Pavel Etingof, and Victor Ostrik, *On Frobenius exact symmetric tensor categories*, Ann. of Math. (2) **197** (2023), no. 3, 1235–1279, DOI 10.4007/annals.2023.197.3.5. With Appendix A by Alexander Kleshchev. MR4564264

[CH15] Michael Cuntz and István Heckenberger, *Finite Weyl groupoids*, J. Reine Angew. Math. **702** (2015), 77–108, DOI 10.1515/crelle-2013-0033. MR3341467

[HS20] István Heckenberger and Hans-Jürgen Schneider, *Hopf algebras and root systems*, Mathematical Surveys and Monographs, vol. 247, American Mathematical Society, Providence, RI, 2020. MR4164719

[HY08] István Heckenberger and Hiroyuki Yamane, *A generalization of Coxeter groups, root systems, and Matsumoto's theorem*, Math. Z. **259** (2008), no. 2, 255–276, `https://doi.org/10.1007/s00209-007-0223-3`. MR2390080

[Hum78] James E. Humphreys, *Introduction to Lie algebras and representation theory*, Graduate Texts in Mathematics, vol. 9, Springer-Verlag, New York-Berlin, 1978. Second printing, revised. MR499562

[Kac77] V. G. Kac, *Lie superalgebras*, Advances in Math. **26** (1977), no. 1, 8–96, DOI 10.1016/0001-8708(77)90017-2. MR486011

[Kac90] Victor G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge University Press, Cambridge, 1990, DOI 10.1017/CBO9780511626234. MR1104219

[Ser96] Vera Serganova, *On generalizations of root systems*, Comm. Algebra **24** (1996), no. 13, 4281–4299, `https://doi.org/10.1080/00927879608825814`. MR1414584

[Skr93] S. M. Skryabin, *A contragredient* 29*-dimensional Lie algebra of characteristic* 3 (Russian, with English and Russian summaries), Sibirsk. Mat. Zh. **34** (1993), no. 3, 171–178, 223, 228, DOI 10.1007/BF00971230; English transl., Siberian Math. J. **34** (1993), no. 3, 548–554. MR1241179

[Str04] Helmut Strade, *Simple Lie algebras over fields of positive characteristic. I: Structure theory*, De Gruyter Expositions in Mathematics, vol. 38, Walter de Gruyter & Co., Berlin, 2004, DOI 10.1515/9783110197945. MR2059133

[VK71] B. Ju. Veĭsfeĭler and V. G. Kac, *Exponentials in Lie algebras of characteristic* $p$ (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **35** (1971), 762–788. MR306282


Iván Angiono

## Credits


Figures 1 and 2 and author photo are courtesy of Iván Angiono.

# Differential Equations for Continuous-Time Deep Learning

*Lars Ruthotto*

Deep learning has been behind incredible breakthroughs, such as voice recognition, image classification, and text generation. While these successes are undeniable, our mathematical understanding of deep learning is still developing. More rigorous insight is needed to overcome fundamental challenges, including the interpretability, robustness, and bias of deep learning, and lower its environmental and computational costs.

Let us define deep learning informally as machine learning methods that use feed-forward neural networks with many (that is, more than a handful) layers. While most traditional approaches use a finite number of layers, we will focus on more recent approaches that conceptually use infinitely many layers. We will explain those approaches by defining differential equations whose dynamics are modeled by trainable neural network components and whose time roughly corresponds to the depth of the network.

Using three examples from machine learning and applied mathematics, we will see how continuous-depth neural network architectures, defined by ordinary differential equations (ODEs), can provide new insights into deep learning and a foundation for more efficient algorithms. Even though many deep learning approaches used in practice today do not rely on differential equations, I find many opportunities for mathematical research in this area. As we shall see, phrasing the problem continuously in time enables one to borrow numerical techniques and analysis to gain more insight into deep learning, design new approaches, and crack open the black box of deep learning. We shall also see how neural ODEs can approximate solutions to high-dimensional, nonlinear optimal control problems.

*Lars Ruthotto is a Winship Distinguished Research Associate Professor at Emory University. His email address is* `lruthotto@emory.edu`.

*Communicated by* Notices *Associate Editor Reza Malek-Madani.*

*For permission to reprint this article, please contact:* `reprint-permission@ams.org`.

DOI: https://doi.org/10.1090/noti2930

This article targets readers familiar with ordinary and partial differential equations and their analysis and who are curious to see their role in machine learning. Therefore, rather than drawing a complete picture of state-of-the-art machine learning, which is shooting at moving targets, we seek to provide foundational insights and motivate further study. To this end, we will inevitably take shortcuts and sacrifice being up-to-date for clarity. Even though designing efficient numerical algorithms is critical to translating the theoretical advantages of the continuous-time viewpoint into practical learning approaches, we will keep this topic for another day and provide some references for the interested reader.

While giving a complete picture of the research activity at this interface between differential equations and deep learning is beyond the scope of this paper, we seek to capture the key ideas and provide a head start to those eager to learn more. This paper aims to illustrate the foundations and some of the benefits of continuous-time deep learning using a few handpicked examples related to the author's activity in the area. The article is written to be self-contained, and it has been an enormous challenge to limit the number of citations to those we believe are most valuable to enable the interested reader an efficient way into this new field. We provide an unedited arXiv version of this paper to provide additional references and context.

## Deep Neural Networks in Continuous Time

Before we begin, let us introduce the main mathematical notation of deep learning for this paper and motivate continuous-time architectures.

Throughout this paper, $F_\theta : \mathbb{R}^n \to \mathbb{R}^m$ denotes a neural network with the subscript $\theta \in \mathbb{R}^p$ representing its parameters, often called weights. We will see that there are different ways of defining $F_\theta$. For example, a common choice is the $L$-layer feed-forward network that maps the input $x_0 = x \in \mathbb{R}^n$ to the output $F_\theta(x) = x_L$ via the layers

$$x_{l+1} = \sigma_l \left(W_l x_l + b_l\right), \ \forall l = 0, \dots, L-1. \tag{1}$$

Here, the activation functions $\sigma_l$ are applied element-wise, and the parameters of the $l$th layer, $\theta_l$, consist of the weight matrix $W_l$ and the weight vector $b_l$. To define $F_\theta$, one must choose the number of layers, activation functions, and the sizes of the weight matrices and vectors, collectively called hyperparameters. Except for the number of columns in $W_0$, which must be $n$, and the number of rows in $W_{L-1}, b_{L-1}$, which must be $m$, the remaining sizes can be adjusted arbitrarily.

With an effective way to choose hyperparameters and identify the model weights, deep networks are perhaps the most efficient high-dimensional function approximators today. Their versatility has enabled their use across various tasks. For example, neural networks have been used in supervised learning to fit given inputs to corresponding outputs, in reinforcement learning to predict optimal actions, and in generative modeling to match simple latent distributions to a complex distribution available only through samples. The success of neural networks across these tasks is rooted in their approximation properties. For example, it can be shown that networks with one hidden layer are universal approximators. Since we will use these networks as an important building block later, let us define them as

$$f_\theta(x) = W_1 \tanh(W_0 x + b_0) + b_1, \tag{2}$$

whose parameter, $\theta$, consists of the weight matrices $W_0 \in \mathbb{R}^{k\times n}, W_1 \in \mathbb{R}^{m\times k}$, weight vectors $b_0 \in \mathbb{R}^k, b_1 \in \mathbb{R}^m$, and whose activation is the hyperbolic tangent function. In other words, for every $\epsilon > 0$, there is a width, $k$, such that they approximate continuous functions to a given accuracy $\epsilon > 0$. Since the width needed to achieve the desired accuracy can be impractically large, most applications today prefer narrower but deeper architectures.

Increasing the depth of neural networks such as the one in (1) to realize approximation results is easier said than done. In practice, it often becomes more and more challenging to identify model weights that accurately approximate the function of interest. For example, it is difficult to approximate the identity function with a network like (1).

Residual neural networks (ResNets) provide an alternative way to define very deep networks and considerably improved the state-of-the-art in computer vision applications recently [HZRS1603]. Their key innovation is often called a skip connection that turns, for example, (2) into the residual layer

$$r_\theta(x) = x + f_\theta(x). \tag{3}$$

Such a ResNet layer can learn the identity map by the trivial choice $f_\theta \equiv 0$. Consequently, increasing the network depth by adding residual layers often improves the approximation result since the weights of the new layers can be chosen to approximate the identity.

One way to motivate deep neural networks that are continuous in time is to view $r_\theta(x)$ in (3) as a forward Euler approximation of $z(1)$ where $z$ solves the initial value problem

$$\frac{d}{dt} z = f_{\theta(t)}(z), \quad t \in (0,1], \quad z(0) = x. \tag{4}$$

Here, $t$ is an artificial time, and with the notation $\theta(t)$, we seek to suggest that the weights can be modeled as functions of time; see [E1703, HR18]. This viewpoint was popularized in the machine learning community by the work [CRBD18], which also coined the term Neural ODEs and demonstrated several new use cases. It is important to remember that ResNets and Neural ODEs are different: one is discrete, and the other is continuous. This has implications both in theory and in practice.

## Supervised Learning in Continuous Time

In this section, we show how using continuous-time models for supervised learning leads to learning problems that can be analyzed and solved using tools from optimal control.

In supervised learning, the goal is to learn $F_\theta$ that approximates the relation between labeled input-output pairs $(x, y) \sim D$ assumed to be independent samples from some data distribution $D$. Once the hyperparameters of the network are chosen, learning the weights $\theta$ is typically phrased as a minimization problem, such as

$$\min_\theta \mathbb{E}_{(x,y)\sim D}[\ell(F_\theta(x), y)], \tag{5}$$

with some loss function $\ell$ whose definition depends on the task; for example, for data fitting, one can consider the regression loss $\ell(v, y) = \frac{1}{2}\|v - y\|^2$. The optimization problem is challenging and interesting in its own right.

A simple way to define a continuous-time model is to define $F_\theta$ as an affine transformation of the terminal state of a neural ODE, that is,

$$F_\theta(x) = Wz(1) + b \tag{6}$$

$$\frac{d}{dt} z = f_{\theta(t)}(z), \quad t \in (0,1], \quad z(0) = x. \tag{7}$$

This model can be interpreted as the affine model (given by the parameters $W$ and $b$) applied to the features evolved by the neural ODE, whose dynamics are governed by the weight function $\theta$. Note that the affine function can be omitted when $m = n$.

Since neural ODEs as in (6) yield an invertible transformation of the data space, it is unsurprising that many functions cannot be approximated by this model. In Figure 1, we demonstrate this in one example that can also illustrate the role of the ODE in the supervised learning problem. Shown here is a classification example obtained by minimizing a logistic regression loss function. Each data point $x$ is associated with a label $y$, either blue or red. The goal is to learn a function $F_\theta$ corresponding to the training data. The center column of the figure shows

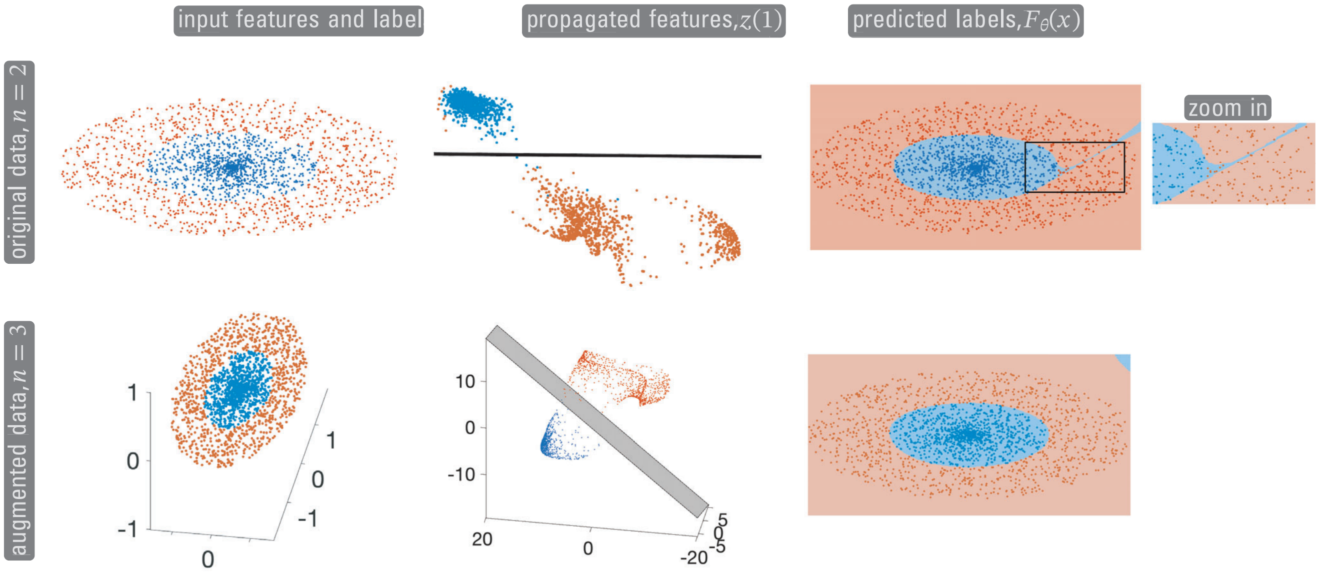



**Figure 1.** Illustration of a continuous-time model for binary classification. Left column: input features and labels. Center column: propagated features given by the final state of ODE and hyperplane given by $W, b$. Right column: Labels predicted by the neural network. The rows show two instances of the problem using the original two-dimensional and augmented features (padded with one zero), respectively. While both models agree closely around the samples, we highlight some errors of the original model that arise from the restriction to invertible maps in $n = 2$. This example demonstrates that augmenting overcomes the need for a noninvertible mapping. Note that the models may not be reliable in regions with no data points, for example, in the top-right corner of the domain.

the propagated features, which are the $z(1)$ associated with each example, as well as the hyperplane parameterized by $W, b$ that seeks to divide the features. The rightmost column visualizes the predictions of the classifier. While the predictions are nearly perfect in both rows, upon close inspection, the limitations of the ODE shine through in the top row, and it can be seen that the network was unable to transform the blue and red points to become linearly separable, which requires a noninvertible transformation. The need for that is alleviated by simply padding the input features with one zero and embedding them into three dimensions. This phenomenon and the importance of augmentation are elaborated in [DDT19].

When $F_\theta$ is defined as in (6) the minimization problem (5) becomes an optimal control problem

$$\begin{aligned} &\min_{\theta, W, b, z} \mathbb{E}_{(x,y)\sim D}[\ell(Wz(1) + b, y)], \\ &\text{s.t.} \frac{d}{dt} z = f_{\theta(t)}(z), \quad t \in (0,1] \\ &z(0) = x. \end{aligned} \tag{8}$$

The relation between learning $\theta$ and solving an optimal control problem has been used to gain insights and obtain more efficient algorithms. Universal approximation results of continuous-time models are derived in [LLS23] by analyzing their flow maps. On the computational side, it is worth mentioning the approach in [LCTE17], which uses control theory to derive new learning algorithms, and [BCE+19], which studies the continuous and discrete versions of the learning problem and proposes schemes that can learn time discretizations and embed other constraints.

With an architecture continuous in time, it is also possible to provide a PDE perspective of the supervised learning problem (8). Following the presentation in [WYSO20], let us take a macroscopic viewpoint and consider the space of examples rather than individual data points. To this end, we model the neural network predictions as a function $u : \mathbb{R}^n \times [0,1] \to \mathbb{R}^m$ whose evolution is governed by the transport PDE with velocity $f_\theta$. In doing so, we formulate supervised learning as a PDE-constrained optimization problem

$$\begin{aligned} &\min_{u, \theta, W, b} \mathbb{E}_{(x,y)\sim D}[\ell(u(x,1), y)], \\ &\text{s.t. } \partial_t u(z,t) + f_{\theta(t)}(z)^\top \nabla u(z,t) = 0 \\ &u(z,0) = Wx + b. \end{aligned} \tag{9}$$

To verify that the problems are equivalent, note that (6) defines the characteristic curves of the transport equation and therefore

$$\begin{aligned} u(x,1) &= u(z(0),1) \\ &= Wz(1) + b \\ &= F_\theta(x) \approx y. \end{aligned}$$

Since the dimensionality of the feature space is usually larger than two or three, problem (9) is intractable. However, this viewpoint can provide new insights and motivate improved learning algorithms. For example, [WYSO20] showed that adding some amount of viscosity to the PDE constraint in (9) can increase the robustness of classifiers to random perturbations of the inputs. They also proposed an efficient method based on the Feyman-Kac formula to scale to high dimensions. The interpretation also allows us to build bridges to optical flow and image registration, which may yield further understanding in the future.

## Continuous-Time Generative Models

This section illustrates the advantages of continuous-time models for building flexible generative models.

In generative modeling, one learns complicated, often high-dimensional, data distributions from examples. The goal typically is to enable sampling and sometimes includes estimating the densities of given examples. Rather than mapping input points to corresponding output points, as in supervised learning, a generative model maps a tractable reference distribution to a target distribution. In deep generative modeling, this mapping is called a generator, and it is represented by a deep neural network.

While the choice of reference distribution is arbitrary (as long as it is easy to sample from), most commonly, it is a standard Gaussian. To motivate the use of continuous-time models, we will set the dimension of the latent distribution, $n$, equal to the data distribution, $m$. We also assume both distributions are proper in $\mathbb{R}^n$, which enables the use of normalizing flows. When this assumption is violated (as is common in practice), other generative models, such as variational autoencoders or generative adversarial networks, are typically superior; a general introduction to generative modeling is given in [RH21]. We illustrate the generative modeling problem in Figure 2.

Under these assumptions, the idea of normalizing flows is to learn a diffeomorphic generator that maps reference to target; that is, we seek to find an invertible $F_\theta$ such that both $F_\theta$ and $F_\theta^{-1}$ are continuously differentiable. Since the reference distribution is a standard Gaussian, its density, which we denote by $\pi_X$, is easy to compute. To estimate the density of a point $y$ from the unknown target distribution under the current generator $F_\theta$, we can use the change of variable formula

$$\pi_Y(y) = \pi_X(F_\theta^{-1}(y)) \cdot \det \nabla F_\theta^{-1}(y). \tag{10}$$

Maximum likelihood training aims to find a parameter $\theta$ that maximizes the expected likelihood over all the samples. One typically considers minimizing the expected negative log-likelihood

$$\mathbb{E}_y \left[ \frac{1}{2} \|F_\theta^{-1}(y)\|^2 - \log\det\left(\nabla F_\theta^{-1}(y)\right) \right], \tag{11}$$

where the first term is (up to an additive constant) the negative log-likelihood of the standard normal $\pi_X$. Even though the functional is convex in $F_\theta^{-1}$, it is not convex in $\theta$

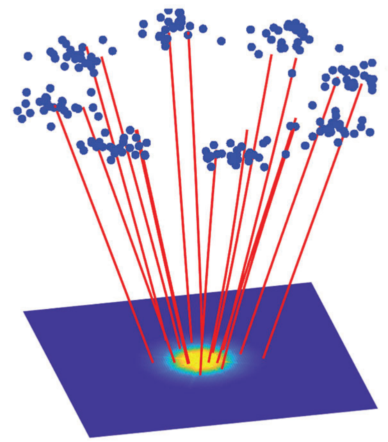


**Figure 2.** Illustrating the generative modeling problem. Given samples from the target distribution (represented by blue dots), we try to find an invertible transformation (represented by red lines) to a simple target distribution (a standard Gaussian).

once we approximate it with a neural network. Therefore, numerical optimization schemes are required to compute an approximate minimizer.

The most critical question for normalizing flows is choosing the neural network architecture $F_\theta$. For example, even though the classical multilayer perception (1) can approximate any generator, it is generally not invertible, so it cannot be trained using (11). Trading off the expressiveness of the model with invertibility and computational considerations has led to various approaches. A key idea to build normalizing flows is to concatenate finitely many layers designed to have easy-to-compute inverses and Jacobian log determinants. However, this construction can limit expressiveness, and often, many layers are needed to approximate mappings in high dimensions. As discussed in more detail in [RH21], one can sometimes increase the approximation power of $F_\theta^{-1}$ by sacrificing the computational efficiency of evaluating the generator $F_\theta$.

Approaches that define the generator using a continuous-time model are known under the term continuous normalizing flows [GCB+18]. When defining $F_\theta(x) = z(1)$ as the terminal state of the neural ODE (4) and $f_\theta$ is sufficiently regular so that the ODE solution is defined uniquely, we can, at least formally, compute the inverse of the generator by integrating backward in time and defining $F_\theta^{-1}(y) = w(0)$, where

$$\frac{d}{dt} w = f_{\theta(t)}(w), \quad t \in [0,1), \quad w(1) = y. \tag{12}$$

One should always be careful when reversing the time in an ODE since, in general, it should not be assumed that the ODE is stable both forward and backward. However, in practice, using fairly arbitrary neural networks to represent $f_\theta$ has been found to induce negligible errors.

Let us also comment on the evaluation of the Jacobian log-determinant. When we define $F_\theta^{-1}(y) = w(0)$ as above, the instantaneous change of variables formula from [CRBD18, Appendix A] implies

$$\log\det(\nabla F_\theta^{-1}(y)) = \int_0^1 \mathrm{tr}\nabla f_{\theta(t)}(w)dt. \tag{13}$$

Numerically integrating the trace of the Jacobian using quadrature rules is often computationally more efficient than computing its log determinant. The computation of the log determinant can also be combined with the numerical ODE solver used to compute the inverse of the generator.

The training of the continuous-time generator $F_\theta$ can now be phrased as an optimal control problem

$$\begin{aligned}\min_{\theta,w} \mathbb{E}_y &\left[\frac{1}{2}\|w(0)\|^2 - \int_0^1 \mathrm{tr}\nabla f_{\theta(t)}(w)dt\right]\\ \text{s.t.} &\frac{d}{dt}w = f_{\theta(t)}(w), t \in [0,1),\\ &w(1) = y.\end{aligned} \tag{14}$$

It turns out that this control problem admits infinitely many solutions. As we shall see in the following section, it is possible for two different networks $f_\theta^{(1)}$ and $f_\theta^{(2)}$ to yield the same generator but follow different trajectories; impatient readers may skip ahead to Figure 3. Even though one often cannot see the differences in the created samples, realizing the nonuniqueness allows one to bias the search toward generators with more regular trajectories. It also bridges generative modeling and optimal transport, which has a rich theory and long history.

The idea of penalizing transport costs has been investigated and shown practical benefits in [YK22, FJNO2011, OFLR21]. Since optimal transport in high dimensions is difficult, a tractable approach is to penalize transport costs, for example, by adding the functional

$$P_{\mathrm{OT}}[w, f_\theta] = \int_0^1 \frac{\alpha}{2}\|f_{\theta(t)}(w)\|^2 dt \tag{15}$$

to the objective function in (14). Here, the parameter $\alpha$ balances matching the distributions (for $\alpha \ll 1$) and minimizing the transport costs (for $\alpha \gg 0$). It is possible to show that we obtain the optimal transport map for an appropriate choice of $\alpha$ and that trajectories become straight. This regularity can translate to practical benefits since the ODEs for the generator and its inverse become trivial to solve.

Adding transport costs to the generative modeling problem also provides exciting opportunities to analyze the training problem. To give a glimpse into this area, we note that the training problem can be written on the macroscopic level as the PDE-constrained optimization problem

$$\begin{aligned}\min_{\rho,\theta} \int_{\mathbb{R}^n}\int_0^1 &\frac{\alpha}{2}\|f_{\theta(t)}(x)\|^2\rho(t,x)dt - \log\rho(1,x)\pi_Y(x)dx\\ \text{s.t.}&\partial_t\rho(t,x) + \nabla\cdot\left(f_{\theta(t)}(x)\rho(t,x)\right) = 0,\\ &\rho(0,x) = \pi_X(x).\end{aligned} \tag{16}$$

Here, the PDE constraint for $t \in (0,1]$ is given by the continuity equation. The formulation above is a relaxed version of the classical dynamic optimal transport formulation. To be concrete, above the terminal constraint $\rho(1,x) = \pi_Y$ is relaxed, and deviations are penalized by the second term of the objective. Similar to the dynamic OT problem, (16) can be reformulated into a convex variational problem, which can provide further insight but becomes impractical when the growth of $n$ requires nonlinear function approximators.

## Neural ODEs for Potential Mean Field Games

This section showcases Neural ODE's promise for overcoming the limitations of other numerical methods for simulating interactions within large populations of agents playing a noncooperative game.

To show how neural ODEs arise naturally in the mean field limit of many games, let us generalize (16) to include objective functions that contain more general cost terms in, for example, using the objective functional

$$\begin{aligned}\mathcal{J}[\rho, f_\theta] = &\int_0^1\int_{\mathbb{R}^n} L(x, f_\theta)\rho(t,x)dxdt\\ &+ \int_0^1 \mathcal{F}(\rho(t,\cdot))dt + \mathcal{G}(\rho(1,\cdot)),\end{aligned} \tag{17}$$

which consists of running cost given by the function $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and the functional $\mathcal{F}$, and a terminal cost functional $\mathcal{G}$. Given an initial state of the population density $\pi_X$, finding the optimal strategy $f_\theta$ amounts to solving the mean field game

$$\begin{aligned}&\min_{\rho,\theta}\mathcal{J}[\rho, f_\theta]\\ &\text{s.t. } \partial_t\rho(t,x) + \nabla\cdot\left(f_{\theta(t)}(x)\rho(t,x)\right) = 0,\\ &\qquad\rho(0,x) = \pi_X(x),\end{aligned} \tag{18}$$

where the continuity equation models the evolution of the population density $\rho$. This more general version of (16) can be used to model various noncooperative differential games played by a large population of rational agents. Furthermore, the formulation allows one to analyze and solve a larger set of generative models beyond continuous normalizing flows [ZK23]. Some examples are listed in Table 1. Also, we provide two one-dimensional instances motivated by optimal transport and crowd motion

| | $L(x, f_\theta)$ | $\mathcal{F}(\rho)$ | $\mathcal{G}(\rho)$ |
|---|---|---|---|
| optimal transport (OT) | $\frac{\alpha}{2}\|f_\theta\|^2$ | | $\mathrm{KL}(\rho, \rho_Y)$ |
| crowd motion | $\frac{\alpha}{2}\|f_\theta\|^2 + Q(x)$ | $\int \rho \log(\rho) dx$ | |
| normalizing flow | | | $\mathbb{E}_y[-\log(\rho)]$ |
| normalizing flow+OT | $\frac{\alpha}{2}\|f_\theta\|^2$ | | $\mathbb{E}_y[-\log(\rho)]$ |

**Table 1.** Examples of different potential mean field games that can be modeled via (18). We also recommend [ZK23] for a more exhaustive list, including score-based diffusion and Wasserstein gradient flows.

to illustrate our problem setup and notation in Figure 3. For simplicity, we focus the observation on deterministic games but note that neural network techniques have also been proposed for stochastic MFGs governed by the Fokker Planck equation [LFL⁺21].

It is important to note that solving (18) in general remains a daunting task, especially when the dimension of the state space, $n$, is larger than three or four. In these cases, the curse of dimensionality affects traditional numerical methods that rely on meshes or grids to solve the continuity equation in (18). Unfortunately, many realistic use cases of MFGs arising in economics, social science, and other fields require considerably larger $n$ to capture the state of the agents.

Deriving a neural ODE formulation for approximating the solution of a class of MFGs requires some calculus and theoretical tools. Here, we will briefly overview our approach in [ROL⁺20]. A key quantity for analyzing and solving the mean field game is its value function $\Phi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$. An intuitive way to define it is via a microscopic perspective. Let $x \in \mathbb{R}^n$ be the state of an arbitrary agent at time $t \in [0, 1)$. Then, the value function $\Phi$ denotes the optimal cost to go for this agent and can be written as

$$\Phi(t, x) = \min_{\theta, z} J_t[f_\theta, \rho, z], \qquad \text{s.t.} \frac{d}{dt} z = f_{\theta(s)}(z), s \in (t, 1], \quad z(t) = x, \tag{19}$$

where $\rho$ is the population density at the equilibrium, and we define the single-agent objective functional

$$J_t[f_\theta, \rho, z] = G(z(1), \rho(1, z(1))) + \int_t^T \left(L\left(z, f_{\theta(s)}(z)\right) + F(z, \rho(s, z))\right) ds, \tag{20}$$

with $F$ and $G$ denoting the $L^2$ derivatives of $\mathcal{F}$ and $\mathcal{G}$, respectively. This makes this an MFG in potential form and ensures that solving the problem from the microscopic and macroscopic perspective leads to the same solution.

Evaluating $J_t$ in (20) requires the agent to estimate the equilibrium density resulting from the collective behavior of the agents around its current trajectory. To this end, we solve the continuity equation in (18). Fortunately, characteristic curves of the continuity equation coincide with the trajectories of the agents. Hence, rather than solving the continuity equation everywhere to compute the density, we can update the densities along the trajectories as the agent's state evolves. Since most of the common choices listed in Table 1 require the log of the density, let us note that along the curves $z(\cdot)$ we have

$$\log \rho(t, z(t)) = \log \pi_X(x) - \int_0^t \mathrm{tr} \nabla f_{\theta(s)}(z) ds. \tag{21}$$

This allows us to eliminate the continuity equation in problem (19) without requiring a grid or a mesh. The resulting Lagrangian approach also has a crucial computational advantage since we can compute the trajectories and objective function values for several agents in parallel without the need to communicate.

The above observations and our experience from the previous section could also be used to obtain a neural ODE approach for potential mean field games. However, the learned solution may violate some theoretical properties. For example, the Pontryagin Maximum Principle shows

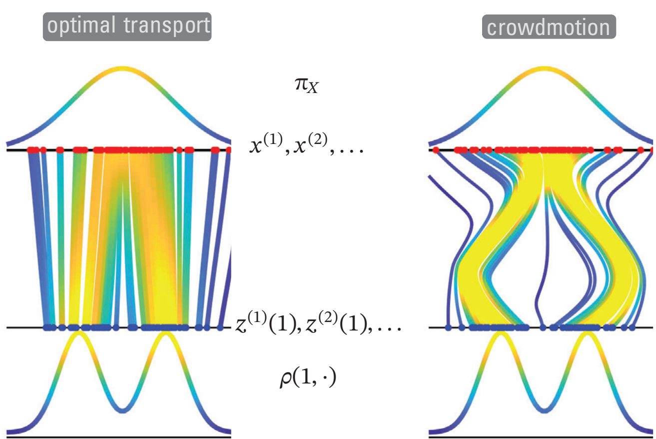



**Figure 3.** Illustration of potential mean field game versions of relaxed dynamic optimal transport (left) and crowd motion problems (right). Both cases use the standard Gaussian reference $\pi_X$ (top) and the same Gaussian mixture as the target (bottom). As expected, in the optimal transport case, the trajectories are straight, whereas in the crowd motion case, the agents are curved to avoid an obstacle in the center of the domain. This example also shows that different dynamics can produce the same map $F_\theta$.

that the optimal control, $f^*$, is related to the value function via the feedback form

$$f^*(x) = -\nabla_p H(x, \nabla\Phi(t,x)). \tag{22}$$

Here, the Hamiltonian $H$ is the Fenchel dual of the running cost $L$ defined by

$$H(x,p) = \sup_{f\in\mathbb{R}^n} \left\{-p^\top f - L(x,f)\right\}. \tag{23}$$

For many choices of $L$ that arise in practice, including the examples in Table 1, $H$ can be computed analytically.

Even when the neural network can approximate the optimal policy, in our experience, finding weights such that properties such as (22) approximately hold is nontrivial. Clearly, choosing the network weights randomly will not yield an approximation that satisfies the feedback form, which means we must solve the learning problem well. Therefore, when $H$ is available and straightforward to compute, we can approximate $\Phi_\theta$ with a scalar-valued neural network and define the control implicitly via (22). Since the value function contains all the information about the MFG solution, approximating it directly can provide helpful insight.

Approximating the value function directly also allows us to incorporate more prior knowledge into the training problem. It is known that the value function solves the Hamilton Jacobi Bellman (HJB) equations

$$\begin{aligned} -\partial_t\Phi_\theta(t,x) + H(x, \nabla\Phi_\theta(t,x)) &= F(x, \rho(t,x)), \\ \Phi_\theta(1,x) &= G(x, \rho(1,x)). \end{aligned} \tag{24}$$

Here, the $-\partial_t$ emphasizes that this equation is backward in time. Lasry and Lions have shown that solving the above PDE in conjunction with the continuity equations is the necessary and sufficient condition for the problem. Even though solving (24) in high dimensions is affected by the curse of dimensionality, we can monitor the violations of the HJB equations along the trajectories and penalize them using some functional $P_{\mathrm{HJB}}$.

To summarize the above observations, we can train the scalar deep neural network, $\Phi_\theta$, that approximates the value functions via the optimal control problem

$$\begin{aligned} &\min_\theta \mathbb{E}_{x\sim\pi_X}\left[J\left[f_\theta, \rho, z\right] + \beta P_{\mathrm{HJB}}[\Phi_\theta, \rho, z]\right] \\ &\text{s.t.} \frac{d}{dt}\begin{pmatrix} z \\ \log(\rho(t,z)) \end{pmatrix} = \begin{pmatrix} f_\theta(z) \\ -\mathrm{tr}\nabla f_\theta(z) \end{pmatrix} \\ &z(0) = x, \quad \log\rho(0,x) = \log\pi_X(x). \end{aligned} \tag{25}$$

## Discussion and Outlook

We demonstrated how differential equations can be used to build continuous-time deep learning approaches. We highlighted a few ways this could lead to new insights and more efficient algorithms for supervised machine learning, generative modeling, and solving high-dimensional mean field games.

The key step is to transform the feature space incrementally using an ODE whose dynamics are represented by a deep neural network. Conceptually, this leads to infinitely deep networks whose artificial time loosely corresponds to the depth of the network. This relation is not precise since, depending on the choice of $f_\theta$, the dynamics of the ODE can be scaled arbitrarily and even depend on trainable parameters. One can consider continuous-time architectures as infinitely deep, which sets them apart from traditional networks consisting of finitely many layers.

Earlier works on continuous-time learning that bring in ideas from differential equations include, for example, [RMKK+92]. Some of the ideas in this work resemble the ones popularized by [CRBD18], but the latter contained other novel ideas, for example, the use of differential equations for generative models later extended in [GCB+18]. The renewed interest in continuous-time models is probably related to the growth of computational resources, advances in numerical methods for solving ODEs and optimal control problems arising in their training, and larger datasets.

We inevitably omitted many important topics to keep the presentation short and coherent. Important examples of continuous-time architectures not governed by ODEs are the controlled differential equations [KMFL20] and non-local networks driven by fractional differential equations [AKLV20]. The above viewpoint can be extended to PDE architectures for input features that can be seen as grid functions.

More could also be said about numerical methods for continuous-time deep learning. An important question is whether first to optimize and then discretize (as, for example, in [CRBD18, GCB+18]) or first to discretize and then optimize (as, for example, done in [OFLR21]). In a first-optimize-then-discretize setting, one solves the adjoint ODE to compute the gradient of the loss function with respect to the weights, the key ingredient for optimization algorithms. While this allows some flexibility in choosing the numerical integrators for the forward and adjoint equation (e.g., one can use different step sizes), the adjoint method requires storing or recomputing the entire trajectory of the features, which is computationally infeasible. In a discretize-then-optimize approach, one selects a numerical time integrator and integration points to discretize the neural ODE (4) and obtains a finite-dimensional optimization problem. Differentiating the discretized loss function with respect to the weights is possible using automatic differentiation (also known as back-propagation) or analytically using the chain rule. The choice of the time integrator is crucial and provides opportunities to design novel network architectures that resemble ResNets but can be tailored to the network model; for example, one can mimic hyperbolic systems and use symplectic time integrators to ensure forward and backward stability [HR18] and save memory costs.

ACKNOWLEDGMENTS. I thank Levon Nurbekyan and Deepanshu Verma for their careful reading of initial drafts and fruitful discussions. I am also grateful for the two anonymous reviewers who made excellent suggestions that improved this paper. My work on this article and related projects was partly supported by NSF awards DMS 1751636, DMS 2038118, AFOSR grant FA9550-20-1-0372, and US DOE Office of Advanced Scientific Computing Research Field Work Proposal 20-023231.

### References

[AKLV20] Harbir Antil, Ratna Khatri, Rainald Löhner, and Deepanshu Verma, *Fractional deep neural network via constrained optimization*, Machine Learning: Science and Technology **2** (2020), no. 1, 015003.

[BCE+19] Martin Benning, Elena Celledoni, Matthias J. Ehrhardt, Brynjulf Owren, and Carola-Bibiane Schönlieb, *Deep learning as optimal control problems: models and numerical methods*, J. Comput. Dyn. **6** (2019), no. 2, 171–198, DOI 10.3934/jcd.2019009. MR4043228

[CRBD18] Ricky T Q Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud, *Neural Ordinary Differential Equations*, 201806.

[DDT19] Emilien Dupont, Arnaud Doucet, and Yee Whye Teh, *Augmented Neural ODEs* (2019), 3140–3150.

[E1703] Weinan E, *A Proposal on Machine Learning via Dynamical Systems*, Communications in Mathematics and Statistics **5** (201703), no. 1, 1–11.

[FJNO2011] Chris Finlay, Joern-Henrik Jacobsen, Levon Nurbekyan, and Adam Oberman, *How to Train Your Neural ODE: the World of Jacobian and Kinetic Regularization*, 202011, pp. 3154–3164.

[GCB+18] Will Grathwohl, Ricky T Q Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud, *FFJORD: Free-form Continuous Dynamics for Scalable Reversible Generative Models* (201810).

[HR18] Eldad Haber and Lars Ruthotto, *Stable architectures for deep neural networks*, Inverse Problems **34** (2018), no. 1, 014004, 22, DOI 10.1088/1361-6420/aa9a90. MR3742361

[HZRS1603] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, *Identity Mappings in Deep Residual Networks*, 201603, pp. 630–645.

[KMFL20] Patrick Kidger, James Morrill, James Foster, and Terry Lyons, *Neural controlled differential equations for irregular time series*, Advances in Neural Information Processing Systems **33** (2020), 6696–6707.

[LCTE17] Qianxiao Li, Long Chen, Cheng Tai, and Weinan E, *Maximum principle based algorithms for deep learning*, J. Mach. Learn. Res. **18** (2017), Paper No. 165, 29. MR3813814

[LFL+21] Alex Tong Lin, Samy Wu Fung, Wuchen Li, Levon Nurbekyan, and Stanley J. Osher, *Alternating the population and control neural networks to solve high-dimensional stochastic mean-field games*, Proc. Natl. Acad. Sci. USA **118** (2021), no. 31, Paper No. e2024713118, 10, DOI 10.1073/pnas.2024713118. MR4305642

[LLS23] Qianxiao Li, Ting Lin, and Zuowei Shen, *Deep learning via dynamical systems: an approximation perspective*, J. Eur. Math. Soc. (JEMS) **25** (2023), no. 5, 1671–1709, DOI 10.4171/jems/1221. MR4592857

[OFLR21] Derek Onken, Samy Wu Fung, Xingjian Li, and Lars Ruthotto, *OT-Flow: Fast and Accurate Continuous Normalizing Flows via Optimal Transport*, 2021, pp. 9223–9232.

[RH21] Lars Ruthotto and Eldad Haber, *An introduction to deep generative modeling*, GAMM-Mitt. **44** (2021), no. 2, Paper No. e202100008, 24, DOI 10.1002/gamm.202100008. MR4270461

[RMKK+92] Ramiro Rico-Martinez, K. Krischer, IG Kevrekidis, MC Kube, and JL Hudson, *Discrete-vs. continuous-time nonlinear signal processing of cu electrodissolution data*, Chemical Engineering Communications **118** (1992), no. 1, 25 –48.

[ROL+20] Lars Ruthotto, Stanley J. Osher, Wuchen Li, Levon Nurbekyan, and Samy Wu Fung, *A machine learning framework for solving high-dimensional mean field game and mean field control problems*, Proc. Natl. Acad. Sci. USA **117** (2020), no. 17, 9183–9193, DOI 10.1073/pnas.1922204117. MR4236167

[WYSO20] Bao Wang, Binjie Yuan, Zuoqiang Shi, and Stanley J. Osher, *Enresnet: Resnets ensemble via the feynman–kac formalism for adversarial defense and beyond*, SIAM Journal on Mathematics of Data Science **2** (2020), no. 3, 559–582, `https://doi.org/10.1137/19M1265302`.

[YK22] Liu Yang and George Em Karniadakis, *Potential flow generator with $L_2$ optimal transport regularity for generative models*, IEEE Trans. Neural Netw. Learn. Syst. **33** (2022), no. 2, 528–538. MR4378312

[ZK23] Benjamin J Zhang and Markos A Katsoulakis, *A mean-field games laboratory for generative modeling*, arXiv 2304.13534 (2023).

Lars Ruthotto

### Credits

All figures are courtesy of Lars Ruthotto.
Photo of Lars Ruthotto is courtesy of Nicole Alice Hart.

# A Feeling for Khovanov Homology

*Andrew Lobb*

One of the pleasures of studying knots and links is that, at some level, everybody understands what you are thinking about.

Formally speaking, a link is a smooth compact oriented 1-dimensional submanifold of the 3-sphere $\mathbb{S}^3$ or of 3-space $\mathbb{R}^3$, and a knot is a link of only one component. Some pictures of knots appear just below in Figures 1–6.

Informally speaking, a knot is just a knotted up piece of string with its ends glued together. Knots and links are most often considered up to isotopy, meaning up to smooth deformations that do not allow the passing of strands through one another. A fundamental question of knot theory is how to tell when two knots are different or really the same—given this knotted up piece of string and that knotted up piece of string can we, without getting out the scissors, arrange this one to look exactly like that one?

Another pleasure open to knot theorists is that their specialism is frequently enriched by contributions from outsiders. Perhaps this is not so surprising: knots give the simplest nontrivial examples of isotopy classes of submanifolds, so concepts from knot theory are encountered in other areas of mathematics and physics that have manifolds and submanifolds as underlying objects of interest. Similarly, concepts originating outside knot theory can often find a testing ground among knots and links.

To a knot theorist, this is all money for jam. A concept in knot theory is understood by physicists or algebraists, they come up with generalizations, these get fed back to you, and all of a sudden you have a whole new family of ideas to play with. Khovanov homology is the best example of such a concept.

At first sight, Khovanov homology seems a fairly innocuous link invariant—perhaps of use to those of us engaged in classifying isotopy classes of links, but with no obvious potential relative to other such invariants to break its bounds. However, in the quarter century or so since it appeared on the scene, Khovanov homology has infiltrated other areas of mathematics and physics in an unprecedented fashion, eclipsing in this way even its own direct forebear—the Jones polynomial.

*Andrew Lobb is a professor of mathematics at Durham University. His email address is* `andrew.lobb@durham.ac.uk`.



DOI: https://doi.org/10.1090/noti2928

**A feeling for Khovanov homology.** Anyone who has seen Holbein's painting *The Ambassadors* will know that the same object can present very differently when viewed from an alternative perspective. Khovanov homology is not an exception to this, and it is a measure of the ubiquity of the theory that it admits understanding from several disparate vantage points.

The purpose of this article is to give the casual reader, who may have a sandwich in one hand as they read, a feeling for Khovanov homology, so that they finish the article and the sandwich with the sense that they know *what kind of thing Khovanov homology is*. To achieve our ends we will approach Khovanov homology from several different directions and record what we see.

To each link $L$, Khovanov homology associates a finitely generated abelian group, $\mathrm{Kh}(L)$, invariant up to isotopy of the link. This splits as a direct sum along a bigrading $(i, j) \in \mathbb{Z} \times \mathbb{Z}$

$$\mathrm{Kh}(L) = \bigoplus_{(i,j)\in\mathbb{Z}\times\mathbb{Z}} \mathrm{Kh}^{i,j}(L).$$

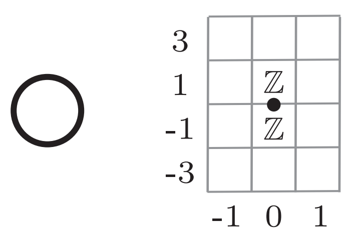



**Figure 1.** The unknot.

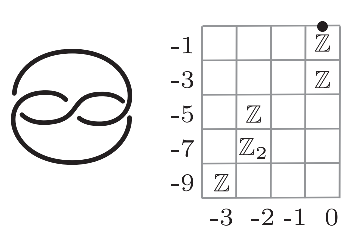



**Figure 2.** The left-handed trefoil.

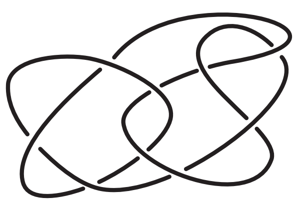
The simplest knots are the unknot, the left- and right-handed trefoils, and the figure eight knot—these are the first four knots $K$ whose homologies $\mathrm{Kh}^{i,j}(K)$ we have given in Figures 1–4, plotted in the $(i, j)$-plane. The $i$-grading is called the *homological grading*, while the $j$-grading is called the *quantum grading*. We write $\mathbb{Z}_n$ for the $n$-element cyclic group. So that the reader may get their bearings quickly, we have included a thick dot in bidegree $(i, j) = (0, 0)$.

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 9 | | | | $\mathbb{Z}$ |
| 7 | | | | $\mathbb{Z}_2$ |
| 5 | | | $\mathbb{Z}$ | |
| 3 | $\mathbb{Z}$ | | | |
| 1 | $\mathbb{Z}$ | | | |

**Figure 3.** The right-handed trefoil.

| | -2 | -1 | 0 | 1 | 2 |
|---|---|---|---|---|---|
| 5 | | | | | $\mathbb{Z}$ |
| 3 | | | | | $\mathbb{Z}_2$ |
| 1 | | | $\mathbb{Z}$ | $\mathbb{Z}$ | |
| -1 | | $\mathbb{Z}$ | $\mathbb{Z}$ | | |
| -3 | | $\mathbb{Z}_2$ | | | |
| -5 | $\mathbb{Z}$ | | | | |

**Figure 4.** The figure eight knot.



| | -4 | -3 | -2 | -1 | 0 | 1 | 2 |
|---|---|---|---|---|---|---|---|
| 7 | | | | | | | $\mathbb{Z}$ |
| 5 | | | | | | | $\mathbb{Z}_2$ |
| 3 | | | | | $\mathbb{Z}$ | $\mathbb{Z}$ | |
| 1 | | | | $\mathbb{Z}$ | $\mathbb{Z}\oplus\mathbb{Z}_2$ | | |
| -1 | | | | $\mathbb{Z}\oplus\mathbb{Z}_2$ | $\mathbb{Z}$ | | |
| -3 | | $\mathbb{Z}$ | $\mathbb{Z}$ | | | | |
| -5 | | $\mathbb{Z}_2$ | | | | | |
| -7 | $\mathbb{Z}$ | | | | | | |

**Figure 5.** The knot $9_{42}$.

The forty-second nine-crossing knot to appear in Rolfsen's knot table, $9_{42}$, is well-known in certain quarters as being the first knot in the table which nevertheless exhibits generic characteristics from various points of view. We have given the homology of $9_{42}$ in Figure 5.

Each of these five knots only admits one orientation up to isotopy, so we have omitted to orient the knots.

## Phenomenology

Most mathematicians are phenomenologists. When trying to understand something, we often think about a toy case, then a toy case with bells on, then a toy case with bells and whistles, before we finally allow ourselves to think generally.

Above we see several examples of homology groups, and just by looking at them the reader will have started to get a feeling for Khovanov homology. What strikes us about these examples? Let's go through them.

Firstly, looking at Figure 1, we see that the homology of the unknot is 2-dimensional and is supported in bidegrees $(0,-1)$ and $(0,1)$. If we were making a guess of what the homology of the unknot ought to be, we might guess that it should be 1-dimensional and supported in bidegree $(0,0)$. This guess is correct for the variant of Khovanov homology called *reduced* Khovanov homology, but we are going to stick with vanilla flavored Khovanov homology for this exposition.

So much for the unknot, let's take a look at the trefoils in Figures 2 and 3. If you were to commission the carving of the following woodblock stamps, then you would be able to use them to produce the homology tables of the unknot, of the left-handed trefoil, and of its mirror image the right-handed trefoil.

| |
|---|
| $\mathbb{Z}$ |
| $\mathbb{Z}$ |

| | |
|---|---|
| | $\mathbb{Z}$ |
| | $\mathbb{Z}_2$ |
| $\mathbb{Z}$ | |

The block on the left is all that you would need to make the homology of the unknot, whereas you would also need the block on the right for the trefoils. This latter is sometimes called the *knight's move* block by chess-playing mathematicians.

We also note that the homological degrees $i$ which support the homology run from $-3$ to $0$ for the left-handed trefoil and from $0$ to $3$ for the right-handed trefoil.

If you choose orientations for the trefoils you will see that the left-handed trefoil has three negative crossings in the given diagram while the right-handed trefoil has three positive crossings, following the conventions below.


To add weight to our musings, the unknot has 0 crossings of either sign and its homology is entirely supported in homological grading $i = 0$. So we might guess that the homological support has something to do with the numbers of positive and negative crossings in a minimal crossing-number diagram.

Finally, there seems to be some sort of duality going on between the left-handed and the right-handed trefoil. These knots themselves are dual in the sense that one is the mirror image of the other. In their homology tables we see that, up to a bit of messing about with the $\mathbb{Z}_2$ summand (which you might hope to explain by some sort of universal coefficient theorem), the homology of one looks like it is obtained from the homology of the other by rotation by $\pi$ around the thick dot that marks the bidegree $(0,0)$.

Moving now to the figure eight knot in Figure 4, we see that our woodblocks still stand us in good stead—we need to use the unknot block once and then the knight's move twice.

Moreover, the homology of the figure eight knot is supported between gradings $i = -2$ and $i = 2$, matching the two negative and two positive crossings in the diagram. And finally the homology has a self-duality around $(0,0)$, which we would expect once we are told that the figure eight knot is its own mirror image.

Next we turn to the knot known as $9_{42}$, pictured in Figure 5. It takes a little bit of scrutinizing, but we see that our woodblocks still suffice. At this point we notice that the unknot block is not getting as much use as the knight's move. For all the examples so far, the unknot block gets used exactly once while the rest of the homology is filled up with knight's moves. We further notice that the unknot block always appears in grading $i = 0$, but its quantum grading $j$ can slide up or down.

These woodblocks, which we commissioned as a time-saving device, are starting to seem quite important. Do they remind us of anything? Relating the groups to the usual singular homology of a space, the unknot block has the total homology of a sphere, while the knight's move block has the total homology of $\mathbb{RP}^3$. We do not see yet what these spaces might have to do with knots or their Khovanov homology.

The homology of $9_{42}$ is supported between degrees $i = -4$ and $i = 2$, while the diagram given has four negative and five positive crossings. So the support of the homology cannot be telling us exactly how many negative and positive crossings there are in a minimal crossing-number diagram of the knot, but maybe it is giving us lower bounds on these quantities?

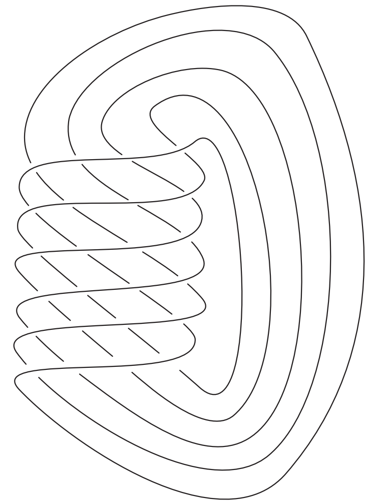
Let's end our phenomological tour by looking at a larger knot—the $(5, 6)$ torus knot $T_{5,6}$—which has 24 positive and zero negative crossings in its minimal diagram $D_{5,6}$ given in Figure 6.

We have not included a thick dot in bidegree $(0, 0)$ in the interests of space. The homology is supported between degrees $i = 0$ and $i = 14$, so our guess about the support possibly providing a lower bound on the numbers of positive and negative crossings in a diagram of the knot is not yet contradicted.

Unfortunately, the reader will notice the groups $\mathbb{Z}_3$ and $\mathbb{Z}_5$ appearing as summands in the higher homological degrees. Before we destroy our woodblocks in frustration, let us also notice that there is still an unknot block summand appearing in degree $i = 0$ and there are also several instances of the knight's move appearing among more general configurations.

And what else? The middle of the unknot block in degree $i = 0$ is placed at quantum degree $j = 20$. Since every second $j$-degree seems to support no homology, this is morally a jump of 10 above $j = 0$. If we look up $T_{5,6}$, the first topological measure of its complexity is its *genus* which is $g(T_{5,6}) = 10$. The genus of a knot is the minimal genus of a compact orientable surface in $\mathbb{S}^3$ whose boundary is the knot. Is this just a coincidence? Looking at the homologies of the earlier knots we have studied tells us that the height of the unknot block in degree $i = 0$ is *not* just recording the genus of the knot, although this does seem to be the case for the genus and homology of any other positive torus knot that we care to look up.

Finally, we should record that in all examples there seems to be a certain *diagonalness* to Khovanov homology, with the support of the homology very roughly lining up along a line from bottom left to top right. There does not seem to be an expected slope nor, in general, an expected $y$-intercept.



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | | | | | | | | | | | | | | | $\mathbb{Z}_3$ |
| 41 | | | | | | | | | | | | | $\mathbb{Z}$ | | |
| 39 | | | | | | | | | | | | | $\mathbb{Z}_2 \oplus \mathbb{Z}_5$ | $\mathbb{Z}$ | |
| 37 | | | | | | | | | | | $\mathbb{Z}_2$ | $\mathbb{Z}^2$ | $\mathbb{Z}$ | $\mathbb{Z}$ | |
| 35 | | | | | | | | | | $\mathbb{Z}^2$ | $\mathbb{Z}_2$ | $\mathbb{Z}_2 \oplus \mathbb{Z}_5$ | $\mathbb{Z}$ | | |
| 33 | | | | | | | | $\mathbb{Z}$ | | $\mathbb{Z} \oplus \mathbb{Z}_2$ | $\mathbb{Z}$ | | | | |
| 31 | | | | | | $\mathbb{Z}$ | | $\mathbb{Z} \oplus \mathbb{Z}_2$ | $\mathbb{Z}^2$ | | | | | | |
| 29 | | | | | | $\mathbb{Z}$ | $\mathbb{Z}$ | $\mathbb{Z}_2$ | $\mathbb{Z}$ | | | | | | |
| 27 | | | | $\mathbb{Z}$ | $\mathbb{Z}$ | | $\mathbb{Z}$ | | | | | | | | |
| 25 | | | | $\mathbb{Z}_2$ | $\mathbb{Z}$ | | | | | | | | | | |
| 23 | | | $\mathbb{Z}$ | | | | | | | | | | | | |
| 21 | $\mathbb{Z}$ | | | | | | | | | | | | | | |
| 19 | $\mathbb{Z}$ | | | | | | | | | | | | | | |

**Figure 6.** The diagram $D_{5,6}$ of the $(5, 6)$ torus knot $T_{5,6}$ and the Khovanov homology of $T_{5,6}$.

Now that we have started to get a feeling for Khovanov homology, let's ask for its historical meaning by discussing the context in which it was discovered.

## History of the Discovery

Mikhail Khovanov put a paper [Kho00] introducing his homology theory on the arXiv at the tail end of the 20th century, a couple of years after completing his PhD at Yale.

Khovanov homology may be viewed as an early exemplar of the success of the program known as *categorification*, a term coined by Louis Crane. Khovanov's PhD advisor Igor Frenkel was another originator and proponent of the nascent subject, and collaborated in a seminal paper with Crane [CF94]. Categorification was originally conceived with a somewhat precise meaning, but is now more often understood as an umbrella term used to refer to various lifts of structure. It is one of those things of which you may say, after having become acquainted with a few examples, that you know it when you see it.

The ideas around categorification included from quite early on the belief that one should be able to categorify the representation theory of quantum groups, a so-far successful enterprise that is still employing mathematicians today. The representation theory of quantum $\mathfrak{sl}_2$ yields up the Jones polynomial $V(L) \in \mathbb{Z}[q, q^{-1}]$, which is an invariant of links $L$ whose discovery earned Vaughan Jones a Fields medal. So, it was believed that the Jones polynomial should eventually admit a "lift" to some stronger invariant, but perhaps this lift would only be accessible a little further down the line once the groundwork had been laid by categorifying the underlying algebraic structure.

On the other hand, as well as arising from quantum representation theory, the Jones polynomial has a particularly simple definition provided by Kauffman [Kau87]. In this reformulation, given a diagram of the link, its Jones polynomial is expressed as an alternating sum of Laurent polynomials in $q$, each of which has nonnegative integer coefficients.

Once you know to listen for it, this definition strongly echoes the notion of Euler characteristic. The Euler characteristic is a topological invariant expressed as an alternating sum of nonnegative integers (the numbers of simplices or cells of each dimension) which are themselves not topological invariants. In what might now be described by the more excitable at the departmental teatime as one of the earliest examples of categorification, one realizes that these numbers are nonnegative because they ought to be interpreted as the dimension of a vector space (or as the rank of an abelian group). These vector spaces or groups fit into a chain complex, of which the homology groups turn out to be strong topological invariants.

Khovanov saw that one could attempt to play the same game with Kauffman's definition of the Jones polynomial. The plan was to leap-frog the business of categorifying quantum representation theory and to jump straight to the answer—a homology theory whose Euler characteristic, appropriately defined, should be the Jones polynomial.

**The quantum world.** The philosophy of categorification would suggest that such a putative homology theory would have something to tell us about 4-dimensional topology and might be a quantum competitor or counterpart to the gauge and Floer theoretic invariants that we shall shortly come to discuss.

When we use the word *quantum* here, we are not thinking of quantum mechanics. Rather we use the term to refer to a collection of combinatorial topological invariants arising out of the representation theory of quantum groups. A quantum group is not a group, but rather a noncommutative algebra which may be thought of as a perturbation by an extra parameter $\hbar$ of the universal enveloping algebra of a Lie algebra. Specializing by setting $\hbar = 0$ recovers the universal enveloping algebra. So we see that the terminology arises by analogy with the fact that setting Planck's constant to zero recovers classical physics from quantum physics.

The reader intent on acquiring a feeling for Khovanov homology should think of two streams of invariants—the analytically defined gauge and Floer theoretic invariants, and the combinatorial and algebraic quantum invariants. The confluence of these two apparently distinct streams will form part of our narrative.

## Khovanov's Chain Complex

Khovanov's construction of a chain complex from a link diagram is so elegant that it has an air of inevitability about it. The original paper [Kho00] is very concrete and readable by anyone who has seen a modicum of homological algebra.

**The Khovanov cube.** We present Khovanov's construction in the case of a link diagram of sufficient complexity so that the reader can recover the general case. In fact, we start with a diagram of the unknot known to the cognoscenti as the *trefail*, pictured in Figure 7.

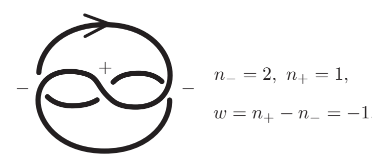
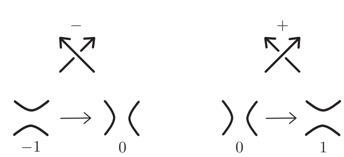



**Figure 7.** The *trefail* diagram of the unknot.

We have included an orientation so that we can collect some sign data from the crossings, following the conventions below. We write $n_- = 2$ for the number of negative crossings, and $n_+ = 1$ for the number of positive crossings. The quantity $w = n_+ - n_- = -1$ is known as the *writhe* of the diagram.




Below the positive and negative crossings you see two *smoothings* of these crossings. Each smoothing is associated with an integer—either −1, 0, or 1. We are going to be interested in smoothings of the diagram which arise from picking a smoothing at each crossing. Each diagram smoothing is nothing more than a collection of circles embedded in the plane.

Since the diagram has three crossings, there are $2^3 = 8$ possible diagram smoothings. Picking an ordering on the negative crossings enables us to associate each diagram smoothing with a vertex of the cube

$$[-1,0]^2 \times [0,1] = [-1,0]^{n_-} \times [0,1]^{n_+}.$$

In Figure 8 we have put the vertex $(-1,-1,0)$ on the far left and the opposite vertex $(0,0,1)$ on the far right, and the edges are oriented from left to right. Moving along any edge will increase exactly one coordinate of the cube by 1, leaving the others fixed. So we see that the coordinate sum of the vertices ranges between $-2 = -n_-$ and $1 = n_+$. The cube is arranged so that vertices with the same coordinate sum lie in a vertical line in the page.

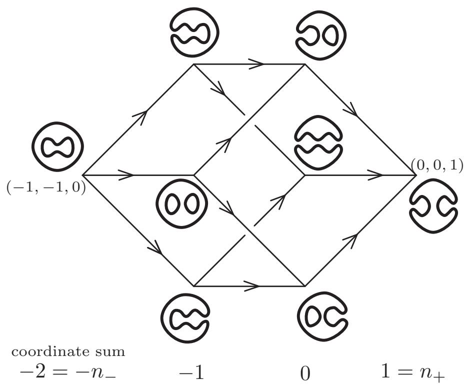



**Figure 8.** The smoothings of the trefail diagram decorate the vertices of a cube.

To pass to algebra, our fundamental building block will be the free 2-dimensional $\mathbb{Z}$-module

$$V = \langle v_-, v_+ \rangle.$$

The module $V$ comes graded by the quantum $j$-grading that we have already encountered. We think of $v_-, v_+$ as being homogeneous of gradings $j = -1, +1$ respectively. A useful shorthand to record this is to write

$$\mathrm{qdim}(V) = q^{-1} + q$$

for the so-called *quantum dimension* of $V$ in which the power of $q$ records the quantum grading and the coefficient records the dimension. In other contexts this shorthand is referred to as the Poincaré polynomial.

To create each chain group we shall be applying three algebraic operations to copies of $V$; these operations are tensor product, direct sum, and quantum grading shift. The quantum grading shift is exactly as it sounds, and is written by appending a square bracket with the degree of the shift. Concretely, we have for each free $j$-graded module $W$

$$\mathrm{qdim}(W[n]) = q^n \,\mathrm{qdim}(W).$$

The chain groups $\mathrm{CKh}^i$ of the Khovanov chain complex coming from the trefail diagram of the unknot are given in Figure 9.

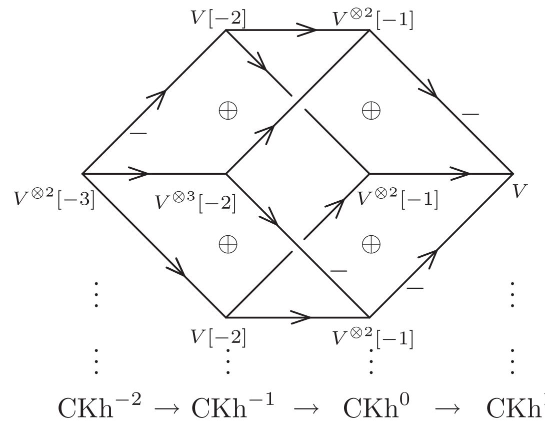



**Figure 9.** The chain groups arising from the trefail diagram.

Recall that we started by decorating each corner of the cube by a smoothing of the trefail diagram. Now each such smoothing has been replaced by a tensor power of $V$ together with some shift. The tensor power corresponds to the number of components of the smoothing, while the shift is by $i + w$ where $i$ is the coordinate sum and $w = -1$ is the writhe. Finally, the chain group in homological degree $i$ is given by taking the direct sum of the groups on corners with coordinate sum $i$.

This decorated cube is known as the *Khovanov cube*, which is a name also applied to the earlier cube of smoothings. It is useful in thinking about the chain complex to bear both in mind.

**The Jones polynomial.** At this point we pause to take stock. We do not yet have a chain complex because we have not given the chain maps between the chain groups

$$\partial^i_{\mathrm{Kh}} : \mathrm{CKh}^i \longrightarrow \mathrm{CKh}^{i+1}.$$

But no matter what chain maps we give, it will of course not affect the Euler characteristic, since this should just be the alternating sum of the dimensions of the chain groups. We note that each chain group is quantum graded. If we give chain maps that preserve the quantum grading then the chain complex will split as a direct sum of chain complexes, one complex for each quantum grading.

$$\mathrm{CKh}^i = \bigoplus_j \mathrm{CKh}^{i,j},$$

$$\partial^i_{\mathrm{Kh}} : \mathrm{CKh}^{i,j} \longrightarrow \mathrm{CKh}^{i+1,j}.$$

If we compute

$$\sum_i (-1)^i \,\mathrm{qdim}(\mathrm{CKh}^i) = \sum_{i,j} (-1)^i q^j \dim(\mathrm{CKh}^{i,j})$$

then we will get a Laurent polynomial in $q$ in which the coefficient of $q^j$ will just be the Euler characteristic of the chain complex summand in quantum degree $j$.

This Laurent polynomial is in fact the *Jones polynomial* of the link, and its expression as an alternating sum of Laurent polynomials in $q$ with nonnegative integer coefficients is exactly Kauffman's reformulation of the Jones polynomial.

Now it is clear where the construction up to this point has come from. If you believe that Kauffman's alternating sum is actually computing the Euler characteristic of a chain complex and you want to guess the chain groups, then a natural choice would be the one that we have given above.

**The differential.** Of course, it is well and good to imagine that we can make a chain complex this way, but to do so we still need to give the differential. We cannot cheat by just using the zero differential, because the aim is to produce homology groups that do not depend on the chosen diagram.

To specify a differential $\partial^i_{\mathrm{Kh}}\colon\ \mathrm{CKh}^i \to \mathrm{CKh}^{i+1}$, we decorate each edge of the Khovanov cube with a map, and then $\partial^i_{\mathrm{Kh}}$ is the direct sum of all those maps decorating edges connecting the summands of $\mathrm{CKh}^i$ to those of $\mathrm{CKh}^{i+1}$.

Looking back at the cube of smoothings, we recall that as we travel along an edge either two circles of a smoothing are being joined into one, or a single circle is being split into two. Each circle of a smoothing corresponds to a tensor factor $V$. So in order to give the edge maps we first write down maps $m\colon V\otimes V \to V$ and $\Delta\colon V \to V\otimes V$ and extend by the identity map on those tensor factors $V$ corresponding to circles that do not get merged or split as we move along an edge.

$$\begin{aligned} m\colon V\otimes V \to V\colon\ & v_-\otimes v_- \mapsto 0,\\ & v_-\otimes v_+ \mapsto v_-,\\ & v_+\otimes v_- \mapsto v_-,\\ & v_+\otimes v_+ \mapsto v_+;\\ \Delta\colon V \to V\otimes V\colon\ & v_- \mapsto v_-\otimes v_-,\\ & v_+ \mapsto v_-\otimes v_+ + v_+\otimes v_-. \end{aligned}$$

These edge maps that we have now specified make each square face of the cube *commute* (this should not be obvious, but it is a relatively easy check). In fact, $m$ and $\Delta$ are one of the very few nontrivial choices that achieve this, which may explain how they were chosen. In order that the differential squares to zero, we want rather that each face of the cube should *anti-commute*, so that travelling one way around any face gives the negative of travelling the other way.

To turn the commutative faces into anticommutative faces we replace some (in this case, four) of the edge maps with their negative, as indicated by the minus signs in the algebraic cube diagram. It can be checked that any selection of negative edges achieving anticommutativity of faces results in a chain homotopic complex.

Finally, note that both $m$ and $\Delta$ are quantum $j$-graded of degree $-1$. Since there is a quantum shift by $i+w$ on the chain group summands, it follows that $\partial^i_{\mathrm{Kh}}$ preserves $j$, as we wanted.

It remains to verify that the resulting homology does not depend on the diagram of the link chosen. Since any two link diagrams of the same link are related by a finite sequence of Reidemeister moves, this amounts to finding a chain homotopy equivalence between two diagrams differing by such a move.

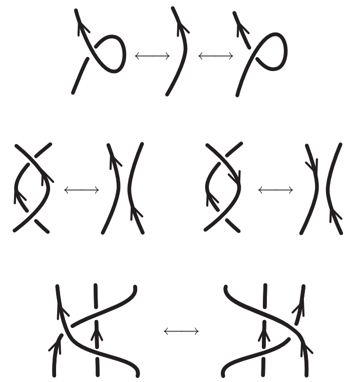


**Figure 10.** The oriented Reidemeister moves.

More precisely, the *oriented Reidemeister moves* are pictured in Figure 10. To each double-headed arrow Khovanov associated a chain homotopy equivalence between the chain complexes of diagrams differing locally by the moves.

**Theorem 1** ([Kho00]). *Suppose that $D$ is a diagram of a link $L$. The chain complex $\mathrm{CKh}^{i,j}(D)$ associated to $D$ is invariant under Reidemeister moves up to chain homotopy equivalence. Consequently, the homology groups $\mathrm{Kh}^{i,j}(L)$ are an invariant of $L$. The graded Euler characteristic*

$$\sum_{i,j}(-1)^i q^j\,\mathrm{rk}(\mathrm{Kh}^{i,j}(L))$$

*coincides with the Jones polynomial of $L$.*

If the aim of Khovanov's paper were simply to produce a bigraded abelian group whose graded Euler characteristic coincides with the Jones polynomial then he could have simply started with the Jones polynomial, forgotten about the link, and defined a bigraded group from there. However, Khovanov homology is strictly stronger than the Jones polynomial (for example the knots $5_1$ and $10_{132}$ have the same Jones polynomial but different Khovanov homologies). And, as we shall soon see, Khovanov homology is more than just an improvement on the Jones polynomial as an isotopy invariant.

**Observations.** Now that we have a definition, let's see whether we can explain some of the phenomenological features of Khovanov homology that we recorded earlier.

From the definition we can immediately see that if the homology of a link is nonzero in degree $i=m$ then any diagram of the link must have at least $-m$ negative crossings if $m<0$, and at least $m$ positive crossings if $m>0$. So this bears out some of our observations on the homological support of the homology.

With a little bit of work, the reader would be able to convince themselves that the complex associated to the mirror of a link diagram is isomorphic to the dual of the complex associated to the unmirrored diagram. This explains the duality in the homology that we observed between the left- and right-handed trefoils, and the self-duality in the homology of the figure eight knot.

What about the copy of the homology of the unknot that seems to appear in homological degree $i = 0$? Let's think about the complex associated to the diagram $D_{5,6}$ that we drew above of the torus knot $T_{5,6}$.

Firstly, $D_{5,6}$ has 24 positive crossings, meaning that the complex $\mathrm{CKh}^i(D_{5,6})$ runs from degree $i = 0$ to degree $i = 24$. We can identify the chain group in degree $i = 0$ very quickly. It is just the group that decorates the vertex $(0, 0, \dots, 0)$ of the Khovanov cube. To construct this group we first take the 0-smoothing of every crossing of $D_{5,6}$. This gives us five nested circles in $\mathbb{R}^2$. Remembering to shift by $i + w = 0 + 24 = 24$ we see that

$$\mathrm{CKh}^0(D_{5,6}) = V^{\otimes 5}[24].$$

Since $\mathrm{CKh}^{-1}(D_{5,6}) = 0$, we have

$$\mathrm{Kh}^0(T_{5,6}) = \ker(\partial^0_{\mathrm{Kh}} : \mathrm{CKh}^0(D_{5,6}) \to \mathrm{CKh}^1(D_{5,6})).$$

Thinking about the vertices of the Khovanov cube in degree $i = 1$, we note that each of them is decorated by 4 circles, arising from merging a pair of the 5 nested circles on the vertex $(0, 0, \cdots, 0)$. Looking back at the definition of the map $m$ associated to edges of the Khovanov cube along which circles merge, we see that the element

$$v_-^{\otimes 5} \in V^{\otimes 5}[24] = \mathrm{CKh}^0(D_{5,6})$$

must be in the kernel of $\partial^0_{\mathrm{Kh}}$ and hence represents a nontrivial homology class in the 0th homology group.

The quantum $j$-grading of this element is $24 - 5 = 19$, so this corresponds to the copy of $\mathbb{Z}$ that appears at $(0, 19)$ in the table of $\mathrm{Kh}^{i,j}(T_{5,6})$. The one task that this article will leave to the reader, in between bites of their sandwich, is to find the generator of the copy of $\mathbb{Z}$ at $(0, 21)$.

Of course, there was nothing particularly special about $T_{5,6}$ among all torus knots. If we were to repeat the arguments for the torus knot $T_{p,q}$ where $0 < p < q$ with $p, q$ coprime, we would see copies of $\mathbb{Z}$ in bidegrees $(0, (p-1)(q-1) \pm 1)$. In other words we would find the unknot homology but shifted up by $(p-1)(q-1)$. This matches twice the knot genus

$$g(T_{p,q}) = \frac{(p-1)(q-1)}{2},$$

although we might make the mistake, at this point in our tour of Khovanov homology, of discounting this as a coincidence rather than recognizing it as foreshadowing.

## Floer Homology

It was recognized almost immediately that Khovanov homology had some similarities with homological invariants of 3-manifolds known as *Floer* homologies. Although combinatorial formulations of several Floer homologies have appeared, these invariants are, at heart, analytic: the differentials of their chain complexes count solutions to differential equations, even if combinatorial means might be found to make the count.

Floer homologies of 3-manifolds arose in the last two decades of the twentieth century from the study of gauge-theoretic invariants of smooth closed 4-manifolds. These gauge theoretic invariants notably include the Donaldson invariant from instanton gauge theory or the Seiberg-Witten invariant from monopole gauge theory.

Removing two open balls from a closed 4-manifold $X$ turns it into a cobordism $\mathring{X}$ from $\mathbb{S}^3$ to $\mathbb{S}^3$. It was realized that the gauge theoretic invariants of the closed manifold are determined by a map on the relevant Floer homology of $\mathbb{S}^3$ (which is always a very simple algebraic structure such as a polynomial ring) induced by the cobordism

$$\mathring{X}_* : H_*^{\mathrm{Floer}}(\mathbb{S}^3) \longrightarrow H_*^{\mathrm{Floer}}(\mathbb{S}^3).$$

Essentially one reproduces the invariants of closed 4-manifolds by using the functoriality of Floer homology.

It is interesting to note the progression in this case was by jumping down from dimension 4 to dimension 3. There is much effort at the moment in trying to jump further down the dimensions by associating invariants of some kind to 2–, 1– , and eventually to 0–manifolds. The idea is to attempt to understand known examples of (3+1)-dimensional topological quantum field theories in the spirit of the cobordism hypothesis laid out by Baez and Dolan [BD95].

From its inception it was expected, in line with the program of categorification, that Khovanov homology should not just be a strengthening of the essentially 3-dimensional Jones polynomial, but should also give rise to 4-dimensional invariants. As we have already mentioned, the Jones polynomial being the Euler characteristic of Khovanov homology suggests singular homology as a useful parallel. Singular homology not only strengthens the Euler characteristic invariant of a space but is functorial for continuous maps. The correct analogue of continuous map turns out to be *link cobordism*.

A link cobordism between the links $L_0, L_1 \subset \mathbb{S}^3$ is a smooth embedding of a compact oriented surface $\Sigma \subset \mathbb{S}^3 \times [0,1]$ which satisfies $\partial\Sigma = L_0 \times \{0\} \cup L_1 \times \{1\}$. We give a picture of this in Figure 11. In fact, the object of interest is usually the isotopy class of a link cobordism, in which we allow smooth deformations of $\Sigma$ while keeping the boundary links fixed. Such a cobordism induces a map

$$\Sigma_* : \mathrm{Kh}^{i,j}(L_0) \longrightarrow \mathrm{Kh}^{i,j+\chi(\Sigma)}(L_1),$$

which preserves the homological $i$-grading and shifts the quantum $j$-grading by the Euler characteristic of the surface.

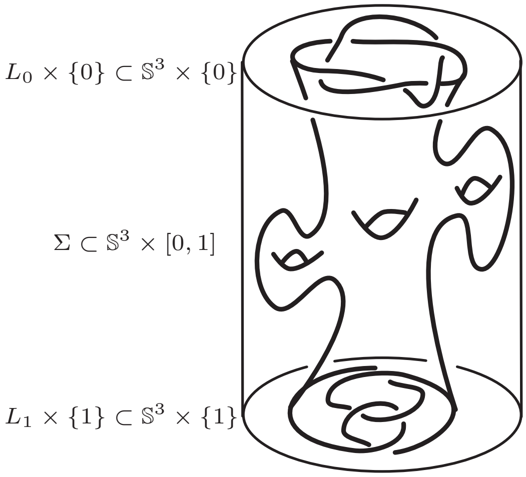



**Figure 11.** A smooth link cobordism $\Sigma$ from $L_0$ to $L_1$.

The map $\Sigma_*$ is constructed combinatorially, beginning with a suitable description of $\Sigma$. Any such cobordism $\Sigma$ can be described by a so-called *movie*, which is a finite sequence of link diagrams (the *frames* of the movie), starting with a diagram $D_0$ for $L_0$ and ending with a diagram $D_1$ for $L_1$. Successive diagrams in the list differ either by an oriented Reidemeister move or by one of the three *Morse moves*, drawn in Figure 12, which represent a local maximum, a local minimum, or a saddle point of $\Sigma$.

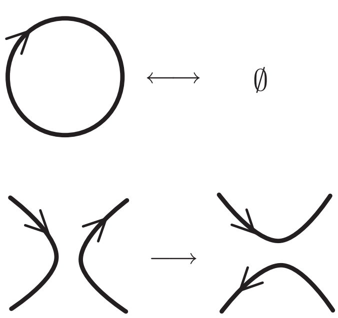



**Figure 12.** The three Morse moves.

To each Morse move there is an associated chain map of Khovanov chain complexes. The chain maps corresponding to maxima and minima are quantum $j$-graded of degree $+1$, while the saddle chain map is quantum $j$-graded of degree $-1$. And in his original paper Khovanov already gave chain homotopy equivalences for each Reidemeister move. Composing all the chain maps coming from Reidemeister and Morse moves in a movie of $\Sigma$, one obtains a chain map

$$\Sigma_\# : \mathrm{CKh}^{i,j}(D_0) \longrightarrow \mathrm{CKh}^{i,j+\chi(\Sigma)}(D_1).$$

**Theorem 2** ([Jac04, CMW09]). *The induced map*

$$\Sigma_* : \mathrm{Kh}^{i,j}(L_0) \longrightarrow \mathrm{Kh}^{i,j+\chi(\Sigma)}(L_1)$$

*on homology is an invariant of $\Sigma$ up to isotopy.*

To prove this theorem, one essentially verifies that equivalent movies define chain homotopic chain maps.

**Heegaard-Floer homology.** From the start, then, 4-dimensional functoriality provided at least a superficial similarity between Khovanov homology and Floer homology. Back at the turn of the century, however, one could still imagine a world in which quantum homological invariants (*i.e.*, generalizations of Khovanov homology) and Floer homological invariants, although similar, would continue to develop in parallel but were essentially unrelated, in the same way that the Jones polynomial and the Alexander polynomial were thought of as being very distinct. This is not how it has turned out.

The meeting of quantum and Floer homological invariants was first presaged by a result relating the Khovanov homology of a link $L$ to the *Heegaard-Floer homology* of its *branched double cover* $M(L)$.

Heegaard-Floer homology is a package of homological invariants; in its simplest incarnation it is an invariant of a closed 3-manifold $M$ that takes the form of a singly-graded $\mathbb{Z}_2$-vector space $\widehat{\mathrm{HF}}(M)$.

Avoiding a formal definition, the branched double cover $M(L)$ is a closed 3-manifold that admits an everywhere 2-to-1 map $M(L) \to \mathbb{S}^3$ apart from at points of $L \subset \mathbb{S}^3$ where it is 1-to-1. Its topology is intimately connected with that of $L$. Ozsváth and Szabó, the progenitors of Heegaard-Floer homology, proved the following result.

**Theorem 3** ([OS05]). *There is a* spectral sequence *from the Khovanov homology of $L$ with $\mathbb{Z}_2$-coefficients to $\widehat{\mathrm{HF}}(-M(L))$.*

Here we have written $-M(L)$ for the branched double cover but with reversed orientation. For readers unfamiliar with spectral sequences, the important content for us is that it means that there is an extra differential on $\mathrm{Kh}(L;\mathbb{Z}_2)$ whose homology gives $\widehat{\mathrm{HF}}(-M(L))$.

This theorem came early on in an era of results that established the existence of spectral sequences between various quantum homological invariants, and from quantum homological invariants to Floer theoretic invariants. Often these spectral sequences were more structural rather than being proved with any concrete topological application in mind. The result is a web of interdependency that ties Khovanov homology and Floer homology tightly together.

What we would like the reader to take from this is that Khovanov homology may be thought of as a first-order combinatorial approximation to more analytic invariants.

**Instanton Floer homology.** Instanton Floer homology is another invariant of a pair $(M, L)$ where $M$ is a 3-manifold and $L \subset M$ is a 1-dimensional submanifold, taking the form of a graded abelian group $\mathrm{I}(M, L)$. Kronheimer and Mrowka observed a curious coincidence which led them to suspect the existence of a spectral sequence from $\mathrm{Kh}(L)$ to $\mathrm{I}(\mathbb{S}^3, L)$ [KM11].

The chain group for $\mathrm{I}(M, L)$ is generated by a perturbed space of flat $\mathrm{SU}(2)$ connections on $(M, L)$ which have a prescribed singularity along $L$. Since there is a correspondence between flat connections and representations of the fundamental group, to get a handle on the chain group for

$\mathrm{I}(\mathbb{S}^3, L)$ we consider the space of homomorphisms

$$\mathrm{Rep}(L) \coloneqq \{\rho :\ \pi_1(\mathbb{S}^3 \setminus L) \to \mathrm{SU}(2) \mid (*)\},$$

where $(*)$ is the condition that $\rho$ should send every meridional element of $\pi_1(L)$ to an element of SU(2) of trace $\mathrm{tr} = 0$.

Topologically, SU(2) is diffeomorphic to the 3-sphere (see Figure 13). The trace function

$$\mathrm{tr} :\ \mathrm{SU}(2) \longrightarrow [-2, 2]$$

has a unique maximum at the identity matrix $I_2$ and a unique minimum at $-I_2$. The traceless matrices are the equatorial 2-sphere $\mathrm{tr}^{-1}(\{0\}) = \mathbb{S}^2$.

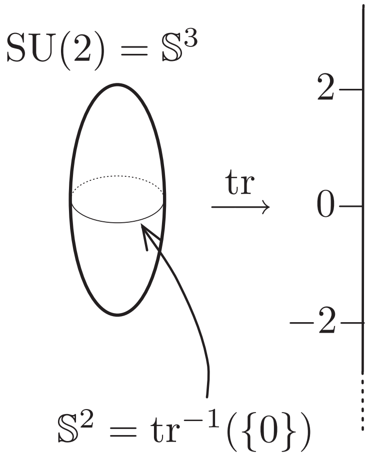



**Figure 13.** The trace function on $\mathbb{S}^3 = \mathrm{SU}(2)$.

In this example, $K$ will be the left-handed trefoil. In Figure 14 we give a diagram giving three elements $a, b, c \in \pi_1(\mathbb{S}^3 \setminus K)$. These are known as *meridional* elements since, forgetting the basepoint, each is homotopic in the knot complement to a small meridional loop to the knot. We have drawn one for each of the three arcs that form the knot diagram. Each of the arcs is incident to each of the crossings. We have labelled the leftmost crossing with a bullet •. If the reader performs a mental isotopy and slides the generators $a, b, c$ towards the crossing •, they will be able to see that

$$b = c^{-1}ac.$$

There are similar relations coming from the other two crossings. In fact, these generators and relations give the group

$$\pi_1(\mathbb{S}^3 \setminus K) = \langle a, b, c | b = c^{-1}ac, c = a^{-1}ba, a = b^{-1}cb\rangle.$$

An element $\rho \in \mathrm{Rep}(K)$ is determined by the images $\rho(a), \rho(b), \rho(c) \in \mathbb{S}^2 = \mathrm{tr}^{-1}(0)$. This is not a completely free choice because we need to make sure that the three group relations are satisfied. Unpacking the equation $\rho(b) = \rho(c)^{-1}\rho(a)\rho(c)$, it turns out to be equivalent to requiring that $\rho(a), \rho(b), \rho(c)$ each lie on the same great circle geodesic of $\mathbb{S}^2$, with $\rho(c)$ lying halfway between $\rho(a)$ and $\rho(b)$.

Since there are two further relations, $\mathrm{Rep}(K)$ corresponds to choices of three points $\rho(a), \rho(b), \rho(c) \in \mathbb{S}^2$ which are equidistantly spaced around a great circle of $\mathbb{S}^2$ (see Figure 15).

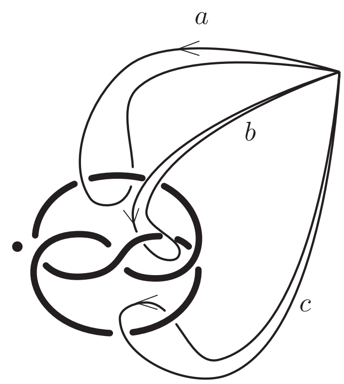



**Figure 14.** Generators for $\pi_1(\mathbb{S}^3 \setminus K)$.

It follows that $\mathrm{Rep}(K)$ falls into two connected components. One of the components parametrizes representations satisfying $\rho(a) = \rho(b) = \rho(c) \in \mathbb{S}^2$ and is topologically just $\mathbb{S}^2$. Points in the other component correspond to choosing a great circle and three nonequal but equidistant points along it. This is equivalent to the choice of the point $\rho(a) \in \mathbb{S}^2$ along with a unit tangent vector in the tangent space $T_{\rho(a)}\mathbb{S}^2$. Specifically, there is a unique great circle running in the direction of the tangent vector, and we place $\rho(b)$ a third of the way along the great circle in this direction, and then $\rho(c)$ two thirds of the way along. In other words, this second component is topologically just the unit tangent bundle to $\mathbb{S}^2$. This is a circle bundle over $\mathbb{S}^2$, and a characteristic class calculation shows that it is $\mathbb{RP}^3$. In conclusion we have

$$\mathrm{Rep}(K) = \mathbb{S}^2 \sqcup \mathbb{RP}^3.$$

This reminds us of something—where did we put our woodblock stamps? The unknot block has the same total homology as $\mathbb{S}^2$, while the knight's move has the total homology of $\mathbb{RP}^3$.

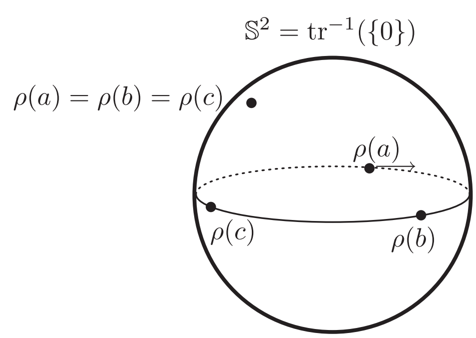



**Figure 15.** Computing the space $\mathrm{Rep}(K)$.

Looking back at Figure 2, we see that

$$\bigoplus_i H_i(\mathrm{Rep}(K)) = \bigoplus_{i,j} \mathrm{Kh}^{i,j}(K).$$

This is, and should be, both surprising and motivational. The definition that we have seen of Khovanov homology was entirely combinatorial and yet here is some relationship with tangible spaces parametrizing representations of the fundamental group of the knot complement.

The coincidence between the singular homology $\bigoplus_i H_i(\mathrm{Rep}(L))$ and $\bigoplus_{i,j} \mathrm{Kh}^{i,j}(L)$ does not hold for all links $L$, this coincidence rather being a shadow for small links of a more general algebraic relationship. Kronheimer and Mrowka proved the following result.

**Theorem 4** ([KM11]). *There is a spectral sequence from* $\mathrm{Kh}^{i,j}(L)$ *to* $I(\mathbb{S}^3, L)$.

If $U$ is the unknot and $K$ is a knot, then Kronheimer-Mrowka further showed that $I(\mathbb{S}^3, K) = I(\mathbb{S}^3, U)$ only if $K = U$. Their spectral sequence then allowed Kronheimer-Mrowka to co-opt this Floer homological result to conclude that $\mathrm{Kh}(K) = \mathrm{Kh}(U)$ only if $K = U$. In other words, Khovanov homology *detects the unknot*. The question of whether the Jones polynomial detects the unknot remains open.

**Observations.** We now have a partial explanation for the appearance of the unknot and knight's move blocks in the homology. The space $\mathrm{Rep}(K)$ always includes an $\mathbb{S}^2$ component corresponding to those $SU(2)$ representations of $\pi_1(\mathbb{S}^3 \setminus K)$ that factor through $H_1(\mathbb{S}^3 \setminus K) = \mathbb{Z}$. Roughly speaking, there should also be many $\mathbb{RP}^3$ components consisting of representations which are path-connected only to their own conjugates within $\mathrm{Rep}(K)$.

Furthermore, although we withhold the details, the internal gradings of various spectral sequences that start at Khovanov homology can give something of an explanation of the "diagonalness" that we saw in the homology tables.

It remains to account for the quantum height of the unknot block summand and its apparent relationship with the genus of torus knots.

## The 4-ball Genus

As was mentioned in the previous section, there are many spectral sequences that start at Khovanov homology or its quantum homological relatives. Some of these limit to Floer homological invariants, while others limit to quantum invariants. The original example of a spectral sequence that remained in the quantum world was due to Lee and was studied and used by Rasmussen.

**Theorem 5** ([Lee05, Ras10]). *For each knot $K$, there exists an even integer $s(K) \in 2\mathbb{Z}$ and a spectral sequence from* $\mathrm{Kh}^{i,j}(K)$ *to* $\mathrm{Kh}^{i,j-s(K)}(U)$ *where $U$ is the unknot.*

(Lee and Rasmussen originally worked over the rationals $\mathbb{Q}$, although the result is now known over $\mathbb{Z}$ as well.) In other words, the quantum $j$-grading of this copy of the unknot that we see, in all our examples, appearing in homological degree $i = 0$, is this even integral knot invariant known as the Rasmussen invariant $s(K)$.

Using the functoriality of Khovanov homology for smooth knot cobordism, Rasmussen showed the following result.

**Theorem 6.** *Suppose that $\Sigma : K_0 \to K_1$ is a connected smooth knot cobordism between the knots $K_0$ and $K_1$. Then we have that*

$$2g(\Sigma) \geq |s(K_0) - s(K_1)|.$$

In other words, the difference in values of the $s$ invariant between two knots provides a lower bound on the genus of a smooth connected knot cobordism between them. What has this got to do with torus knots?

For our example of the $(5,6)$-torus knot we see that $s(T_{5,6}) = 20$. Suppose that $T_{5,6} = \partial S$ where $S \subset B^4$ is a smooth, connected, oriented surface in the 4-ball. Such a surface is called a *slice surface* for $T_{5,6}$ and the minimal genus of such a surface is called the *slice genus* $g_*(T_{5,6})$.

Puncturing by removing a small ball centered at an interior point of $S$ gives rise to a knot cobordism $\mathring{S} : T(5,6) \to U$ where $U$ is the unknot. We then have

$$g(S) = g(\mathring{S}) \geq \frac{|s(T_{5,6}) - s(U)|}{2} = \frac{|20-0|}{2} = 10.$$

So we have $g_*(T_{5,6}) \geq 10$. Since there is a surface in the 3-sphere of genus 10 whose boundary is $T_{5,6}$, pushing the interior of this surface down into the 4-ball shows that

$$g_*(T_{5,6}) = 10 = g(T_{5,6}).$$

Similar arguments using $s(T_{p,q})$ work for all torus knots $T_{p.q}$ so that we can conclude the following result which is sometimes known as the Milnor Conjecture, and sometimes as the Local Thom Conjecture.

**Theorem 7.** *For any torus knot $T_{p,q}$ we have*

$$g_*(T_{p,q}) = \frac{(p-1)(q-1)}{2} = g(T_{p,q}).$$

This result was first proved thirty years ago [KM94], and was one of the high points of gauge theory as applied to the understanding of smooth 4-manifolds. Rasmussen's purely combinatorial proof, which we have just outlined, was something of a surprise to the low-dimensional topology community. It was hitherto believed that very little of substance in smooth 4-dimensional topology could be proved while sidestepping serious analysis.

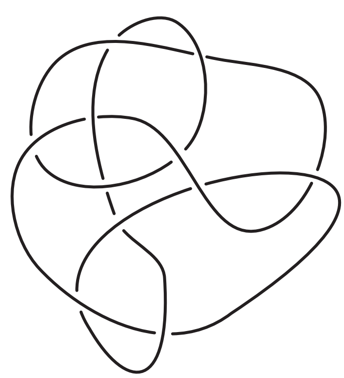


**Figure 16.** The Conway knot $C$.

Of course the structure of the argument which produces from $s(K)$ a lower bound on $g_*(K)$ applies equally well to knots $K$ other than torus knots. In one of the more celebrated results of recent years, Piccirillo used the $s$ invariant to show that the Conway knot $C$ (pictured in Figure 16) does not bound a smooth disk in the 4-ball [Pic20] (in other words

$g_*(C) \neq 0$). Although it was known that the Conway knot satisfies $g_*(C) \in \{0,1\}$, it had hitherto stubbornly resisted attempts to determine its slice genus precisely.

Piccirillo's work was more than just the computation of $s(C)$; in fact $s(C) = 0$ so does not obstruct $g_*(C) = 0$. Piccirillo gave a knot $\overline{C}$ that she showed satisfies $g_*(\overline{C}) = 0$ if and only if $g_*(C) = 0$. Then an electronic computation gave $s(\overline{C}) = 2$.

Although we have seen that quantum and Floer theoretic invariants are becoming increasingly tightly woven together, at the time of writing there are no Floer theoretic invariants that can be combined with Piccirillo's topological arguments to demonstrate that $g_*(C) \neq 0$.

## Further Reading

Khovanov homology is so wide-ranging that one would need a book, rather than an article, to attempt to chase down its influence on all the fields that it has colonized. In particular, we are sorry to forgo consideration of the role of Khovanov homology in quantum representation theory and in mathematical physics; we refer the interested reader to elegant and comprehensible discussions of these in [LS22] and [GS16] respectively.

On the other hand, we hope that we have realized our objective of giving readers who have persevered to this point a feeling for Khovanov homology. We entrust those who now wish to read a little more to the seductive work of Bar-Natan [BN05]. In this paper Bar-Natan takes as his object of study the first, topological, Khovanov cube that we drew above in Figure 8. He shows how this cube itself, in a suitable category, already yields a link invariant. In this way he opens the door to a much subtler understanding of Khovanov homology, and one that is more amenable to abstraction and generalization.

ACKNOWLEDGMENT. We thank Mikhail Khovanov, Aaron Lauda, Mark Powell, and Raphael Zentner for helpful comments on an early draft.

## References

[BD95] John C. Baez and James Dolan, *Higher-dimensional algebra and topological quantum field theory*, J. Math. Phys. **36** (1995), no. 11, 6073–6105, DOI 10.1063/1.531236. MR1355899

[BN05] Dror Bar-Natan, *Khovanov's homology for tangles and cobordisms*, Geom. Topol. **9** (2005), 1443–1499, DOI 10.2140/gt.2005.9.1443. MR2174270

[CF94] Louis Crane and Igor B. Frenkel, *Four-dimensional topological quantum field theory, Hopf categories, and the canonical bases*, J. Math. Phys. **35** (1994), no. 10, 5136–5154, DOI 10.1063/1.530746. Topology and physics. MR1295461

[CMW09] David Clark, Scott Morrison, and Kevin Walker, *Fixing the functoriality of Khovanov homology*, Geom. Topol. **13** (2009), no. 3, 1499–1582, DOI 10.2140/gt.2009.13.1499. MR2496052

[GS16] Sergei Gukov and Ingmar Saberi, *Lectures on knot homology and quantum curves*, Physics and mathematics of link homology, Contemp. Math., vol. 680, Amer. Math. Soc., Providence, RI, 2016, pp. 59–97, DOI 10.1090/conm/680. MR3591643

[Jac04] Magnus Jacobsson, *An invariant of link cobordisms from Khovanov homology*, Algebr. Geom. Topol. **4** (2004), 1211–1251, DOI 10.2140/agt.2004.4.1211. MR2113903

[Kau87] Louis H. Kauffman, *State models and the Jones polynomial*, Topology **26** (1987), no. 3, 395–407, DOI 10.1016/0040-9383(87)90009-7. MR899057

[Kho00] Mikhail Khovanov, *A categorification of the Jones polynomial*, Duke Math. J. **101** (2000), no. 3, 359–426, DOI 10.1215/S0012-7094-00-10131-7. MR1740682

[KM11] P. B. Kronheimer and T. S. Mrowka, *Khovanov homology is an unknot-detector*, Publ. Math. Inst. Hautes Études Sci. **113** (2011), 97–208, DOI 10.1007/s10240-010-0030-y. MR2805599

[KM94] P. B. Kronheimer and T. S. Mrowka, *The genus of embedded surfaces in the projective plane*, Math. Res. Lett. **1** (1994), no. 6, 797–808, DOI 10.4310/MRL.1994.v1.n6.a14. MR1306022

[Lee05] Eun Soo Lee, *An endomorphism of the Khovanov invariant*, Adv. Math. **197** (2005), no. 2, 554–586, DOI 10.1016/j.aim.2004.10.015. MR2173845

[LS22] Aaron D. Lauda and Joshua Sussan, *An invitation to categorification*, Notices Amer. Math. Soc. **69** (2022), no. 1, 11–21, DOI 10.1090/noti2399. MR4353348

[OS05] Peter Ozsváth and Zoltán Szabó, *On the Heegaard Floer homology of branched double-covers*, Adv. Math. **194** (2005), no. 1, 1–33, DOI 10.1016/j.aim.2004.05.008. MR2141852

[Pic20] Lisa Piccirillo, *The Conway knot is not slice*, Ann. of Math. (2) **191** (2020), no. 2, 581–591, DOI 10.4007/annals.2020.191.2.5. MR4076631

[Ras10] Jacob Rasmussen, *Khovanov homology and the slice genus*, Invent. Math. **182** (2010), no. 2, 419–447, DOI 10.1007/s00222-010-0275-6. MR2729272

Andrew Lobb

## Credits

All figures are courtesy of the author.

Photo of the author is courtesy of Jane Lobb.

# EARLY CAREER

The Early Career Section offers information and suggestions for graduate students, job seekers, early career academics of all types, and those who mentor them. Krystal Taylor and Ben Jaye serve as the editors of this section. Next month's theme will be Jobs and Grants.

# Research

## Lessons from Our Advisor Christopher K. R. T. Jones

### *Jonathan Rubin and Amitabha Bose*

Like many aspects of being a professor, advising PhD students is often part of the job but is rarely taught. Those of us in this profession encounter many instructors during our academic careers, and hence have deep wells of experience to draw from when developing our own approaches to teaching. Yet most of us only have one doctoral advisor, so an advising style tends to be a highly individualized trait that strongly reflects an advisor's personality and values. Both of us had the opportunity to work with Chris Jones as our PhD advisor, yet it was only from conversations after the fact that we came to fully realize how special and beneficial this experience was. Preparing to write this article reinforced this realization, as we reached out to Chris's other advisees and many shared their deep appreciation for their experiences with Chris and his role in helping them to attain their professional goals. Chris's advisees represent a diverse array of people: those who grew up in the US as well as students from around the world; women and men; people of different races and ethnicities. How was it that Chris, who grew up and completed his Bachelor's degree in England, was able to connect with all of them in ways that made them feel unique? What were the lessons he passed on to each of us, either through words or actions, that have helped to shape us as professionals? We've tried here to piece together a few thoughts about all this and underlying them all is the notion that when Chris took on a student, he never viewed it as a limited-time relationship, but rather one in which he was committed to the student for the long term, or at the very least until they were professionally and personally settled.

Prior to his entrance to university as an undergraduate student pursuing both math and philosophy, Chris had spent time traveling the world, enjoying the experience of meeting and interacting with people. These early experiences no doubt shaped his ability to connect with so many of us later on. Chris completed his PhD with Charles Conley in the late 1970s. While doing so, he clearly absorbed a strong geometric perspective on dynamical systems. For example, he learned of the Wazewski principle, a tool that became a staple of his graduate ODE classes and that both of us continue to teach in our own courses. This geometric perspective pervades Chris's work and that of many of his trainees.

We met Chris in the early 1990s at Brown University. He joined the faculty in the Division of Applied Math in 1990, bringing his PhD student Todd Kapitula along for good measure. Among the graduate students, Chris immediately developed a reputation as someone who cared about and was willing to spend time with students. From the feedback and clarity of his written comments on our graded assignments, to his holding extra recitation sections to go over homework problem solutions, to his organization of topics seminars for his students, postdocs, and visitors, his actions spoke for themselves. For one of us (AB) a particular training method stands out. Chris made me review a set of seminal papers by presenting them in excruciating detail to him on the blackboard in his office. These meetings were intense for me, but critical to my development in understanding the depth of knowledge and expertise that Chris expected of me. I distinctly recall him telling me "You have to develop a desire to know the (mathematical) truth!" To this day, I use exactly the same blackboard sessions with each of my PhD students. It takes time. The student initially is hesitant, but as time goes by, the student becomes more confident and clearly benefits.



*Jonathan Rubin is a professor of mathematics at the University of Pittsburgh. His email address is* jonrubin@pitt.edu.

*Amitabha Bose is a professor of mathematics at the New Jersey Institute of Technology. His email address is* bose@njit.edu.

DOI: https://doi.org/10.1090/noti2939

Another aspect of Chris's caring has been his respect for students' unique individual interests. The other of us (JR) came to Brown hoping to work on mathematical biology. Knowing this, when Chris went on sabbatical just as I was looking for a thesis topic, he set up the opportunity for me to travel weekly from Brown to visit the group of Nancy Kopell at Boston University. That experience provided me with an invaluable introduction to mathematical neuroscience and opportunity to interact with Kopell, her group and visitors, and many of the other dynamicists at BU, which has had a lasting impact on my career. When that experience did not lead me to a suitable thesis topic, Chris agreed to connect me with a problem from nonlinear optics, but he specifically steered me to consider dynamics inside a semiconductor laser cavity as modeled by a coupled system of reaction-diffusion equations, because experience with that model would likely be transferable to problems in mathematical biology.

Coupled with our own reflections, the comments from Chris's former students helped us to identify several major themes that underlie his advising philosophy, which we detail below.

## High Standards

Chris has a clear view of what he aims for in his own work: results that are mathematically interesting and rigorous or at least provide mathematical insight, that shed new light on an interesting phenomenon, and that are expressed clearly and elegantly, ideally in a way that includes a geometric perspective. He applies these same standards to his students' performance across all domains. Importantly, he does so in a way that we would describe as humanistic: he acknowledges when problems are challenging, he respects students for putting in effort, and he compliments them for success; when effort does not produce the desired outcome, he does not belittle his students, but he also does not lower his expectations. In this vein, an anecdote from Todd Kapitula is illustrative: "When I was a graduate student and learning how to use some of the analytical/topological tools associated with the Evans function, Chris had me look at a couple of problems to see what I could do with them. I ended up solving one of them. Surprisingly to me, at the time the result was unpublished. Chris complimented me on completing the work, but said nothing about writing it up and submitting it for publication. I asked him about this, and his comment was along the lines of, "not every result needs to be published." As time went on it became clearer to me that he was looking for good work on problems that he thought were interesting, and he wanted to publish only if the problem and results crossed a fairly high threshold. It was not to say that if the threshold was not reached the work was not good—it's just that he was looking for a little bit more."

## Individualized Attention to Students and Finding Suitable Paths for Each of Them

Each student comes with their own goals and interests. Chris embraces that gestalt and customizes his approach accordingly. Chris himself recently said to us, "I was invested in the student's career path first and foremost, before the research that followed." The experience of JR described above exemplifies this principle. Other students of Chris's shared their own versions of this experience. For example, Olivia Chandrasekhar "approached Chris as a second year student, emboldened by my recent success on my first comprehensive exam and looking for a way to meaningfully apply all the mathematical theory I'd devoted my time to studying. I'd decided I wanted to study models of wildfire propagation, although at the time I had little idea what that actually entailed. I reached out to Chris because of his interest in climate dynamics, which I decided was sufficiently close to my area of interest." Chris not only agreed to work with Olivia, despite not having his own experience with wildfire research, but went further, as Olivia describes, "After our first Zoom call, Chris sent me a paper, gave me a brief rundown of tipping phenomena in dynamical systems, and suggested that we apply for a fellowship to support my work. . . We worked tirelessly on my fellowship application, bouncing email drafts back and forth in real time, Chris patiently clarifying technical details I bungled or didn't fully understand. To my absolute surprise, our application was successful. After another potential advisor had discouraged me from applying for a similar fellowship because he felt my academic record wasn't impressive enough, Chris's investment and belief in me was extremely heartening. The funding from my fellowship has allowed me to pursue my own research agenda, but I believe Chris would have encouraged this independence regardless of my funding source. . . As a rule, Chris is supportive of each student's individual path and respectful of their career aspirations. I'm very grateful to have a mentor who understands and supports my goals, rather than trying to fit my graduate work and experience into their mold of success." In brief, as his recent student Katie Slyman told us, "I appreciate that Chris let me dictate my future career and path, and he found opportunities for me that aided in that process."

## Constantly Seeking Opportunities for Mentees

Olivia's experience with her fellowship application illustrates this principle, which many of Chris's other students also highlighted. It is almost uncanny how adept Chris has been at pairing his advisees with opportunities that are well-attuned to their goals and interests and have long-lasting impact on their career trajectories. For example, once JR was settled on a nonlinear optics thesis topic, Chris supported his participation in a multiweek summer school and workshop in Edinburgh on nonlinear optics. Similarly, when AB was deep in the throes of his thesis, Chris had a conversation with Nancy Kopell about the problem which broke my logjam and ultimately led to a postdoctoral appointment for me with Kopell. Similarly, Katie Slyman reported that, "I was able to partake in two academic summer schools, which shaped my network of collaborators, as well as gave me a flavor for different types of dynamical systems research and extensions to the field. Additionally, Chris provided the opportunity to comentor an undergraduate senior thesis project. This mentoring role gave me the chance to gain experience and learn what I could expect if I chose a job in academia." Todd Kapitula also described this experience, "I remember him getting me invited to various workshops, and him working on getting me speaking invitations at various conferences, for the express purpose of meeting the right people, and better understanding what others were doing. As the years go by, I appreciate more and more how he looked out for me at the beginning of my poststudent career…doing what he could to help me succeed."

## Team-Building with Vertical and Disciplinary Breadth

Chris is a strong believer in team science. As an applied mathematician, he recognizes that research is most effective when it starts from the problems that arise in an application area, and that the best way to ensure this connection is to engage directly with the experts in this area. There are two aspects to this approach that he emphasizes with students. First, it's important to fully embrace the terminology and "culture" prevalent in the field. Second, it's critical to balance applied relevance with mathematical intricacy; that is, research should identify and tackle the mathematical challenges that are needed for the applied challenges. In doing so, we create new math and new science. Chris fully includes his students in this multidisciplinary team engagement, and through this experience, his advisees unknowingly develop a sense of how to organize their own professional relationships. He has used this approach very effectively over the years, for example, building diverse networks of collaborators in nonlinear optics, geophysical fluid dynamics, and most recently, climate change. Examples of this principle come from the comments of former students. Colin Grudzien told us, "Chris Jones showed that theoretical mathematics can have a direct impact in other sciences and real-world applications. The key that Chris showed was in engaging scientists from other disciplines with openness and respect, and by assembling diverse teams that valued the differences in members' expertise and perspectives." Monica Romeo wrote to us, "I appreciated Chris' community-oriented style of mentoring and practicing mathematics. Knowing a wide range of applied mathematicians helped me become more comfortable talking about mathematics with different people. It also connected me with a great group of individuals." Finally, Olivia Chandrasekhar writes, "Another hallmark of Chris's advising style is his willingness to collaborate, both with other mathematicians and researchers outside his field. Throughout my time as his student, he has encouraged my ongoing collaboration with researchers in wildland fire modeling at a national lab. This work furthers both my current research interests and my future career goals."

## "Mathematics is a Social Activity"

In May 2023, at the recent "SIAM Conference on the Applications of Dynamical Systems" in Portland, Oregon, the two of us had a chance to sit down in a lovely park with Chris to hear directly from him about his advising style (Figure 1). We didn't share with him any of the comments of his former students, but simply asked him if he had a specific advising philosophy. In response, one thing he mentioned that resonated with the two of us was the phrase "Mathematics is a social activity." Chris's elaboration on this simple statement was very consistent with what his students had independently communicated to us, as documented above. Undoubtedly his advising style has centered around this notion while being informed by and evolving due to his interactions with students and colleagues. The two of us, and likely all of his advisees, understand this statement in light of our witnessing the depth of Chris's involvement with the community of applied mathematicians as a whole, as well as the commitment, care and attention that he has individualized for each of his students.

**Figure 1.** Jonathan Rubin, Christopher K. R. T. Jones, and Amitabha Bose.

### Credits

Figure 1 is courtesy of Amitabha Bose.

# Research with Graduate Students

*Robert Fraser*

## How and Why I Decided to Involve Students in My Research

When I was a PhD student at the University of British Columbia, a more senior student in our department, Kyle Hambrook, heard that I was working on a problem in $p$-adic harmonic analysis. He approached me with a problem he wanted to consider—adapting a construction from the Euclidean setting to the $p$-adic setting. He mentioned in passing that he thought that PhD students in our department did not do enough collaboration and he was hoping that we would be able to do this project together. Now, half a decade later, we are still collaborating and working on problems together, and some of my best work has been joint with Kyle. I don't know if I could ever pay Kyle back for the trust he placed in me as an inexperienced PhD student, but as the saying goes, if you can't pay it back, pay it forward. Thus my piece of advice to young researchers: **always look for opportunities to involve graduate students in your research.**

One of the most striking things I noticed as a postdoc was how *lonely* it can get. I had an excellent postdoctoral supervisor and a very welcoming research group and department, but even so I found it difficult to see myself as part of the department. I was a newcomer who had no understanding of the culture of the department I had suddenly become a part of, and knew that in a few short years I would have to leave again. As a postdoc, I found it *difficult* to get involved with the research group outside of interacting with my supervisor even though the department as a whole could not have been more welcoming. My postdoctoral years were exciting from a mathematical perspective, but knowing that I would be leaving soon made it difficult to integrate into the department.

One observation I made as a postdoc was that the graduate students at the University of Edinburgh learned very different background material from the graduate students at the University of British Columbia. In my second year at Edinburgh, I taught a course on additive combinatorics—an area that essentially all of the harmonic analysis students at UBC had encountered, but that was more-or-less unknown to the graduate students at Edinburgh. As a postdoc, you should try to get a feel for what the graduate students in your department know. Often it will be very different from what you learned as a PhD student.

*Robert Fraser is an assistant professor of mathematics at the Wichita State University. His email address is* `robert.fraser@wichita.edu`.

DOI: https://doi.org/10.1090/noti2935

While at Edinburgh, I had an idea for a new research project that I could tackle. I decided that, in order to become more involved in the department, I would send out an email asking if any PhD students were interested in working with me. I was happy when PhD student Reuben Wheeler expressed interest. This is something that might surprise newer researchers—many graduate students will jump at the chance to be a part of a new project.

Unfortunately, the project that I had proposed did not work out—the idea I had come up with to tackle the problem had a flaw, and at the time I felt that I had wasted Reuben's time. However, I later came to realize that **even an unsuccessful side project can be a good experience for a PhD student**. Many PhD students have their first few projects hand-picked by their supervisors, and learning when and how to fail at a project is a valuable skill.

Later, I proposed a second project to Reuben Wheeler, as well as two number theorists at the University of York—Sanju Velani and Evgeniy Zorin. Sanju and Evgeniy encouraged Reuben and me to write up the project ourselves and it ended up being Reuben's second publication. We are currently writing a follow-up project and this has become a very fruitful direction of research.

## Why You Should Involve Graduate Students in Your Work

Giving PhD students a chance to work on a project with you can be a valuable opportunity for them to gain new research perspectives. Further, involving graduate students in your work can be a huge benefit to **you** as well.

If you are intending to pursue a research career in mathematics, you are at some point going to be supervising graduate students of your own. Bringing a graduate student onto a research project is a good way to build your skills as a supervisor. Everyone has their own quirks in terms of how they work on mathematics projects—for example, some people do their best work in the morning, and others do their best work in the evening. Some people think most clearly in the office, and some people think most clearly when out on an errand. Everyone is different, and when you start to supervise graduate students of your own, you are going to need to know how to supervise students who will not work in the way that is the most comfortable for you. It is important to practice this skill early in a relatively low-stakes setting so that you are ready to supervise students when you are hired as a professor.

Starting a joint project with a graduate student is a great way to expand your perspective on mathematics. If you work on a project with an experienced mathematician, they are likely familiar with a broad range of literature and will likely have read many of the same papers as you and many other subject matter experts. However, a graduate student will likely have read only a few papers, which may shape their perspective on your research area in a unique way. It can be advantageous to work with someone who has only read a few papers because it is likely that their perspective will be vastly different from someone who has been working in the field for a long time. Often I have found that working with a graduate student on a project gives me a broader perspective on problems that I thought I knew a lot about.

During your postdoctoral years (and perhaps even as early as the end of your PhD), you should be thinking about potential research projects for master's and PhD students. I was actually surprised at how many problems generated by my research would be appropriate for someone early in their career. For example, it is often possible to sharpen an existing result by obtaining better estimates on an error term. Often, a theorem that works in one setting can be straightforwardly adapted to another setting. These kinds of simple but nontrivial arguments are perfect problems for future graduate students. Taking the time to work on problems with graduate students will help train you to think of good problems for them.

## How to Involve Graduate Students in Your Work

No matter what stage you are in your career, you have an opportunity to involve graduate students in your research. If you are currently a graduate student, you can always work on a joint project with another graduate student in your department. Other students—especially more junior students—will usually jump at the opportunity to work with you.

Later in your career, if you want to work with a graduate student in your department, you can simply send an email and see if anyone is interested.

Another interesting idea is to talk to graduate students at *other* institutions. This is something you should be doing at every conference. When you attend a conference, it is tempting to try to talk only to people you already know, or experts who you are interested in working with. However, it is often rewarding to talk to graduate students about their work as well. Often, a graduate student at another institution will be working on an interesting problem you've never heard of. This is an excellent opportunity for the student as well—most graduate students do not have the opportunity to coauthor a paper with someone at another institution.

A great way to meet graduate students at conferences is through problem sessions. If there is a problem session at a conference, try to present a problem that will be within the reach of a graduate student. You might be surprised at how many graduate students are willing to work on problems with you. If you attend a conference that does not have a problem session on the schedule, you can always suggest holding one. Even an informal problem session will probably be very popular, especially with younger researchers.

## Lightning Round—Some Miscellaneous Advice

It can be hard, but try to keep a regular work schedule. Think about what conditions lead to getting the most work done and try to do most of your work in a favorable environment. Never neglect your teaching duties—teaching well takes a lot of planning. Do not use class time to discuss homework problems. Try to maintain a positive relationship with your colleagues, if possible. Do not neglect your health. Make sure you take the time to exercise. Do not dwell on your mistakes or failures. Do not be intimidated by other researchers. Do not underestimate the impact of your research on the community. Do not assume you know everything about your research area. Do not assume you will have nothing to bring to a new research area. Do not be afraid to supervise a graduate student working on a problem that is slightly outside of your area of expertise. Do not be afraid to ask the senior professors in your department for advice. Always be grateful for the help and kindness you received from others, and try to provide the same help to people who might need it. Respect the wisdom of more experienced mathematicians, but remain humble when working with students or junior mathematicians. Before you start your first permanent position, make sure you know when your grant proposal deadlines are. Work on problems that will be of interest to other researchers in your area, but be wary of chasing fads. If you are going to work on a popular problem, try coming at it from a different angle from the rest of the experts. Your job will be stressful, but never forget the joy of teaching or doing research.

Robert Fraser

### Credits

Photo of Robert Fraser is courtesy of Gloria Gartner.

# How to Start a Career in Mathematical Biology

*Avner Friedman*

Mathematical biology is a multidisciplinary field of research where questions that arise in biology are addressed with mathematics. In this sense, it is very much like mathematical physics, which has a long history of advancing physics, while at the same time has given rise to new areas of research in mathematics. However, biology is far more complex than physics. Take, for example, a single cell with its millions of proteins and DNA molecules: It has to navigate its life by defending itself while supporting the organ to which it belongs, continuously sensing its ever-changing environment and keeping in contact with other cells. And there are billions of billions of such cells even in a mouse, or smaller animals.

There has been gigantic progress in the biological and biomedical sciences in recent decades, generating an immense amount of data. The need to extract knowledge from these data has been a challenge and an opportunity for mathematics to step in. This, in fact, was the reason why I founded, in 2002, the Mathematical Biosciences Institute (MBI) at The Ohio State University. The MBI, which was one of the eight institutes funded by the National Science Foundation at the time, brought biologists and mathematicians together in many workshops, and hosted short- and long-term visitors, and long-term postdocs, to discuss and collaborate on research projects in mathematical biology. Based on my experience in working in this field over the last 20 years, I would like to advise graduate students on how to start a career in mathematical biology, focusing primarily on how to proceed with the first project, toward a PhD thesis. I assume that you have already been exposed to mathematical biology at the undergraduate level, in terms of simple models and numerical simulations. Such material can be found in the textbooks [1], [2] that we are using at OSU.

The field of mathematical biology is immensely broad. It is important that you choose a mentor who is actively engaged in research in mathematical biology and, preferably, not too narrow. Expect your mentor to suggest reading material, books and research articles (they are mostly online). Your mentor is expected to meet with you frequently, to respond to your questions, and to hold informal discussions that will arise from your readings. Such discussions will naturally lead to biological questions that have the potential to be addressed with mathematics. Addressing one of these questions, I expect, could become a project for your PhD thesis. So here is some advice how to develop this project, illustrated by an example of a model represented by a dynamical system.

1. *Build a chemical reaction network based on the biological mechanism*

You need to build a chemical reaction network as a basis for addressing your biological question. This necessitates a solid background on the biological details related to your question, and only then decide which variables (e.g., cells, proteins) have to be included in order to address the question, and which may be dropped; it is important not to make the model unnecessarily complicated, but also not to oversimplify the biology. Draw a reaction network with the selected variables as nodes, and with edges indicating their functions (activation, inhibition, catalyzation, etc.). For example, let $X$ and $Y$ be two types of proteins, a sharp arrow from node $X$ to node $Y$ indicates that $X$ activates $Y$, while an arrow with blocked edge indicates that X deactivates (or blocks) $Y$.

2. *Develop a mathematical model*

To develop a mathematical model on the basis of a biological network, you need to represent each edge by a mathematical expression. For example, a sharp arrow from cells $Y$ to proteins $X$ means that $Y$ produces $X$, and we write

$$\frac{dX}{dt} = aY - bX,$$

where a is the rate of production and $b$ is the rate of degradation of X (every protein eventually degrades). If a drug $Z$ blocks the degradation of $X$, we replace $b$ by $\frac{b}{1+cZ}$, where $c$ is the "strength"of the drug.

Decisions have to be made, based on the particular situation, as to what level of detail to incorporate into the model. An example of a more complex case is the enhancement of reaction by ligands. Here, let $L$, $R$, and $P$ denote a ligand, receptor, and product respectively. When the ligands $L$ outside cells $X$ attach to proteins R on the surface (membrane) of cells $X$ and, as a result, $R$ changes its structure and starts a chain reaction (signaling cascade) of proteins that eventually lead to the production, by $X$, of some new proteins $P$. How do we model this process? The answer depends on what is relevant to the question we are addressing. If all we need to know is that $L$ caused the production of $P$ by $X$, then we represent this process by writing,

$$\frac{dP}{dt} = a\frac{XL}{K+L},$$

where a is the rate by which $X$ can "eat" the proteins $L$ (with $R$), and the 'eating' expression is the famous Michaelis-Menten functional response. If the biological question involves space (e.g., tissue growth), then we need to use a

*Avner Friedman is a Distinguished University Professor, Faculty Emeritus, at the Ohio State University. His email address is* `friedman.158@osu.edu`.

DOI: https://doi.org/10.1090/noti2936

PDE model which includes diffusion of some of the variables.

3. *Parameter estimates*

The mathematical model has many parameters, and they have to be correctly estimated before the model can be reliably used. Some parameters can be easily estimated. For example, consider the case $\frac{dX}{dt} = a - bX$, where X is a protein with constant source $a$. The half-life $s$ of $X$ is defined by the relations

$$\frac{dX}{dt} = -bX,\, X(s) = \frac{X(0)}{2},$$

so that $b = \frac{ln2}{s}$. Since the quantity $s$ is recorded for many known substances, the parameter $b$ can be reliably determined. As for the parameter $a$, we try to find the average density $X_0$ of $X$, and then infer $a = bX_0$, assuming the biological system is at average state.

But there are usually many other parameters that cannot be found, nor inferred from available literature, and the only way to estimate them is by making some assumptions. It may be good to discuss with your mentor or biologists which assumptions are biologically acceptable and which are not. Such a case arises, for instance, in an equation of the form

$$\frac{dX}{dt} = aY - cZ - bX,$$

where both $a$ and $c$ are unknown rates. If $Y$ is assumed to be more efficient than $Z$ in activating $X$, then you make an assumption that $a$ is one order of magnitude greater than $c$, for instance, and you can reduce the situation to the case of one unknown parameter, as in the case considered above.

4. *Model validation*

You first simulate the model extensively to show that the predictions agree with known experimental studies. This model validation process is essential, since there is always some uncertainty in the parameters, and perhaps even in some of the model equations. For experimental data you look either into animal studies (e.g., mouse models) or in vitro studies. Agreement with experimental data has to be at least qualitatively, but also 'not far off' quantitatively in terms of the range of known measurements.

Unless you are extremely fortunate, you are more than likely to find some disagreements with known literature in your first try. This often means that more than one of your uncertain parameters are wrong and need to be changed. To this end, one must make an educated guess as to which parameter is the most critical and which should be the first to change. This is a judgement call that improves with intuition and experience with the mathematical models; your mentor could be very helpful.

After you have made the changes in some parameters, you repeat the simulations. This process can take many steps, but eventually converges to a model that has shown to replicate known experimental results. Now you are ready to use the model to address the biological question.

5. *Addressing your biological question: A case study*

Cancer is treated with a combination of two drugs, chemotherapy $D$ and immune therapy $G$. Both drugs are injected intravenously once a month, $D$ at day 1 of the month and $G$ by day 14.

Question: Is this the best schedule, or are there other schedules that yield more benefits to the patient?

You begin by choosing the major players of the model. Clearly, cancer cells and the drugs $D$ and $G$ need to be included, but, since $G$ activates the immune response to cancer, also the most relevant immune cells; we probably need also to include proteins (cytokines) that these cells secrete in their interactions among each other and with cancer cells.

In order to address the biomedical question, you need to quantify the benefits of treatment, i.e., the efficacy of the drugs, in mathematical terms. If your model consists of a system of PDEs, then you may take the efficacy E(D,G) to be the percentage of tumor volume reduction, by a specified time horizon $T$. We can now simulate the model and compute E(D,G) for different schedules, and thus generating quantitative answer to address the original question.

But you can do much more with model. You can try changing/optimizing the amount of drugs to reduce side effects, or changing the ratio $D/G$. You can also modify the model to fit the profile of a specific type of cancer. You may end your project with a paper that makes suggestions for future clinical trials. In [3] you will find more details and examples on how to build mathematical models.

As you have completed a few projects successfully, you have now finished your thesis and are going to receive a PhD degree. The time has come to apply for a job in a university or in some research institute. Take a look at the book "A PhD is not enough," [4]; it has good advice for early careers in the physical sciences, and some of it applies also to careers in mathematical biology. In your application for a job, you will include a CV and a short essay describing your research accomplishments and plans for future research. Your application will show that you have done innovative work in an area of interest and promise. Ask your mentor to review your application!

If your project has attracted the attention of fellow researchers, you may be invited for one or multiple job interviews, and you are going to give a one-hour job talk. It is critically important that you prepare your presentation very carefully. Giving a talk is not the same as writing a paper. Prepare a talk that will make people want to hire you. Start by describing a broad area of research, of broad interest and critical importance, with challenges and opportunities for progress. And then present your own research, as naturally embedded in this general context. Your future research plans should be ambitious, but not unrealistic. Tell a good story. **Rehearse your talk in front of an audience who are not shy to give critical feedback, prepare a response to questions.**

In your job interview, you are likely to be asked to describe your research in a few sentences (e.g., by the dean or chair of the department); prepare how to respond. It is sometimes more difficult to explain your work in five minutes than to explain it in one hour.

Although, in this article, I gave examples of modeling with ODE or PDE, there are many other areas of mathematics which offer PhD projects and a career in mathematical biology, including:

(i) Algebraic methods of symbolic computations, used in genomics, proteomic, and analysis of molecular structure of genes [5];
(ii) Knot theory, applied to folding of proteins and nucleic acids [6];
(iii) Topological data analysis, used to generate shapes of biomolecules and cells [7];
(iv) Computational biology, with data analysis, mathematical modeling and computational simulations, which is used to study protein-protein interaction and metabolic processes [8], could be a project more suitable in a multidisciplinary PhD program, as could be a project in artificial intelligence in biology, applied to agriculture, medicine, and bioindustry [9].

### References


[1] Avner Friedman and Chiu-Yen Kao, *Mathematical modeling of biological processes*, Lecture Notes on Mathematical Modelling in the Life Sciences, Springer, Cham, 2014, DOI 10.1007/978-3-319-08314-8. MR3290270

[2] Ching-Shan Chou and Avner Friedman, *Introduction to mathematical biology: Modeling, analysis, and simulations*, Springer Undergraduate Texts in Mathematics and Technology, Springer, [Cham], 2016, DOI 10.1007/978-3-319-29638-8. MR3468665

[3] Avner Friedman, *Mathematical biology: Modeling and analysis*, CBMS Regional Conference Series in Mathematics, vol. 127, Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2018. MR3793187

[4] P. I. Feibelman, *A PhD is Not Enough!: A Guide to Survival in Science*, Addison Wesley, Reading, MA, 1993.

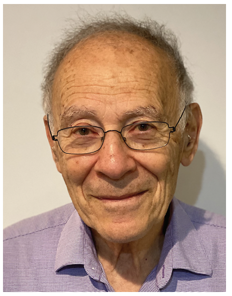
[5] Wikipedia, Mathematical and theoretical biology.

[6] Michael Monastyrsky, *Riemann, topology, and physics*, Birkhäuser Boston, Inc., Boston, MA, 1987. Translated from the Russian by James King and Victoria King; Edited and with a preface by R. O. Wells, Jr.; With a foreword by Freeman J. Dyson, DOI 10.1007/978-1-4899-3514-4. MR896908

[7] Y. Skaf & R. Laubenmacher, *Topological data analysis in biomedicine: A review*, J. Biomedical Informatics **130** (2022).

[8] Wikipedia, Computational Biology.

[9] A. Bhardwaj, S. Kishore & D. Pandey, Artificial Intelligence in Biological Sciences, MDPI, 2022.

Avner Friedman

### Credits

Photo of Avner Friedman is courtesy of Alissa Friedman.

# A Guide for Mathematics PhDs: How to Incorporate Education into Your Scholarship

*Melinda Lanius*

Let me introduce myself: I am an assistant professor at Auburn University, where I was specifically hired to conduct discipline-based education research (DBER) in a mathematics and statistics department. While my current research is solely in mathematics education, my undergraduate and graduate work, as well as my postdoctoral research, was in Mathematics. You may be wondering, *why the drastic change?* In the second year of my postdoc, I realized that I was fascinated by *math as a verb*, rather than *math as a noun*. To clarify, while I found mathematical objects to be interesting (math as a noun), I was deeply enthralled with the different ways people do, learn, and teach mathematics (math as a verb). Fortunately, I didn't need to go back to grad school to earn a second PhD in mathematics education; I have been able to pick up the tools and techniques of education research as I go, while also continuing to leverage the skills I developed as a mathematician.

In this article, I will share lessons I learned making the transition from research in mathematics to research in undergraduate mathematics education (RUME). I assume most readers won't want to entirely leave their mathematics research program to become education researchers and, accordingly, will provide advice on how you could build

*Melinda Lanius is an assistant professor in the department of mathematics and statistics at Auburn University. Her email address is* `melinda.lanius@auburn.edu`.

DOI: https://doi.org/10.1090/noti2937

familiarity with education research or how you might incorporate mathematics education research as a portion of your overall research program.

## Blurring "Teaching vs Research"

In 1990, Ernest L. Boyer, a former US commissioner of education (i.e., secretary of education) argued that academia should reconsider the siloing of teaching apart from research activities [Boy90]. A variety of initiatives, such as the scholarship of teaching and learning (SoTL) movement, have since emerged as a way for faculty to combine scientific inquiry with their own teaching practice. This is precisely where I recommend you start your journey, by considering teaching and learning in your own courses.

One way we can blur the line between teaching and research is to purposefully bring our mathematics research skills into the context of our teaching-related activities. In particular, in mathematics research we strive to be systematic by verifying and proving claims. We also share the results of our work broadly with our community, for example through seminar talks or posting our preprints to the arXiv.

Let's consider how we might be more systematic and public with a teaching-related material, which could be your lecture notes, a digital demonstration (e.g., GeoGebra), a handout for group work, or an assessment. While one may be quite careful and thoughtful in designing something to use in a class, if it is used over and over again without explicit consideration of an outcome goal and evidence that it will achieve that goal, then I will position that activity as informal and private. I've also been very vague here in terms of outcome goals. You could, for example, intend an assignment to support students in developing understanding of a concept or in developing a positive outlook on mathematics.

One way to be more systematic in the creation of your classroom materials is to collect feedback from a colleague. Alternatively, reading literature could help you think about your course materials to see if anyone has written about your scenario and desired outcome. You may find numerous suggestions that you can leverage when developing or refining your teaching. To be even more systematic, you can collect information that will suggest if you achieved your goal or not. For example, if you want a handout to support students in developing their understand of a concept, you might ask them to answer a short question at the beginning and at the end of class to see if there was a change after engaging with your handout.

The other important consideration for scholarship is dissemination. You can share your teaching materials online, either on your personal webpage or in a widely used repository such as *AMS Open Math Notes*. If you have collected data concerning the effectiveness of your materials in achieving your outcome goal, you can analyze your data and publish it in a wide variety of outlets, depending on both the scope of your data collection and the rigor of your methods. To delineate between SoTL and RUME, SoTL research tends to be more limited in scope, for example by focusing on one person's classroom. What we call RUME or DBER generally aims to make broader generalizations, for example by studying a learning phenomenon with college students coming from many types of institution—liberal arts colleges, universities, and community colleges. Frequently, the data and methods presented in RUME or DBER publications will be reviewed more critically than if they were to be presented as SoTL.

There are numerous grant opportunities that support education research. Seeing what education-related grants have been funded can provide you with a broad perspective on the current state of the field and help you in developing your support network. The most relevant NSF solicitation is called *Improving Undergraduate STEM Education (IUSE)*; I recommend typing "IUSE" and "mathematics," or a more specific search term if you are interested in say linear algebra, into the NSF award search: `https://www.nsf.gov/awardsearch/`. Note that you do not need a grant in order to engage in education research.

Above, I focused on creating and disseminating teaching materials, but you might be more interested in teaching-related activities that fall outside of the classroom, such as professional development for instructors. The private and informal version of this activity might be reflecting on your teaching by yourself. This could become more systematic by writing a teaching statement, reflecting on your student course evaluations and adjusting your teaching practice accordingly, or participating in a workshop on teaching and learning. To leverage your interest in this activity for scholarship, you can explore organizing teaching professional development for your community and reporting on the impact. Teaching professional development can range from resource-heavy options, like organizing a seminar series or a weeklong workshop with external experts, to low-cost options, like forming a community of colleagues that meets regularly to discuss teaching or visits each others' classrooms to provide teaching feedback.

After reading about these two possible directions, if you would like to start incorporating education into your scholarship, you will need to build your support network. In the next section, I will share the communities that supported me in making my transition from mathematics to mathematics education research.

## Building Your Support Network

**Locally.** If you are at a large university, your institution will likely have seminars or workshops related to teaching and learning that you can attend. If you are at a smaller institution, you may not be able to attend this type of event on your campus, but you can find virtual opportunities on the AMS or Mathematical Association of America (MAA) webpages. These resources likely focus on practical tips for practitioners which has been distilled from research. I highly recommend these; they will help you to develop as an educator and you will meet others who are focusing particularly on their teaching. On the other hand, this type of event will not reveal *how* to do education research. To gain familiarity with the mechanics of research, you should look for talks that are specifically intended to communicate research. You may have to work harder to find these events, for example by reaching out to another department on campus to see if you can join their colloquium email listserv. Even better, if your department has a seminar/colloquium budget, you could ask if you can invite an undergraduate mathematics education researcher to visit and give a talk.

As I mentioned before, SoTL as a movement is gaining a lot of traction. You may be able to locally find a SoTL institute/academy, an organized group with a facilitator to walk you through the steps of conducting your first research project. This could provide you with a cohort of colleagues who also are interested in engaging in education research and who come from different disciplinary backgrounds. A SoTL academy would likely walk you through the process of securing Institutional Review Board (IRB) approval for human subjects research, which may be required if you publish data related to a class. IRBs were put in place as an ethical safeguard after numerous instances of research abuse came to light in the 1900s (e.g., the Tuskegee Syphilis Study); you can read more about the history of IRBs in the the *Belmont Report*.

By federal law (Title 45 Code of Federal Regulations Part 46), every academic institution that receives federal funding is in charge of forming their own IRB, so it is important that you find a local "IRB mentor" who can share their experience of the process with you. While it may be tempting to ask a friend who you know has an approved IRB protocol at another institution about the process, don't do it! What may be true for me at my home institution is likely to not be true for you at your home institution. The forms you need to fill out and the training you must complete will be different.

**Nationally.** If you are attending a national meeting, such as the Joint Math Meetings or MAA Mathfest, try attending some sessions on education research. As I discussed above with respect to seminars and workshops on your local campus, some of these sessions will be more about relaying information and less about *how* to do research, so you may need to ask around about the purpose of the session.

You can also attend a conference specifically dedicated to education research. My two favorites are the Annual Conference on Research in Undergraduate Mathematics Education put on by the Special Interest Group of the MAA on RUME (`http://sigmaa.maa.org/rume/Site/Conferences.html`) and the annual conference of the Research Council on Mathematics Learning (RCML) (`https://www.rcml-math.org/`). RUME conference attendees have a variety of academic backgrounds; there will be mathematics PhDs and Mathematics Education PhDs presenting on their education research. I enjoy the RCML conference because it focuses on mathematics learning for both children and adults, broadening my perspective on what is happening at all levels of mathematics education research, not just in my particular area of focus. Similar to the annual RUME conference, researchers with a wide variety of backgrounds attend the RCML conference. You'll meet mathematics education faculty, mathematics faculty, and K–12 school leaders.

If, after attending some research talks, you feel like education research is something you would like to try, I would strongly encourage you to apply to attend a weeklong NSF-funded field school called Professional Development for Emerging Education Researchers (PEER) (`https://peerinstitute.org/`). These field schools typically happen in December or over the summer. Some locations may be more competitive to get into based on local demand, so I encourage you to apply again if you don't get invited the first time (One of my graduate students had to apply twice in order to participate). I attended a PEER field school in Rochester, NY, and it felt like summer camp for DBER! The program is extremely well-thought-out and hosted by a team of highly dedicated facilitators. I formed a pod of mathematics education researchers at PEER that I still keep in touch with.

## Audacity and Its Challenges

At this point, I've given you a mixture of terms all to describe education research: SoTL, DBER, and RUME. For me, these areas are not well-defined in the mathematical sense. I've experienced that what one person calls DBER, another will call RUME. For me, these terms have been most helpful in identifying myself and how I relate to a particular education research community, as well as helping me to set reasonable expectations for my work.

For those just getting started in education scholarship, SoTL can be a useful identity. SoTL is intended for faculty from noneducation backgrounds to become more reflective and scholarly teachers and to share their findings publicly. Because the focus of SoTL is on your classroom and your teaching practices, SoTL is a friendly introduction to education scholarship.

On the other hand, DBER/RUME have more expansive aspirations, meaning the skills will be much more challenging to learn. I currently identify most with DBER, because this identity acknowledges that I am coming at education from a disciplinary background. In math grad school, we slowly learn the norms of our community, from the conventions of mathematical writing to the structure of a research talk. We have a structured program of study, that makes sure we develop a common foundation of skills that will prepare us to be independent and successful mathematicians. As a discipline-based education researcher, I am learning new skills in a very ad hoc way, meaning I make mistakes that would be obvious to a second- or third-year mathematics education PhD student.

This is why I have called this section "Audacity and its Challenges." Audacity can mean both willingness to take a big risk as well as impudence. It is a testament to the welcoming and nurturing culture of the research in undergraduate mathematics education community that my audacity and I have been able to thrive.

## Why Try Education Scholarship?

The vast majority of mathematicians employed at a college or university are also responsible for teaching undergraduate or graduate mathematics courses. Additionally, many faculty members in mathematics departments will have to do teaching-focused service, making decisions about a broad range of issues from textbooks and course placement exams to teaching professional development and awards. Engaging with education research can open new opportunities and increase the effectiveness of your efforts. If you feel drawn to put a lot of effort into teaching and learning, writing up your work to share it with others could help you to meet your expected annual scholarly contributions (make sure that your department will count this before deciding if or how much you want to work in this area!). Most importantly, by engaging in this conversation, you can contribute your unique insights to our generalized understanding of mathematics teaching and learning.

As a mathematician, I've found that learning how to do education research is definitely challenging, but with perseverance and patience, it is doable. If you choose to integrate education into your scholarship, there are supportive communities waiting to welcome you!

### References

[Boy90] Ernest L. Boyer, *Scholarship reconsidered: Priorities of the professoriate*, Princeton University Press, 3175 Princeton Pike, Lawrenceville, NJ 08648, 1990.

Melinda Lanius

### Credits

Photo of Melinda Lanius is courtesy of College of Sciences and Mathematics, Auburn University.

## Dear Early Career

I'm on the job market but I'm not sure what kind of job I want. Where can I find job ads with a variety of types of jobs?

—*Ambivalent*

Dear Ambivalent,

This is a broad question, but we'll do our best! In our careers we have typically used MathJobs as our primary resource to navigate the academic job market, and while it performs a great service to the community, it is by no means the only resource. We'll mention some relevant sites that have proven useful in getting some people we know into jobs that they love, and which might help you start your research on jobs.

First of all, the aforementioned MathJobs (`https://www.mathjobs.org/jobs`) is sponsored by the AMS and aims to support job applicants with advanced degrees in mathematics, and employers seeking mathematicians. The success of MathJobs has led to the development of AcademicJobs (`https://academicjobsonline.org`), which seeks to provide a similar service for other areas—it is currently run by Duke's Math Department. AcademicJobs posts listings in a much broader range of disciplines, and you may find positions related to (for instance) data science or machine learning which are not listed on MathJobs. Most employers on both MathJobs and AcademicJobs are universities and major research labs.

While it is fair to call Mathjobs the primary resource for research-focused academic positions in mathematics in the US, other job sites list many more teaching-focused positions, in addition to some research-oriented positions which are not on MathJobs. Certainly, it is worth checking out the MAA sponsored site (`https://www.mathclassifieds.org/jobs/`), The Chronicle for Higher Education jobs site (`https://jobs.chronicle.com`), and the HigherEd jobs site (`https://www.higheredjobs.com/`). These sites also host many nonacademic jobs from employers looking for people with advanced math degrees.

In terms of understanding what jobs in business, industry and government are available to you, two great sites are the Erdos Institute (`https://www.erdosinstitute.org`) and the BIG Math Network (`https://bigmathnetwork.org`) (and, of course, LinkedIn). You can find out a bit more about the Erdos Institute in the July 2023 EC article by its cofounder Roman Holowinsky!

Additional important resources come from job boards hosted by other professional organizations. We'd like to mention here the Institute of Mathematical Statistics Career Center (`https://jobs.imstat.org`), and the SIAM (Society for Industrial and Applied Mathematicians) job board (`https://www.siam.org/careers/job-board`). Also well worth checking out are the job listings at the National Association of Mathematicians (`https://www.nam-math.org/job-announcements`) and the Math Alliance (`https://www.mathalliance.org/job-postings.html`).

Good luck with your job search!

*—Early Career editors*

Have a question that you think would fit into our Dear Early Career column? Submit it to `Taylor.2952@osu.edu` or `bjaye3@gatech.edu` with the subject Early Career.

DOI: https://doi.org/10.1090/noti2938

# Memories of Mikio Sato (1928–2023)

*Pierre Schapira, Takahiro Kawai, Yoshitsugu Takei, Masaki Kashiwara, Barry M. McCoy, Craig Tracy, Tetsuji Miwa, and Michio Jimbo*

Mikio Sato (born April 18, 1928, dec. January 9, 2023) was a mathematician of great depth and originality. Initiator of the theory of hyperfunctions and *D*-modules, he made influential contributions in diverse areas of mathematics ranging from analysis, number theory to mathematical physics.

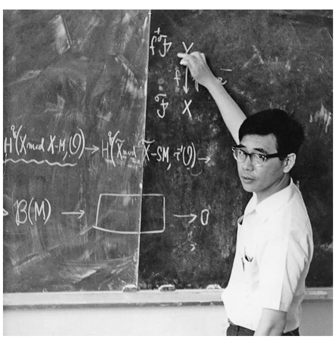


**Figure 1.** Mikio Sato, Nice, France 1970.

He earned his degree in 1963 at the University of Tokyo. After holding appointments as professor at Osaka University and the University of Tokyo, in 1970 he moved to the Research Institute for Mathematical Sciences (RIMS), Kyoto University, where he served as director from 1987 to 1991. He was awarded the Asahi Prize in 1969, the Japan Academy Prize in 1976, the Fujiwara Prize in 1987, the Rolf Schock Prize in 1997 and the Wolf Prize in 2003. He was also chosen a Person of Cultural Merit by the Japanese government in 1984.

*Communicated by* Notices *Associate Editor Han-Bom Moon.*

*For permission to reprint this article, please contact:* `reprint-permission@ams.org`.

DOI: https://doi.org/10.1090/noti2931

## Mikio Sato, a Visionary of Mathematics

### *Pierre Schapira*

Mikio[1] Sato's passing on January 9, 2023 was very sad news for all of us who had the chance to meet him and share his vision of mathematics. Sato's vision was radically new and revolutionary, too much so to be immediately understood by the mathematical community, in particular, by the analysts. Sato's aim in developing the theory of "algebraic analysis" was to treat problems of analysis using the tools of algebraic geometry. However, Sato never made any special effort to propagate his ideas widely.

Sato did not write a lot, did not communicate easily, and attended very few meetings, but he invented an influential new way of doing analysis, and opened a new horizon in mathematics: the microlocal approach. Thanks to Sato, we understand that many phenomena which appear on a manifold are in fact the projection on the manifold of things living in the cotangent bundle to the manifold. Being "local" becomes, in some sense, global with respect to this projection.

*Pierre Schapira is a professor emeritus at the Sorbonne Université, CNRS IMJ-PRG, France. His email address is* `pierre.schapira@imj-prg.fr`.

[1]*This paper is a modified version of a text that already appeared in the* Notices of the AMS, *February 2007 after a first publication in French, in* La Gazette des Mathématiciens **97** *(2003) on the occasion of Sato's reception of the 2002/2003 Wolf Prize. The part concerned with number theory of these publications had benefited from the scientific comments of Jean-Benoît Bost and Antoine Chambert-Loir, and also Pierre Colmez for the last version. I warmly thank all of them.*

Mikio Sato also created a school, "the Kyoto school," among whom Masaki Kashiwara, Takahiro Kawai, Tetsuji Miwa, and Michio Jimbo should be mentioned.

Born in 1928,[2] Sato became known in mathematics in 1959–1960 with his theory of hyperfunctions. Indeed, his studies had been seriously disrupted by the war, particularly by the bombing of Tokyo. After his family home was burned down, he had to work as a coal delivery man and later as a school teacher. At age 29 he became an assistant at the University of Tokyo. He studied mathematics and physics, on his own.

To understand the originality of Sato's theory of hyperfunctions, one has to place it in the mathematical landscape of the time. Mathematical analysis from the 1950s to the 1970s was under the domination of functional analysis, marked by the success of the theory of distributions. People were essentially looking for existence theorems for linear partial differential equations (LPDE) and most of the proofs were reduced to finding "the right functional space," to obtain some a priori estimate and apply the Hahn–Banach theorem.

It was in this environment that Sato defined hyperfunctions [2] in 1959–1960 as boundary values of holomorphic functions, a discovery which allowed him to obtain a position at the University of Tokyo thanks to the clever patronage of Professor Iyanaga, an exceptionally open-minded person and a great friend of French culture. Next, Sato spent two years in the USA, in Princeton, where he unsuccessfully tried to convince André Weil of the relevance of his cohomological approach to analysis.

Sato's method was radically new, in no way using the notion of limit. His hyperfunctions are not limits of functions in any sense of the word, and the space of hyperfunctions has no natural topology other than the trivial one. For his construction, Sato invented local cohomology in parallel with Alexander Grothendieck. This was truly a revolutionary vision of analysis. And, besides its evident originality, Sato's approach had deep implications since it naturally led to microlocal analysis.

The theory of LPDE with variable coefficients was at its early beginnings in the years 1965–1970 and under the shock of Hans Lewy's example which showed that a very simple first order linear equation $(-\sqrt{-1}\partial_1 + \partial_2 - 2(x_1 + \sqrt{-1}x_2)\partial_3)u = v$ had no solution, even a local solution, in the space of distributions.[3] The fact that an equation had no solution was quite disturbing at that time. People thought it was a defect of the theory, that the spaces one had considered were too small to admit solutions. Of course, often just the opposite is true and one finds that the occurrence of a cohomological obstruction heralds interesting phenomena: the lack of a solution is the demonstration of some deep and hidden geometrical phenomena. In the case of the Hans Lewy equation, the hidden geometry is "microlocal" and this equation is microlocally equivalent to an induced Cauchy–Riemann equation on a real hypersurface of the complex space.

**Figure 2.** Mikio Sato and Pierre Schapira, ICM, Nice 1970.

In mathematics, as in physics, in order to treat phenomena in a given (affine) space, one is naturally led to compute in the dual space. One method to pass from a vector space to its dual, which is most commonly used in analysis, is the Fourier transform. This transform, being not of local nature, is not easily adapted to calculus on manifolds. By contrast, Sato's method is perfectly suited to manifolds. If $M$ is a real analytic manifold, $X$ a complexification of $M$, what plays the role of the dual space is now the conormal bundle $T^*_M X$ to $M$ in the cotangent bundle $T^*X$, something local on $M$. (Note that $T^*_M X$ is isomorphic to $\sqrt{-1}T^*M$.) In order to pass from $M$ to $T^*_M X$, Sato constructed a key tool of sheaf theory, the microlocalization functor $\mu_M(\cdot)$, the "Fourier–Sato" transform of the specialization functor $\nu_M(\cdot)$. This is how Sato defines the analytic wave front set of hyperfunctions (in particular, of distributions), a closed conic subset of the cotangent bundle, and he shows that if a hyperfunction $u$ is a solution of the equation $Pu = 0$, then its wave front set is contained in the intersection with $T^*_M X$ of the characteristic variety of the operator $P$ [3]. It is then clear (but it was not so clear at the time) that if you want to understand what happens on a real manifold, you better look at what happens on a complex neighborhood of the manifold.

Of course, at this time other mathematicians (especially Lars Hörmander) and physicists (e.g., Daniel Iagolnitzer) had the intuition that the cotangent bundle was the natural space for analysis, and in fact this intuition arose much earlier, in particular in the work of Jacques Hadamard, Fritz John, and Jean Leray. Indeed, pseudo-differential operators did exist before the wave front set.

In 1973, Sato and his two students, M. Kashiwara and T. Kawai, published a treatise [4] on the microlocal analysis of LPDE. Certainly this work had a considerable impact, although most analysts didn't understand a single word. Hörmander and his school then adapted the classical Fourier transform to these new ideas, leading to the now popular theory of Fourier-integral operators (see for example [5]).

[2] *See* [1] *for more details about Sato's life.*

[3] *The slightly simpler equation* $(\partial_1 + \sqrt{-1}x_1\partial_2)u = v$ *does not have any solution in the space of germs at the origin of distributions in* $\mathbb{R}^2$ *either, nor even in the space of germs of hyperfunctions.*

The microlocalization functor is the starting point of microlocal analysis but it is also at the origin of the microlocal theory of sheaves, due to Kashiwara and the author [6]. This theory associates to a sheaf on a real manifold $M$ its microsupport, a closed conic subset of the cotangent bundle $T^*M$ and allows one to treat sheaves "microlocally" in $T^*M$. The theory of systems of LPDE becomes essentially sheaf theory, the only analytic ingredient being the Cauchy-Kowalevsky theorem. One of the deepest results of [4] was the involutivity theorem which asserts that the characteristic variety of a microdifferential system (in particular, of a $\mathcal{D}$-module, see below) is co-isotropic. Similarly the microsupport of a sheaf is proved to be co-isotropic, which makes a link between sheaf theory and symplectic topology that is the origin of numerous important results.

Also note that by a fair return of things, microlocal analysis, through microlocal sheaf theory, appeared quite recently in algebraic geometry, under the impulse of Sasha Beilinson [7].

Already in the 1960s, Sato had the intuition of $\mathcal{D}$-module theory, of holonomic systems and of the $b$-function (the so-called Bernstein–Sato $b$-function). He gave a series of talks on these topics at Tokyo University but had to stop for lack of combatants. His ideas were reconsidered and systematically developed by Masaki Kashiwara in his 1969 thesis.[4] As its name indicates, a $\mathcal{D}$-module is a module over the sheaf of rings $\mathcal{D}$ of differential operators, and a module over a ring essentially means "a system of linear equations"[5] with coefficients in this ring. The task is now to treat (general) systems of LPDE. This theory, which also simultaneously appeared in the more algebraic framework developed by Joseph Bernstein, a student of Israel Gelfand, quickly had considerable success in several branches of mathematics. In 1970–1980, Kashiwara obtained almost all the fundamental results of the theory, in particular those concerned with holonomic modules, such as his constructibility theorem, his index theorem for holomorphic solutions of holonomic modules, the proof of the rationality of the zeroes of the $b$-function, and his proof of the (regular) Riemann-Hilbert correspondence.

The mathematical landscape of 1970–1980 had thus considerably changed. Not only did one treat equations with variable coefficients, but one treated systems of such equations and moreover one worked microlocally, that is, in the cotangent bundle, the phase space of the physicists. But there were two schools in the world: the $C^\infty$ school, in the continuation of classical analysis and headed by Hörmander who developed the calculus of Fourier integral operators,[6] and the analytic school that Sato established, which was almost nonexistent outside Japan and France.

France was a strategic place to receive Sato's ideas since they are based on, or parallel to, those of both Jean Leray and Alexander Grothendieck. Like Leray, Sato understood that singularities have to be sought in the complex domain, even for the understanding of real phenomena. Sato's algebraic analysis is based on sheaf theory, a theory invented by Leray in 1944 when he was a prisoner of war, clarified by Henri Cartan, and made extraordinarily efficient by Grothendieck and his formalism of derived categories and the *six opérations*.

Sato, who was motivated by physics as usual, then tackled the analysis of the $S$-matrix in light of microlocal analysis. With his two new students, M. Jimbo and T. Miwa [9], he explicitly constructed the solution of the $n$-points function of the Ising model in dimension 2 using Schlesinger's classical theory of isomonodromic deformations of ordinary differential equations. This naturally led him to the study of KdV-type nonlinear equations. In 1981, he and his wife Yasuko Sato (see [10] and [11]) interpreted the solutions of the KP-hierarchies as points of an infinite Grassmannian manifold and introduced his famous $\tau$-function. These results would be applied to other classes of equations and would have a great impact in mathematical physics in the study of integrable systems and field theory in dimension 2.

In parallel with his work in analysis and in mathematical physics, Sato obtained remarkable results in group theory and in number theory. He introduced the theory of prehomogeneous vector spaces, that is, of linear representations of complex reductive groups with a dense orbit. The important case where the complement of this orbit is a hypersurface gives good examples of $b$-functions (see [12, 13]).

In 1962, using a construction of auxiliary (Kuga-Sato) varieties, Sato also discovered how to deduce the Ramanujan conjecture on the coefficients of the modular form $\Delta$ from Weil's conjectures concerning the number of solutions of polynomial equations on finite fields. His ideas allowed Michio Kuga and Goro Shimura to treat the case of compact quotients of the Poincaré half-space and one had to wait another ten years for Pierre Deligne to definitely prove that Weil's conjectures imply Ramanujan and Petersson's conjecture.

Mikio Sato shared the Wolf Prize with John Tate in 2002/03. They also share a famous conjecture in number theory concerning the repartition of Frobenius angles.

[4] *See* [8] *for an overview of Kashiwara's work, a part of which was deeply influenced by Sato's ideas.*

[5] *According to Mikio Sato (personal communication), at the origin of this idea is the mathematician and philosopher of the 17th century, E. W. von Tschirnhaus.*

[6] *Many names should be quoted at this point, in particular those of Viktor Maslov and Vladimir Arnold.*

Let $P$ be a degree-3 polynomial with integer coefficients and simple roots. Hasse has shown that for any prime $p$ which does not divide the discriminant of $P$, the number of solutions of the congruence $y^2 = P(x) \pmod{p}$ is like $p - a_p$, with $|a_p| \leq 2\sqrt{p}$. When writing $a_p = 2\sqrt{p}\cos\theta_p$ with $0 \leq \theta_p \leq \pi$, the Sato–Tate conjecture predicts that these angles $\theta_p$ are distributed with the probability density $(2/\pi)\sin^2\theta$ (in absence of complex multiplication). Note that Tate was led to this conjecture by the study of algebraic cycles and Sato by computing numerical data.

Sato's most recent works are essentially unpublished (see however [14]) and have been presented in seminars attended only by a small group of people. They treat an algebraic approach of nonlinear systems of PDE, in particular of holonomic systems, of which theta functions are examples of solutions!

Looking back, 50 years later, we realize that Sato's approach to mathematics is not so different from that of Grothendieck, that Sato did have the incredible temerity to treat analysis as algebraic geometry and was also able to build the algebraic and geometric tools adapted to his problems.

His influence on mathematics is, and will remain, considerable.

# Personal Reminiscences of the Late Professor Mikio Sato

## *Takahiro Kawai and Yoshitsugu Takei*

Concerning the concrete description of Sato's works (including those in number theory) we refer the reader to the "Commentary" of the collected papers of M. Sato [15], and here we note some impressive episodes which lie behind them.

1) In 1973 in Nice, one of us (T. Kawai) was waiting for a friend to come to bring him to the airport. Suddenly Sato appeared and urged Kawai to take a taxi. As a result, Kawai was just in time for the flight. It was surprising as Sato was notorious for being rather loose in time: He himself said he had been amazed to learn the word "punctual."

2) In 1974, Kawai asked Sato whether he should begin the study of nonlinear differential equations or study the analytic $S$-matrix theory. His response was very clear: "You might study nonlinear differential equations at any time you would wish to do, but you should study the $S$-matrix theory before 30 years of age." It was a critically useful suggestion in Kawai's career.

**Figure 3.** Sato's family at the coast of the Mediterranean Sea near CIRM, Marseille, France, 1991.

3) These facts show how carefully Sato observed the surroundings. We guess it was tightly tied up with the fact that Sato had once been a night-school teacher to support his family.

4) At the same time, Sato always enjoyed discussions with young people: As the late Shintani said, "How to use Sato? Bring one to him, then he will give you back ten."

5) The most remarkable example was the "Sato-Komatsu seminar" in 1968 and 1969. It was a part of a very well-prepared project of the late Professor Komatsu to call Sato back to the theory of partial differential equations from number theory. It included an explanation of his results to Sato in New York in 1966 and his introductory lectures on the theory of hyperfunctions in 1967. The historically marvelous result of this seminar was Sato's creation of microfunction theory in 1969. We tremble to imagine what if Komatsu had not organized the seminar.

6) Parenthetically we note that [Kawai-Takei; Adv. Math. **80** (1990), 110–133] is a nice successor to F. Suzuki's talk in the seminar on first order equations.

**Figure 4.** At the party celebrating Kawai's 60th birthday in 2005: Sato, Kawai and his wife, and Kashiwara (from the left) for the first, and Pham and Sato (from the left) for the second.

7) Apparently Sato recalled the importance of the Sato-Komatsu seminar. When he was inspired by F. Pham's talk at the Delphi symposium in 1987 and became confident that the singular perturbation theory fortified with the Borel resummation was important in analysis, he urged the surrounding people to study this subject with him. His enthusiasm for this trial was very impressive particularly in view of the fact that he was busy as the director of the Research Institute for Mathematical Sciences of Kyoto University.

*Takahiro Kawai is a professor emeritus at RIMS, Kyoto University.*
*Yoshitsugu Takei is a professor in the Science and Engineering Department of Mathematical Sciences at Doshisha University. His email address is* ytakei@mail.doshisha.ac.jp.

8) Fortunately his enthusiasm found a nice resonator: T. Aoki figured out Stokes curves with the help of a computer. In parenthesis, we feel nostalgic to see the first figure which consists of many tiny arrows indicating the vector fields.

**Figure 5.** At the party celebrating Komatsu's 80th birthday in 2015: Komatsu, Kataoka, and Sato (from the left).

9) A great person has a broad view of life: Sato left the study of the singular perturbation theory to young students when he felt they had begun to study the novel subject, and he tried to attack new issues. "Papers are like decorations," he once said. We refer the reader to [16] for the later development of the theory.

Many thanks for all your support and advice, Mentor Sato!

# Mikio Sato, My Mentor

## *Masaki Kashiwara*

I first saw Professor Mikio Sato when I was a senior in the University of Tokyo in 1968. He was then starting to construct the theory of microlocal analysis.

He was the main speaker of the seminar on algebraic analysis. This seminar was organized by Professor Hikosaburo Komatsu (1935–2022) in order to understand the work of Sato. It was a weekly seminar just started in 1968. Sato gave numerous talks explaining his theory. One of his main ideas was to consider a real manifold inside a complex manifold (a complex neighborhood). This already appeared in his theory of hyperfunctions. However he expanded this idea to a great extent. Hyperfunctions are obtained as sums of boundary values of holomorphic functions defined in a tubular neighborhood (in a complex neighborhood). We can think that a hyperfunction has milder singularities if these tubular neighborhoods are bigger. It was the starting point of his theory of microlocal analysis. Through these considerations, he found that those generalized functions live in the cotangent bundle. Of course we can study functions locally on a real manifold. We can study functions on the cotangent bundle much more precisely. This is the essence of microlocal analysis.

In order to construct the theory, he had to start from the beginning to establish a basic theory by hand. For example, he constructed the theory of relative cohomologies of sheaves in his own way in order to introduce microfunctions. When he constructed microfunctions, he introduced the notion of the Fourier-Sato transform of sheaves. The original Fourier transform is a correspondence between functions on a vector space and ones on the dual vector space. Sato found a correspondence between sheaves on a vector space and ones on the dual vector space, which is now called the Fourier-Sato transform.

He also noticed the importance of $D$-modules in the study of linear partial differential equations. Especially, he introduced the notion of maximally overdetermined systems (of linear partial differential equations), which is now known under the name of "holonomic $D$-modules." These new ideas were presented in the Algebraic Analysis Seminar. I was very impressed by the overflowing ideas of Sato, and I decided to study algebraic analysis when I became a master course student.

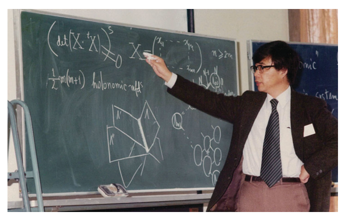


**Figure 6.** Mikio Sato, Katata conference, 1971.

In 1970, Professor Sato moved to Research Institute for Mathematical Sciences (RIMS), Kyoto University. In the next year, I followed him to RIMS. I worked with him and Takahiro Kawai to complete the theory of microlocal analysis. Inspired by the work of Maslov and Egorov, Sato introduced the notion of microdifferential operators which live on the cotangent bundle and act on microfunctions. Moreover, he discovered that microfunctions are invariant by symplectic transformation on the cotangent bundle. This opened a new understanding in analysis.

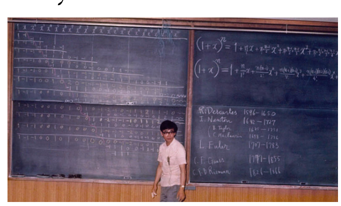


**Figure 7.** Mikio Sato at RIMS Public Lecture, Kyoto 1976.

It is very sad that Professor Sato passed away on January 9, 2023. However, he remains my mentor and his marvelous work will remain forever.

*Masaki Kashiwara is a professor at the Research Institute for Mathematical Sciences at Kyoto University. His email address is* `masaki@kurims.kyoto-u.ac.jp`.

# Mikio Sato and the Creation of the Tau Function

## *Barry M. McCoy*

I first met Professor Sato in 1980 at a week-long workshop sponsored by Erhart Seminar Training in San Francisco where Sato and his two junior collaborators, Michio Jimbo and Tetsuji Miwa, reported on their recent work which united the mathematical theory of isomonodromic deformation theory with the physics of the statistical mechanics of the Ising model. This was made possible by the invention of the tau function.

The journey to the invention of the tau function begins with the 1964 paper of John Myers [17] who discovered that the scattering of electromagnetic radiation by a finite conducting strip can be described with the use of a particular Painlevé III function. This was the first time which Painlevé functions had appeared in Physics.

The next step was published in 1976 by Wu, McCoy, Tracy, and Barouch [18] who found that Myers' work could be adapted to make an exact computation of the two point correlation function of the two-dimensional Ising model in the continuum limit near the critical temperature. Remarkably, exactly the same Painlevé III function found by Myers was needed for the construction.

This was the background for the invention of the tau function. In a series of six remarkable papers from 1978 to 1980, Sato and his collaborators Miwa, Jimbo, and Mori [9],[19] put these previous computations into the vastly more general mathematical setting of isomonodromic deformation of first order systems of $m \times m$ ordinary differential equations with $n$ singular points $a_j$

$$\frac{dY(z)}{dz} = \{\sum_{j=1}^{n}\sum_{k=0}^{r_k} \frac{A_{j,k}}{(z-a_j)^{k+1}} - \sum_{k=1}^{r_\infty} A_{\infty,k} z^{k-1}\} Y(z).$$

The study of these isomonodromic problems was first posed by Riemann and is one of Hilbert's 24 problems. In the Fuchsian case where $r_k = r_\infty = 0$, the conditions on the $A_j$ for the deformations of the singular points $a_j$ to be isomonodromic (the Schlesinger equations) were known by the beginning of the 20th century. The great advance made by Sato et al was the discovery in paper 2 of [9] of a closed one form which can be extended to all the Fuchsian systems

$$\omega = \frac{1}{2}\sum_{j\neq k} TrA_j(a)A_k(a)d\,\ln(a_j - a_k)$$

where the $A_j(a)$ satisfy Schlesinger's equations and the eigenvalues of $A_j$ are all distinct with no integer differences. From this, the tau function is defined as

$$d\ln\tau = \omega.$$

This is generalized to non Fuchsian cases in paper 3 of [9] and [19]. Furthermore, in [19] it was shown that this one form is a Hamiltonian which characterizes the tau function.

For a particular $2\times2$ linear system, this tau function was then shown to characterize the scaled two point function of the Ising model and was computed from deformation theory in terms of a Painlevé V function which is equivalent to the Painlevé III function of [18]. Even more impressive is the result obtained soon after by Jimbo and Miwa [20] that the diagonal two point function of the Ising model $C(N,N)$ on the lattice is the tau function of a $2\times2$ linear problem with four Fuchsian singularites which gives a Painlevé VI function.

Following the discoveries of [9] Jimbo, Miwa, Mori, and Sato [19] found that correlation functions for the impenetrable Bose gas in one dimension are tau functions of another $2\times2$ system which characterizes Painlevé V functions. Subsequently tau functions have revolutionized the study of random matrix theory by expressing distributions of eigenvalues of random matrices in terms of Painlevé functions. More recently, further applications of tau functions to the Ising model have been made by Witte [*Nonlinearity* **29** (2016) 131–160] who found that the row correlation $C(0,N)$ is the tau function of a $2\times2$ system with 6 singularities.

In the years following the discovery of the tau function, there have been many mathematical studies of the one forms $\omega$ for $m\times m$ systems with $m>2$. In 1984, Myers [*Physica D* **11** (1984) 51–89] studied the wave scattering from a broken corner by use of the exact same methods he previously used for the strip to obtain a solution which involves a $2\times2$ matrix version of the Painlevé III equation found in [17], which is based on a $4\times4$ linear system.

Subsequent to the Erhart Seminar Training workshop, I had the privilege of meeting Professor Sato many times at the Research Institute for Mathematical Sciences in Kyoto and attending his seminars. I am convinced that the applications of Sato's tau function to physics have not yet been fully realized.

*Barry M. McCoy is a Distinguished Professor Emeritus at the CNYang Institute for Theoretical Physics at StonyBrook University. His email address is* mccoy@max2.physics.sunysb.edu.

# In Memoriam of Mikio Sato

## *Craig Tracy*

I spoke with Professor Sato only a couple of times so my comments here necessarily lack first-hand personal stories. What I can do is highlight some of Professor Sato's profound work in integrable systems and statistical physics.

In the 1970s, T. T. Wu, B. M. McCoy, E. Barouch, and I [18, 21] were involved in the analysis of the two-point correlation function, $\langle\sigma_{00}\sigma_{MN}\rangle$, of the two-dimensional, zero field Ising model in the *massive scaling limit*.[7] Using work of J. Myers [17] on scattering from a finite strip we derived the following

$$\frac{\langle\sigma_{00}\sigma_{MN}\rangle}{\mathcal{M}_T^2} \longrightarrow \begin{Bmatrix} \sinh(\frac{1}{2}\psi(r)) \\ \cosh(\frac{1}{2}\psi(r)) \end{Bmatrix}$$
$$\times \exp\left[\frac{1}{4}\int_r^\infty \left(\sinh^2\psi(s) - (\frac{d\psi(s)}{ds})^2\right) s\, ds\right] \qquad (1)$$

where the upper (lower) equation holds for $T \to T_c$ from above (below) and $\mathcal{M}_T$ is a normalization factor.[8] The function $\psi$ satisfies the differential equation in the scaling variable $r$

$$\frac{d^2\psi}{dr^2} + \frac{1}{r}\frac{d\psi}{dr} = \frac{1}{2}\sinh(2\psi) \qquad (2)$$

with boundary condition $\psi(r) \sim \frac{2}{\pi}K_0(r)$ as $r \to \infty$. The substitution $\eta(r) = e^{-\psi(r)}$ transforms (2) into a Painlevé III equation.

In the period 1978–1981, Sato, Miwa, and Jimbo [9] vastly extended the 2-point analysis to $n$-point functions, $\langle\sigma_{a_1}\cdots\sigma_{a_n}\rangle$. Introducing monodromy preserving deformations of the 2D Euclidean Dirac equation, they derived a set of deformation equations (function of the points $a_j, j = 1, \dots, n$). When $n = 2$, the SMJ deformation equations reduce to the Painlevé III equation; thus "explaining" its appearance in the WMTB result. We remark that the analysis of the SMJ deformation equations is still open, e.g., connection formulas. For $n = 2$ the connection problem is solved in [21, 22].

Sato together with Jimbo, Miwa, and Môri [19] studied $\tau(a) \coloneqq \det(1 - K)$ where $K$ is the integral operator with kernel

$$\frac{1}{\pi}\frac{\sin(x-y)}{x-y}\chi_I(y).$$

Here $I = \bigcup_{j=1}^n (a_{2j-1}, a_{2j})$ and $\chi_I(y)$ is the characteristic function of the set $I$. In random matrix theory, $\tau(a)$ is the probability no eigenvalue of the rescaled GUE lies in $I$. A deformation theory for the Fredholm integral equation is derived [19]; and for the special case of $I = (-t/2, t/2)$, the deformation equations reduce to Painlevé V. This is the first appearance of Painlevé transcendents in random matrix theory.

The notion of a $\tau$-function is now a central object in integrable systems thanks to the work of Professor Sato and his colleagues Miwa and Jimbo.

*Craig Tracy is a Distinguished Professor Emeritus in the department of mathematics at UCDavis. His email address is* tracy@math.ucdavis.edu.

[7] *When the vertical and horizontal interactions are equal, the massive scaling limit is* $R = \sqrt{M^2 + N^2} \to \infty$, $\xi(T) \to \infty$ *with* $R/\xi(T) \to r$, $r > 0$. *Here* $\xi(T)$ *is the correlation length,* $T$ *is temperature, and* $T_c$ *the critical temperature where* $\xi(T) \to \infty$ *as* $T \to T_c$.

[8] *For* $T < T_c$, $\mathcal{M}_T$ *is the spontaneous magnetization, see* [18] *for details.*

# Memories from 2019 and 1980

## *Tetsuji Miwa*

**Figure 8.** Hiroko and Masaki Kashiwara, Mikio and Yasuko Sato, Setsuko Miwa (from the left), Kyoto 2019.

My last meeting with Professor Mikio Sato was in November 2019, at Hakusa-Sonso, a Japanese restaurant near Ginkakuji. A photo is at hand, taken at a luncheon meeting. I sang a song of Mikis Theodrakis in the meeting.

In the summer of 1980, during the Soliton Meeting in Crete, Greece, one day was spent in an excursion up to the mountains and down to the beach, where an evening concert was held. There were several huge speakers, which blasted songs including some by Mikis Theodrakis. I was there in the first line of the audience with my family, and Mikio and Yasuko Sato and Michio Jimbo.

# Professor Mikio Sato: Personal Recollections

## *Michio Jimbo*

1. In 1973, I was an undergraduate student at Tokyo, wondering whether to pursue my graduate study in Tokyo or Kyoto. Upon hearing my hesitation, Professor Sato kindly offered me an opportunity to visit him in Kyoto. This was

*Tetsuji Miwa is a professor emeritus at Kyoto University. His email address is* sanlunz442@gmail.com.

*Michio Jimbo is a professor emeritus at Rikkyo University. His email address is* jimbomm@rikkyo.ac.jp.

quite a surprise, because I had never met him except at the oral examination earlier that summer. He took pains to explain to me in informal language what he had been doing with Kawai and Kashiwara, i.e., microlocal analysis. I did not understand much, but Sato's enthusiasm and vision left a strong impression on me.

Some time after that, I had a chance to chat with Daisuke Fujiwara, whose lectures I was attending. When I mentioned my meeting with Sato, he said "If I were you I would go to Kyoto without a moment's hesitation." Realizing how much Sato was esteemed by young mathematicians, I followed Fujiwara's advice.

2. I was extremely lucky to be able to join Sato's research project, together with Miwa, during the years 1975–1980. This was the time when he had finished the monumental work [4] with Kawai and Kashiwara, and his interests were turned to problems in theoretical physics. The outcome of his research during this period, "holonomic quantum fields," is summarized in the commentary on Sato's collected papers [15].

Sato aimed at applying microlocal analysis to the study of S matrices and Green's functions in quantum field theory. In order to gain insight into the problem, he spent some time looking for a simple nontrivial model where computations can be done explicitly. This led him to study the scaling limit of the Ising model.

Then he learned about the work of Wu, McCoy, Tracy, and Barouch [18] which gives the two point correlation functions in terms of a Painlevé transcendent. He soon recognized the role of monodromy-preserving deformation of linear differential equations behind the Ising model. The relevant deformation theory was formulated smoothly, but it remained unclear how to relate it to the calculation of spin-spin correlators. The missing key was found when he came up with the idea of inserting fermions into the correlators. (Though this may sound commonplace today, one should bear in mind that it was long before the advent of conformal field theory.)

**Figure 9.** Tetsuji Miwa, Hermann Flaschka, Mikio Sato, and Jimbo (from the left). Clarkson College, USA, August 1979.

During the course, I was able to witness the way his thoughts shaped and developed. It was the most valuable occasion for me.

3. When I talked to Sato in private, he was always very kind and willing to explain on the blackboard. If it was over a meal he would start writing formulas on the back of a chopstick bag. I then started to feel that the ideas behind his work are very natural and even intuitively clear.

By all accounts, his lectures and conference talks were energetic and fascinating. He would start talking about the big picture, often meander into various related subjects, and eventually come to the point—only when the time was up.

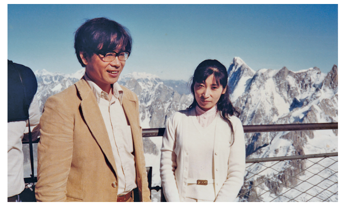


**Figure 10.** Mikio Sato and Yasuko Sato. Aiguille du Midi, France, September 1979.

4. Reading mathematical literature, Sato liked concisely written books and papers containing succinct statements, rather than those giving lengthy explanations including all the technical details. When proofs became necessary it was easier for him to produce them on his own. Since his youth, one of his favorite readings had been the *Encyclopedia of Mathematics* edited by the Japan Mathematical Society. He carefully chose genuinely important papers on a subject and carried copies with him. One such example was Lax's paper on the Lax pair for soliton equations.

**Figure 11.** Leon Takhtajan, Alexander Zamolodchikov, and Alexander Belavin (front from the left); Jimbo, Tetsuji Miwa, Mikio Sato and his son Nobuo (back from the left). Kyoto, October 1988.

5. Sato got interested in soliton theory because of its many structural similarities with holonomic quantum fields. He was particularly impressed by Hirota's original method for solving various soliton equations.

Most soliton equations, such as the KdV and Toda equations, have an infinite number of mutually commuting flows, constituting hierarchies of higher order equations. Around 1979–1980, Mikio and Yasuko Sato performed a thorough-going study of Hirota's bilinear equations. At a workshop at RIMS, they presented a list of *all* Hirota equations for KdV, KP, and other hierarchies to rather high orders, and reported that their numbers are given by partition numbers. It took them more than 1000 hours to carry through the computation with a pocket calculator. (One has to recall again that this was the time long before personal computers became available.) It was a mystery to everyone else why they were going through these elaborate computations.

The answer was revealed in Sato's talk in January 1981. He explained that the space of solutions to the KP hierarchy forms an infinite-dimensional Grassmannian, and Hirota's equations are the Plücker relations defining this manifold. This beautiful result, obtained after a long inductive process, is reminiscent of Euler's work.

Sato once made the following comment about Euler: *At first glance, his work might seem a collection of miscellaneous results. However I cannot but think that all was coherent in himself—that he had been led by something clearly visible to his own mind. His work is marvelous precisely for that reason. I feel that he was really a great master, greater than is usually considered.*

It seems to me that these words best describe Sato's own mathematics.

### References

[1] Emmanuel Andronikof, *Interview with Mikio Sato*, Notices Amer. Math. Soc. **54** (2007), no. 2, 208–222. MR2285125

[2] Mikio Sato, *Theory of hyperfunctions. I*, J. Fac. Sci. Univ. Tokyo Sect. I **8** (1959), 139–193. MR114124

[3] Mikio Sato, *Regularity of hyperfunctions solutions of partial differential equations*, Actes du Congrès International des Mathématiciens (Nice, 1970), Gauthier-Villars Éditeur, Paris, 1971, pp. 785–794. MR650826

[4] Mikio Sato, Takahiro Kawai, and Masaki Kashiwara, *Microfunctions and pseudo-differential equations*, Hyperfunctions and pseudo-differential equations (Proc. Conf., Katata, 1971; dedicated to the memory of André Martineau), Lecture Notes in Math., Vol. 287, Springer, Berlin-New York, 1973, pp. 265–529. MR420735

[5] Lars Hörmander, *The analysis of linear partial differential operators. III*: *Pseudodifferential operators*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 274, Springer-Verlag, Berlin, 1985. MR781536

[6] Masaki Kashiwara and Pierre Schapira, *Sheaves on manifolds*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 292, Springer-Verlag, Berlin, 1990. With a chapter in French by Christian Houzel, DOI 10.1007/978-3-662-02661-8. MR1074006

[7] A. Beilinson, *Constructible sheaves are holonomic*, Selecta Math. (N.S.) **22** (2016), no. 4, 1797–1819, DOI 10.1007/s00029-016-0260-z. MR3573946

[8] Pierre Schapira, *Fifty years of mathematics with Masaki Kashiwara*, Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. I. Plenary lectures, World Sci. Publ., Hackensack, NJ, 2018, pp. 97–110. MR3966724

[9] Mikio Sato, Tetsuji Miwa, and Michio Jimbo, *Holonomic quantum fields. I*, Publ. Res. Inst. Math. Sci. **14** (1978), no. 1, 223–267, DOI 10.2977/prims/1195189284. MR499666

[10] Mikio Sato and Yasuko Sato, *Soliton equations as dynamical systems on infinite-dimensional Grassmann manifold*, Nonlinear partial differential equations in applied science (Tokyo, 1982), North-Holland Math. Stud., vol. 81, North-Holland, Amsterdam, 1983, pp. 259–271. MR730247

[11] Mikio Sato, *The KP hierarchy and infinite-dimensional Grassmann manifolds*, Theta functions—Bowdoin 1987, Part 1 (Brunswick, ME, 1987), Proc. Sympos. Pure Math., vol. 49, Amer. Math. Soc., Providence, RI, 1989, pp. 51–66, DOI 10.1090/pspum/049.1/1013125. MR1013125

[12] Mikio Sato and Takuro Shintani, *On zeta functions associated with prehomogeneous vector spaces*, Ann. of Math. (2) **100** (1974), 131–170, DOI 10.2307/1970844. MR344230

[13] M. Sato and T. Kimura, *A classification of irreducible prehomogeneous vector spaces and their relative invariants*, Nagoya Math. J. **65** (1977), 1–155. MR430336

[14] M. Sato, *$\mathcal{D}$-modules and nonlinear systems*, Integrable systems in quantum field theory and statistical mechanics, Adv. Stud. Pure Math., vol. 19, Academic Press, Boston, MA, 1989, pp. 417–434, DOI 10.2969/aspm/01910417. MR1048603

[15] Collected Papers, Mikio Sato, Springer, *in Press*.

[16] Takahiro Kawai and Yoshitsugu Takei, *Algebraic analysis of singular perturbation theory*, Translations of Mathematical Monographs, vol. 227, American Mathematical Society, Providence, RI, 2005. Translated from the 1998 Japanese original by Goro Kato; Iwanami Series in Modern Mathematics, DOI 10.1090/mmono/227. MR2182990

[17] John M. Myers, *Wave scattering and the geometry of a strip*, J. Mathematical Phys. **6** (1965), 1839–1846, DOI 10.1063/1.1704731. MR186000

[18] Barry M. McCoy and Tai Tsun Wu, *The two-dimensional Ising model*, Harvard University Press, Cambridge, MA, 1973, DOI 10.4159/harvard.9780674180758. MR3618829

[19] Michio Jimbo, Tetsuji Miwa, Yasuko Môri, and Mikio Sato, *Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent*, Phys. D **1** (1980), no. 1, 80–158, DOI 10.1016/0167-2789(80)90006-8. MR573370

[20] Michio Jimbo and Tetsuji Miwa, *Studies on holonomic quantum fields. XVII*, Proc. Japan Acad. Ser. A Math. Sci. **56** (1980), no. 9, 405–410. MR603054

[21] Barry M. McCoy, Craig A. Tracy, and Tai Tsun Wu, *Painlevé functions of the third kind*, J. Mathematical Phys. **18** (1977), no. 5, 1058–1092, DOI 10.1063/1.523367. MR473322

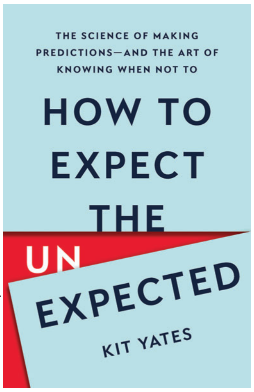
[22] Craig A. Tracy, *Asymptotics of a $\tau$-function arising in the two-dimensional Ising model*, Comm. Math. Phys. **142** (1991), no. 2, 297–311. MR1137066

### Credits

Figure 1 and Figure 2 are courtesy of Pierre Schapira.
Figure 3 and Figure 5 are courtesy of Takashi Aoki.
Figure 4 is courtesy of Minoru Nakano.
Figure 6 and Figure 7 are courtesy of Masaki Kashiwara.
Figure 8 is courtesy of Tetsuji Miwa.
Figures 9, 10, and 11 are courtesy of Michio Jimbo.

# BOOKSHELF

*New and Noteworthy Titles on our Bookshelf*
**May 2024**



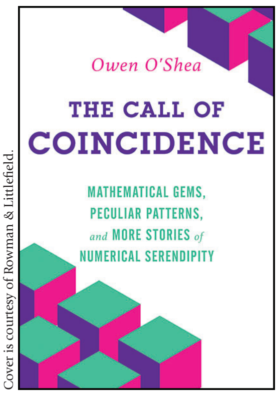

Cover is courtesy of Basic Books.

***How to Expect the Unexpected***
*Basic Books, 2023, 448 pp.*
By Kit Yates

I'll start by considering the word "unexpected." What do you think it means? Does an unexpected event have an element of surprise? Is it a phenomenon that happens rarely or with low probability? In the book *How to Expect the Unexpected*, something that is unexpected plays into our personal biases of what we expect to occur. This nonfiction, general-interest mathematics book takes the reader on a journey of the biases, effects, and psychological phenomena that can often be explained or resolved with mathematics.

Yates starts the book by detailing common psychological tricks used by psychics to give a satisfactory reading and superstitious rituals utilized by professional athletes. While that sounds nonacademic, soon Yates begins to weave in mathematical topics to illustrate his points. As the chapters progress, the reader learns about more mathematical topics, such as linearity bias (the brain's desire to extrapolate linearly) and illusory correlation (where one perceives a relationship when there isn't one, such as finding a pattern in a random data set or attributing a deeper meaning to a coincidence). Bayes' theorem and Nash equilibrium are discussed without technical definitions, and game theory is applied to both the plots of movies and the world's political stage.

The book includes well-known counterintuitive facts and thought experiments such as the birthday paradox, the stockbroker scam, and the prisoner's dilemma, where the results start to make sense after a careful analysis of the numbers. An instructor might use these examples to introduce a mathematical concept to the class, perhaps in a math for liberal arts course. You could add this recreational book to your personal library or gift to your graduating students for them to enjoy.




Cover is courtesy of Rowman & Littlefield.

***The Call of Coincidence***
*Prometheus, 2023, 192 pp.*
By Owen O'Shea

For many years, Owen O'Shea has been fascinated by numerical patterns and coincidences, and his latest book continues this tradition. The book features mathematical tidbits, ranging from arithmetic oddities to surprising number theory results that could be used in a discrete mathematics course. Parts of this recreational math book are more traditional; because historical anecdotes are juxtaposed with facts involving $\sqrt{2}$, including a proof of its irrationality, I believe the intended audience for this book is interested in the calculations as well as the musings.

The chapters in this book have whimsical titles, such "Geometry Gems," "A Few Words about Simplicity, Mathematics, and Everything There Is," and "Two Lightning Calculation Tricks and Sundry Other Matters." Interludes with a fictional numerologist are entertaining. This numerologist is particularly skilled at finding numerical coincidences that might belong on memes or internet posts, such as "Martin Gardner was born on the $293 + 1$ day in the year 1914. There are $293 \times 1$ primes less than 1914." Some of the chapters discuss much deeper and possibly unanswerable questions, such as "Does mathematics exist outside of human minds?" and "Why is there *something* rather than *nothing*?" Examples from physics and philosophy are woven within the mathematical anecdotes, and O'Shea offers opinions and theories that will leave the reader thinking.

I enjoyed the inclusion of various puzzles and problems for the reader to solve, some numerical and others logical. After reading this recreational mathematics book, you are sure to believe that you can find "surprising" coincidences anywhere if you look hard enough.

---

*This Bookshelf was prepared by* Notices *Associate Editor Emily J. Olson.*

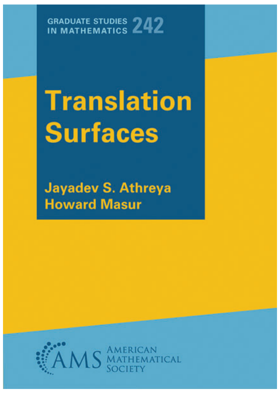
*Appearance of a book in the* Notices *Bookshelf does not represent an endorsement by the* Notices *or by the AMS.*

*Suggestions can be sent to* `notices-booklist@lists.ams.org`.

DOI: https://doi.org/10.1090/noti2942

# AMS BOOKSHELF

*The AMS Book Program serves the mathematical community by publishing books that further mathematical research, awareness, education, and the profession while generating resources that support other Society programs and activities. As a professional society of mathematicians and one of the world's leading publishers of mathematical literature, we publish books that meet the highest standards for their content and production.* Visit **bookstore.ams.org** to explore the entire collection of AMS titles.




***Translation Surfaces***
By Jayadev S. Athreya
and Howard Masur.
GSM/242, 2024, 179 pp.,
`https://bookstore.ams.org/gsm-242`

A translation surface is a two-dimensional manifold obtained from a collection of polygons in the Euclidean plane, where the sides of the polygons are grouped into pairs that are parallel and of the same length. The sides in each pair are oriented so that the Euclidean translation taking one to the other preserves this orientation, and the polygon of one of the oriented sides lies to the left of the oriented side and the polygon corresponding to the other side to the right. The archetypical example of a translation surface is a "flat torus" which is obtained from a parallelogram by identifying the opposite sides by translations, which inherits the flat structure from the plane. A translation surface has its origins in the study of billiards, and has connections to algebraic geometry, number theory, low-dimensional topology, and dynamics. The study of translation surfaces and their moduli spaces has seen enormous growth in recent decades, and results in related areas have been recognized with prestigious awards such as Fields Medals, Breakthrough Prizes, and Clay Research Awards. There are several excellent surveys on this topic, but this is the first book that gives a comprehensive introduction to the subject.

In Chapter 1, the authors analyze the geometry of the flat torus and associated dynamical and counting problems. The results about flat tori and their moduli spaces described in Chapter 1 motivate the discussion in higher-genus settings. Chapter 2 introduces three different definitions of translation surfaces as polygons in the Euclidean plane, holomorphic differentials on Riemann surfaces, and (singular) geometric structures on two-dimensional manifolds. The equivalence of these three perspectives is at the heart of the study of the geometry and dynamics of translation surfaces and their moduli spaces but is often opaque in the literature. Athreya and Masur do a wonderful job explaining this equivalence in a clear way.

Chapter 3 starts with a brief discussion of Teichmüller and moduli spaces of Riemann surfaces and delves into the moduli space of translation surfaces where the equivalence of two translation surfaces is carefully defined in the previous section. This moduli space of translation surfaces admits a natural "stratification" by topological data (the genus, the number of cone points, and the excess angle at each cone point). Fixing this data, one obtains "strata of translation surfaces." The components of strata admit natural coordinates, and a natural Lebesgue measure for which the total volume of the space of *unit-area surfaces* is finite.

The linear action of the group $\mathrm{GL}^+(2,\mathbb{R})$ of real matrices with positive determinant on the Euclidean plane, induces an action of $\mathrm{GL}^+(2,\mathbb{R})$ on the space of translation surfaces which preserves strata, and where the subgroup $\mathrm{SL}(2,\mathbb{R})$ preserves the area 1 locus. The dynamics and ergodic theory of subgroups of $\mathrm{SL}(2,\mathbb{R})$ acting on strata, and their role as renormalization dynamics for flows on translation surfaces, takes up most of Chapter 4 and Chapter 5.

The remaining chapters give a broad overview of the field since the 1990s while concealing technical details in well-chosen "Black Boxes." Instead, precise references and exercises are given to lead the reader to the correct places. The book ends with a discussion of "Lattice surfaces" which are translation surfaces with "maximal symmetry" and "optimal dynamics," originally studied by Veech. Although they form a measure zero subset in any stratum, they are nevertheless dense in each stratum and many natural examples studied in the literature have this property.

*Translation Surfaces* by Athreya and Masur is a wonderful addition to the literature with lots of beautiful pictures and clear exposition. I expect it to become the standard reference book for researchers and graduate students working in the area.

*This AMS Bookshelf is by Caglar Uyanik, an assistant professor at the University of Wisconsin, Madison. His email address is* `cuyanik@wisc.edu`.



DOI: https://doi.org/10.1090/noti2933

# Are University Budget Cuts Becoming a Threat to Mathematics?

*Edgar Fuller*

## Introduction

Mathematics as an area of study occupies an important place in higher education. Due in part to its utility in other disciplines as well as its role in student learning, institutions of higher education (IHEs) often have large numbers of mathematics faculty with different balances of teaching and research in different ranks and appointment structures. Most flagship IHEs, especially state land-grant institutions, have large undergraduate populations taking mathematics courses in many cases built around the widespread use of calculus [Bre12] and the connections between mathematics and science, technology, and engineering. These connections have made mathematics departments essential to universities [OR12] and emphasized the critical role math plays in supporting student success [RRN20, KFW$^+$23] in all areas of postsecondary education. We tend to take that essential nature of mathematics at the undergraduate level, and for research universities at the graduate level, as a given, but that characterization no longer holds for some IHEs.

In September of 2023, the Board of Governors for West Virginia University (WVU) voted to discontinue the graduate program in mathematics, ending a period of 35 years of doctoral studies in mathematics and a much longer history of providing master's-level opportunities to students at the institution. The vote resulted in the loss of 16 of the 48 faculty lines in the School of Mathematical and Data Sciences [Spi23] including the termination of tenure-track as well as teaching faculty. The changes at WVU, characterized as "Academic Transformation," have received significant national attention summarized here [Ful23] in more detail. The president and provost at WVU based these cuts on recommendations from the consulting firm rpkGroup [Moo23] hired in response to declining enrollment. They echo similar recommendations made by the same consulting group to the University of Kansas in 2022, and represent growing concerns in higher education about projected future decreases in enrollment known as the "enrollment cliff" [Kno23]. In Kansas, 42 academic programs were recommended for discontinuation, with reasoning, as in prior evaluations in Arkansas and Missouri [Ful23], that questioned the funding of mathematics and other STEM degree programs. In yet another review, the Kansas Board of Regents suggested that mathematics requirements should be removed altogether [Ful23] for some majors.

Data from the Integrated Postsecondary Education Data System (IPEDS) [Ful23] shows that since the overall postsecondary enrollment maximum in 2010, student populations at both land-grant and research universities have increased as shown in Figure 1, even after experiencing small declines during the Covid-19 lockdown. Total enrollments at four-year institutions have decreased recently, but the decreases at WVU follow neither of these patterns.

While these trends may change, overall student interest in attending large land-grant and research focused institutions remains consistent with the enrollments that supported their programs over the last decade. That broader pattern nationally suggests that factors other than

*Edgar Fuller is associate director of the STEM Transformation Institute and Distinguished University Professor of Mathematics at Florida International University. He served as chair of the Department of Mathematics at West Virginia University from 2008–2018. His email address is* `efuller@fiu.edu`.

*For permission to reprint this article, please contact:* `reprint-permission@ams.org`.

DOI: https://doi.org/10.1090/noti2934

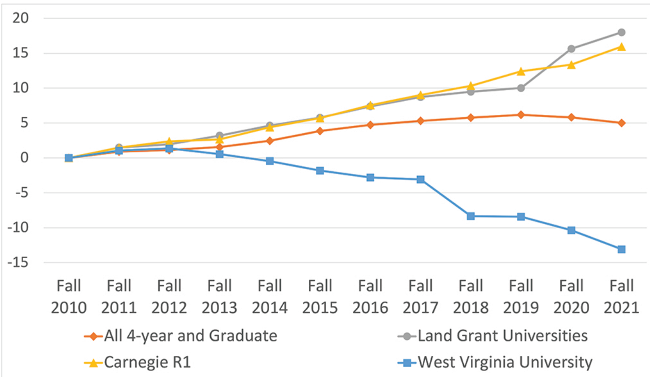



**Figure 1.** Percent change in enrollment since 2010.

enrollment trends, such as already existing plans to restructure programs, may be responsible for these choices. Even in the context of the decreases in student enrollment at WVU, the removal of the graduate programs in mathematics represents a fundamental change in the mission and values of an institution of higher education. Calls to shift resources away from mathematics education and research in particular ignore the powerful role it plays within an institution and in society.

At WVU, the restructuring runs even deeper. A total of 28 of 338 degree programs, including both the MS and PhD in mathematics [Uni23] have been discontinued and more than 146 of the tenured and nontenured faculty at WVU, or roughly 16% of total faculty university-wide, will either resign by the end of the 2023–2024 academic year, retire early during the 2023–2024 academic year, or be terminated. Within this group, 44 tenured or tenure-track faculty will be involuntarily separated at the end of the academic year, leading the American Association of University Professors to state in a recent letter [Ful23] that with these measures enacted "...tenure, and by extension academic freedom, cannot be said to exist at West Virginia University." At least 112 of the positions will be removed from academic colleges, or more than 15% of the 880 academic faculty at WVU. Concerns presented by consulting firms [Ful23] focus on the belief that universities should only allocate resources to disciplines with some specific value [Moo23], but how that value is measured is often inconsistent and not reflective of the true state of an institution or its mission. Administrators at WVU ignored the fact that tuition revenue generated from mathematics courses averaged $19.3 million per year over the last three years, making it one of the most efficient departments on campus with operational costs annually around 35% of generated revenue. Instead they chose to focus solely on the fact that the total value of grants obtained by faculty was less than 1 million per year even though the total was above that cutoff as recently as 2018. Afterward, President Gordon Gee stated that "...I don't believe every aspect of math is essential" [Ful23] and others in the Office of the Provost have also disputed any need for theoretical mathematics. The university has since begun removing mathematics coursework from degree programs in business and other areas.

When universities, or the consultants they hire, overlook the significant components of what mathematics provides to higher education, it signals a fundamental change in the strategic goals of those institutions. The true impact of a discipline like mathematics becomes lost in the narratives used to justify decision making, and these trends begin to affect mathematics departments within large flagship research universities as well as universities which have already seen dramatic resource reallocations. Leaders in mathematics both inside and aligned with higher education must work together to correct these narratives and reinforce the transformative potential of mathematics.

Land-grant institutions, for example, were created by the Morrill Act "in order to promote the *liberal and practical education* of the industrial classes in the several pursuits and professions in life." [Arc62]. Policymakers at land-grant IHEs may discuss the practical uses of education as the primary driver of programmatic choices, but the use of the word liberal here implies a specific mission as well. The government sought to increase access to education in a way that would broaden the opportunities available to citizens, and increase their freedom to choose new occupational pathways. The act provided a mechanism to create institutions of higher education so that communities would have access to as many areas of learning as possible. Land-grant IHEs in particular represent essential paths to higher education for students in their respective states, a role that should be maintained. Many land-grant and Hispanic Serving Institutions (HSIs) have developed mathematics departments which excel [Gar19] in research while meeting the needs of the communities they serve, providing indispensable access to general education mathematics courses as well as research and graduate study beyond those needs. To overlook the added value of graduate study in mathematics at an IHE to its overall research and education missions, especially at these universities, neglects the critical place they occupy in our communities. We can't predict when the next Katherine Johnson, a graduate student in mathematics at WVU in 1940 [Mal20], will arrive from Appalachia asking to study the subjects that will be used to launch a rocket to the stars, or when the next Addison Fischer, who earned bachelor's and master's degrees in Mathematics at WVU from 1966–1972, will want to study mathematics and then go on to build multi-billion dollar corporations. IHEs are founded on the notion of providing that access, and the strong support of mathematics amplifies academic quality and breadth across every program. Universities, and the mathematics departments within them, should certainly operate efficiently and provide measurable returns, but support for graduate programs fulfills the higher purpose of making strong educational opportunities broadly accessible to the people we serve. As a profession we must continue to demonstrate this impact, even when consultants lack the expertise to recognize it. We have an obligation to work to make sure that the systems in which we operate see the power of mathematics to elevate and inspire consistently and indisputably. Not long before his passing, Bob Moses summarized the power of mathematics [Mil21], saying that

> *Math literacy will be a liberation tool for people trying to get out of poverty and the best hope for people trying not to get left behind.*

As a profession, we often view the importance of our discipline as self-evident since the role of mathematics in other disciplines, or just as a practical tool in our lives, generally remains well-established. Reinforcing that importance requires ongoing work, however, since changing priorities, enrollment trends, and funding stresses at IHEs create challenges that recur. In the mid 1990s, the University of Rochester proposed the elimination of its mathematics graduate programs due to financial difficulties [JLL96] as a part of its "Renaissance Plan," but the national response from multiple science and higher education organizations led to a reversal of the decision. Years later, the department was thriving [Jac97], giving some indication of departmental strengths that the initial plan overlooked and providing some sense that targeting graduate programs as a cost saving measure may not be the best strategy. The events at WVU and other IHEs highlight a potentially changing landscape in which even core disciplines such as mathematics may be targeted for reduction when not aligned with strategic goals that have departed from the historical mission of universities and colleges. Multiple factors have led higher education to this place, and it is worth considering research and graduate study in mathematics in the context of some of the drivers of that change.

## Changes in Enrollment and State Appropriations

Enrollment declines at some IHEs have influenced many of the current discussions of program efficiency, but public IHEs have already been experiencing resource strain in recent years [Off23] as decreasing state appropriations have forced more dependence on tuition and federal funding [Ful23]. Figure 2 shows the change in state-appropriated funding for the operation of public IHEs by state from 2001 to 2022. Over the last twenty years, the majority of states have decreased appropriations substantially, in constant dollars over the last twenty years. Data from the Association of Public and Land-grant Universities (APLU), shown in Figure 3, illustrates the decline in the state funding allocations to four-year public IHEs over the last four decades, where it has decreased to approximately half the level of the late 1980s.

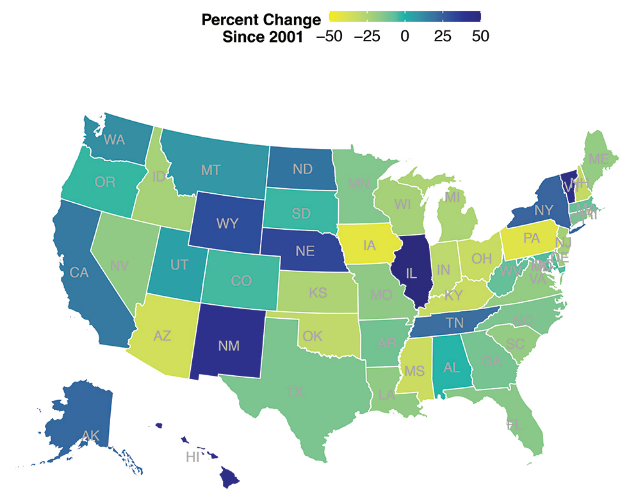



**Figure 2.** Percent change in state-appropriated funding for higher education by state from 2001 to 2022 in CPI adjusted constant dollars (data from [Off23]).

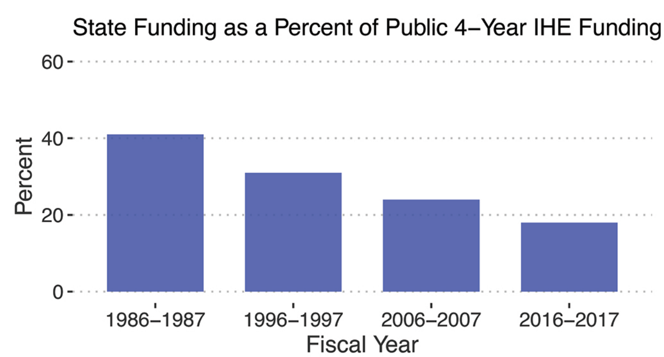



**Figure 3.** Percent of total university budgets for public four year IHEs from state appropriations—APLU.

State-appropriated funds, particularly for land-grant universities, remains an important part of the funding model for public universities [Ful23] though this source has decreased for IHEs in 36 of 50 states for the last two decades, leading universities to increase tuition and turn to other sources of revenue to make up for shortfalls. In WV, for example, state funding has decreased by 8.5% in that period. As a result, tuition has steadily risen at state institutions, increasing the student share of the cost of attending an IHE [Ful23].

As the total of state funding has decreased, the appropriated funds can be used in different ways by IHEs as well. From 2001 to 2021 the fraction of state-appropriated funding for higher education used for IHE operations nationally remained relatively constant, averaging 78.2%, while in WV the fraction specifically used for operations decreased from 65.5% to 56.2%. The reallocated 10% of the overall already decreasing state appropriations moved to the research, agricultural extension programs, public health care services, and medical education (RAM) category indicating a use outside of academic operations. In 2022, WV allocated 18.08% of total higher education appropriations specifically to public health care services and medical education, followed by state funding used to support agriculture extension programs (5.27%), university affiliated hospital support (1.67%), and state support for research (0.38%). Revenue and expense data from WVU [Uni23] show that medical education programs operate at a loss, and so this shift of state appropriations toward that portion of the university appears to be a strategic effort to support medical activity at the expense of core academic operations. The combined decrease in appropriations overall with allocation to other uses applies stress to the ability of state IHEs to meet mission goals. IHEs then begin to look at reductions to even high enrollment departments such as mathematics that normally are seen as essential to the proper functioning of a university in an effort to balance budgets. The resulting reviews often focus on the cost of a department, major enrollments, or other metrics, but they ignore many of the important impacts of academic departments such as mathematics at an IHE.

## University and Mathematics Research Activity

As resources have tightened at even larger universities, arguments made in support of mathematics research and graduate programs sometimes focus on the classification of an IHE as a Research—Very High Activity institution, commonly referred to as R1 status. A total of 146 IHEs are currently classified as R1, or about 3.7%, making this an important distinction for some universities. Those within the R1 group form a fairly broad collection of private and public institutions and for those IHEs R1 status is considered a measure of prestige that enhances the reputation of an IHE as well as that of its faculty. Both IHEs and mathematics departments within them appeal at times to this high-level research status for recruitment and to procure additional research contracts. These contracts in turn provide additional revenue to universities that can be critically important during times of financial stress, leading some administrators to emphasize grant-funded research over other departmental contributions. R1 status, however, does not necessarily capture the specifics of research at an IHE.

Grouping IHEs into private, public, Hispanic-serving (HSI) and land-grant (LG) institutions, some differences in the characteristics of the institutions appear in Table 1. Land-grant institutions (35%), which include some public, private, and HSI IHEs, form a large subset of R1s. HSIs which are R1s (11 of 146, 7.5%) are a growing subset that includes four land-grant IHEs. A *Forbes* analysis estimated that HSIs have been underfunded in 16 states by almost $13 billion in the last 30 years [BW23], with the mean number of faculty [Ful23] reported as 63.5% of the average number for other R1s as of 2021 though their undergraduate enrollment is considerably higher. They award a substantial number of PhDs and significantly more undergraduate degrees than other R1s.

The table summarizes the mean values of several key aspects of R1 IHEs including total undergraduate enrollment (UG), total graduate enrollment (GR), number of Bachelors (BS) and PhDs (PhD) awarded, and the number of full-time faculty in ladder rank (assistant, associate, and full professors). The mean number of students enrolled at the undergraduate level provides the clearest discrepancy between the public and private R1s. External funding listed here as Science and Engineering Research and Development Expenditures (SERD, in millions of dollars) are considerably higher for private and land-grant institutions, while the average number of faculty and the number of PhDs awarded are almost the same. This metric along with the number of postdoctoral and nonfaculty research staff with doctorates (PD) is the largest factor impacting R1 ranking as of the 2021 analysis. PD numbers are calculated differently at some IHEs, with some counts including faculty equivalent appointments that involve teaching and others large numbers of laboratory support staff with no academic duties.

| | **Private** | **Public** | **HSI** | **LG** |
|---|---|---|---|---|
| Number | 39 | 107 | 11 | 51 |
| SERD | 579.7 | 388.8 | 235.6 | 483.7 |
| UG | 8966 | 25241 | 28382 | 26331 |
| GR | 9315 | 7787 | 6838 | 7792 |
| Mean PhD | 331 | 345 | 287 | 415 |
| Mean BS | 2629 | 6228 | 6882 | 6700 |
| Mean FAC | 1471 | 1474 | 962 | 1624 |
| Mean PD | 887 | 443 | 299 | 580 |

**Table 1.** Characteristics of R1 universities by group.

As a part of their "Academic Transformation" process, administrators at WVU identified departments with graduate programs with less than $1,000,000 per year in grant related research expenditures as measured by the SERD data in Table 1, and reviewed student enrollments in their programs. The subsequent evaluation of the School of

Mathematical and Data Sciences led to the discontinuation of all graduate mathematics programs. NSF HERD survey data indicate that if only federal sources are included, as with WVU, more than 50% of R1 mathematics departments would be under similar reviews and in jeopardy. SERD funding levels contribute to the rankings as described in [Ful23] and in particular to the Aggregated Research Index Score (ARIS) used as one factor for R1 status since it is compiled from external funding, postdoctoral researchers and PhDs conferred. This index and the rank of an IHE increase as external funding grows, but the contributions from mathematics departments and graduate programs in mathematics, however, extend well beyond research funding, and many other factors derived from a department also contribute to ranking and reputation. Figure 4, for example, shows that the number of TT faculty in mathematics by itself correlates positively with increases in the ARIS aggregate measure of research productivity in the R1 grouping.

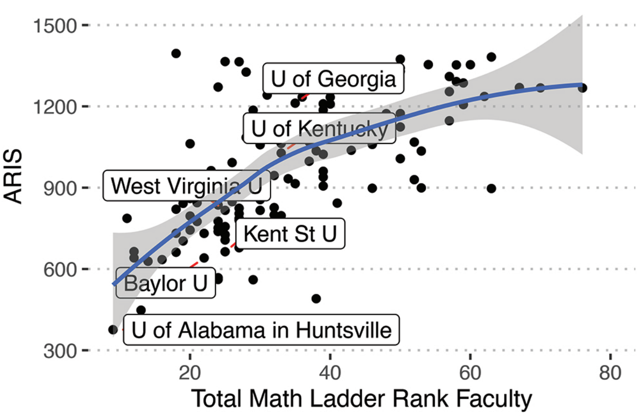



**Figure 4.** Scatterplot of the ARIS weighted ranking by the total of tenure line math faculty.

Similar analysis [Ful23] shows that other factors such as lower student-faculty ratio in undergraduate mathematics courses, higher numbers of graduate students in mathematics, and higher numbers of mathematics PhDs produced also positively correlate with ARIS ranking, supporting the notion that an active and well-resourced research program in mathematics supports R1 status in ways that are independent of research funding levels. These correlations also give some insight into why 97% (142 of 146) of R1 IHEs maintained graduate programs conferring doctorates in mathematics up until 2023. Students and other researchers benefit from access to advanced mathematics, and this access enhances research programs in multiple STEM departments, economics and business as well as health sciences.

In order to meet their instructional needs in mathematics, mathematics departments within R1 IHEs appoint a combination of tenured or tenure-track faculty (TT) with non-tenure-track teaching faculty (NTT). At most of the IHEs in the United States, departments engage in the instruction of service courses such as calculus, differential equations, or linear algebra to sometimes thousands or tens of thousands of students per year and this demands a strategic balance of faculty with different teaching loads and research expectations. The populations of TT and NTT mathematics faculty along with the ratio of undergraduate students in all math courses to total math faculty (UG/Math FAC) are shown in Table 2. Other metrics such as the ratio of TT to NTT faculty, average federal research funding (RF, in $1000s) in mathematics, the mean number of postdoctoral researchers in math (Mean PD), the five-year mean number of mathematics PhDs conferred (Mean PhDs), UG mathematics majors (UG math %) as a percent of the total population of UG students taking math courses, and graduate PhD and MS mathematics students (GR math %) as a percentage of all students in graduate mathematics courses are shown as well.

| | Private | Public | HSI | LG |
|---|---|---|---|---|
| UG/Math FAC | 61.5 | 131.8 | 131.2 | 124.5 |
| Math RF | 2927 | 1948.7 | 727.1 | 2748.6 |
| Mean TT | 28.7 | 37.9 | 44.6 | 43.4 |
| Mean NTT | 7.4 | 15.3 | 15.6 | 16.7 |
| TT/NTT | 5.5 | 3.9 | 4.3 | 5.0 |
| Mean PD | 13.1 | 7.5 | 4.4 | 10.6 |
| Mean PhDs | 10.0 | 10.6 | 9.1 | 12.3 |
| UG math % | 31.5% | 30.9% | 25.7% | 32.2% |
| GR math % | 4.6% | 5.2% | 10.3% | 4.8% |

**Table 2.** Characteristics of mathematics departments at R1s.

For land-grant IHEs, the ratio of undergraduates in math courses to total department faculty in the first row of Table 2 correlates negatively with the ARIS research ranking of the university for values greater than 150:1 [Ful23]. As the ratio increases, aggregate research indicators begin to drop, suggesting that an institution with ratios nearing 200:1 may have structural issues with the ability of research faculty to participate in research projects and to procure external funding. It is unclear how effectively an R1 can then lower the number of both TT and NTT faculty in mathematics as WVU plans to do and still serve its educational community properly or even meet the instructional needs of its student body in mathematics. With ratios nearing or above 300:1, all STEM disciplines begin to struggle with insufficient access to needed mathematics coursework for their students, and the profile of the university as a whole changes. R1 status depends, at least in part, on the trends shown here for mathematics resources. The way in which R1 mathematics departments meet instructional needs interacts synergistically with measures of research productivity and these include both the number of TT faculty in a department as well as the numbers of graduate students, PhDs conferred and grant funding. The actions proposed at the IHEs listed in the introduction, however, and in particular at WVU, indicate policy choices that assume that R1 IHEs can operate without core graduate and research capacities in mathematics as well as other areas. The motivations for these changes reflect more than funding shortfalls, and ultimately indicate a belief that portions of those departmental missions can be sacrificed without damaging others. Increases in R1 ranking factors can be obtained from other sources of external funding such as health sciences, and, as noted in the previous section, state funding can be realigned to support the RAM portions of a university mission. This appears to be the case for WV. The actions taken in response to recent funding concerns within land-grants such as WVU conflict with their core missions primarily because they limit student access to the best undergraduate course opportunities, but they also eliminate access to graduate study and potentially limit research capacity overall. Ignoring the role of mathematics as discussed above becomes especially problematic in the context of land-grant R1s where the research and teaching missions exist primarily to provide the best opportunities to the citizens of a state. Relying on R1 status alone as an indicator of the excellence in academic programs, or as a basis for maintaining any program at an IHE, then becomes problematic since that status may only reflect research expenditures that are completely independent of mathematics, physics, engineering or any scholarly work within those programs. For land-grant, HSI, HBCUs, and other public IHEs with much broader missions, it is essential that resource allocation take all missions into account, and that measures of excellence incorporate multiple factors including grant-supported research programs as well as a variety of other indicators. Departments with research programs must continue to seek funding for their research, but more generally departments should work to increase the appeal of mathematics as a major, incorporate more computational and data science components into mathematics programs, and work to retain majors at critical departure points. Penalizing departments with program discontinuation and terminating productive and talented faculty, however, punishes students by limiting their options unnecessarily.

## Conclusions

Internal or external forces may change resource allocations and departments must become more adept at working to demonstrate impact in order to have a voice in those allocations. A vice president for research can reallocate the funding for a graduate program to another department or a provost can redirect faculty resources away from STEM programs and into other areas in order to meet what they feel is a better strategic goal for a university, and mathematics departments must be ready to engage with other departments and administrators to demonstrate the implications of those choices. We must also clearly articulate outside of the university in legislative circles and with the public the value of the contributions of mathematics to education and research in general. We must be ready to provide evidence that supports the effectiveness of our teaching as well as outcomes that demonstrate the impact of our research and the value of our degrees [Ful23]. Without this dialogue, we risk being characterized as expendable. At the core of large public institutions, our mission is to serve the people of our communities by providing opportunities to study and learn in areas that will benefit society in ways that sometimes are not reflected in the number of people that graduate from a specific degree program or the total amount of grant funding supporting the research of a department. Narratives that focus on a single aspect of a department or ignore significant portions of what a department does will not effectively support the missions of an IHE.

The administrators at WVU and the state leaders of West Virginia overlooked the contributions of mathematicians to the educational opportunities of the people in that state [Ful23] and to its flagship university. It may seem at first glance that this is a local concern dependent on financial or enrollment issues, but these arguments appear in other states and localities, spread in part by consultants with no commitment to the communities those universities serve. External funding from grants or federal appropriations are important for the support of research and even graduate programs, but the pursuit of that funding for its own sake may reinforce alarming shifts away from academics by creating the notion that all efforts must be aligned with that pursuit. The metrics in that case become the goal [Ful23] of those operations with no consideration for the quality of core academics. In the example of WVU, leaders have aligned their choices with health science operations, medical school support, and with degrees they perceive to be the most employable options for students. These supposed efforts to "modernize" programs or provide students with the "best" career options become arguments that ignore the essence of what an IHE does. How academia responds may determine which universities continue to be institutions of true learning and which become minimally effective professional training programs. It is possible to build a "Very High Activity" research institution around a health sciences paradigm or industrial applications, but in the case of at least public land-grant universities, it ignores the fundamental intent of their creation. Gordon Gee, president of WVU, responded [Ful23] to concerns about discontinuing the PhD in Mathematics by saying "Someone else is going to have a great PhD program in mathematics. And you know what? God bless them." True flagship universities with effective leadership see opportunity in providing the most *liberal and practical education* possible and move towards it, not away. The students of states like West Virginia will have to look elsewhere to find that education, and for those students the lack of access is a failure that is not reflected in any financial analysis, research status, or ranking. The "academic transformation" of WVU is a model for changes to the underlying strategic mission of IHEs that does not reflect the original intent of publicly funded higher education as spending in essential areas such as mathematics, world languages or libraries are further reduced to compensate for interest payments on growing debt or private aircraft travel for administrators [Ton23].

It remains to be seen what the long-term impact of the program cuts and faculty reductions will be at WVU. The overall structure of large, public R1s continues to attract students as shown in Figure 1 for academic as well as nonacademic reasons. IHEs such as WVU can currently maintain R1 status and emphasize other programs, at least in principle, but the students and faculty will ultimately decide if the quality of what remains meets their needs or if they will search elsewhere for academic opportunity. The guidelines for Carnegie classification are being revised to include "a wider set of dimensions that define an institution, and not distort the process by overweighting one single aspect of an institution's purpose and activities" [Led23] so it may be that a university that focuses solely on the value of grants obtained may be classified as that, and one with a broad collection of graduate offerings will be recognized that way. Removing graduate mathematics redefines WVU, and it is arguable that the loss of mathematics as well as world languages and other fields diminishes the university in important ways. The outcomes at other IHEs such as Emporia State University in Kansas where student enrollments dropped another 12.5% [Ful23] after program changes suggest that universities considering similar actions should do so carefully. Mathematically speaking, they may accelerate the declines to which they are responding instead of turning them around.

Just as the value of an education is more than the credential acquired, the value of mathematics extends from more than just mathematics course requirements in a degree program or even a degree in mathematics. Mathematics gives students an opportunity to understand ideas ranging from differential equations to graph theory and to use these ideas to solve problems in finance, renewable energy technology or molecular structure prediction. Mathematics is necessary for IHEs to make those educational opportunities available, and other disciplines need advanced mathematics to facilitate their research. It exposes students to a wealth of knowledge that begins with mathematics but spreads out across all areas of scientific inquiry. Planning and resource allocation at an IHE must account for both the core educational impact of mathematics as well as the broader, sometimes harder to measure, impact of mathematics research on other disciplines and student opportunities in order for an institution to excel.

ACKNOWLEDGMENTS. Many thanks to Cristian Potter for his help in gathering the data related to mathematics departments at R1s, and to Karen Saxe for helpful suggestions and comments during the writing of this article. Additional thanks are extended to the many investigative reporters, journalists, faculty, and students who have published the many works documenting different aspects of the situation at WVU collected more completely here [Ful23].

## References

[Arc62] National Archives, *Morrill Act of 1862*, accessed September 5, 2023, `https://www.archives.gov/milestone-documents/morrill-act`.

[Bre12] David M Bressoud, *Second year calculus: from celestial mechanics to special relativity*, Springer Science & Business Media, 2012.

[BW23] Jeremy Bauer-Wolf, *University of Kansas looks to cut 42 academic programs*, Higher Ed Dive, 2023, accessed September 19, 2023, `https://www.highereddive.com/news/16-states-underfunded-land-grant-hbcus-by-over-12b-biden-admin-says/694020/`.

[Ful23] Edgar J. Fuller, *Supplementary Material for Are Math Departments At Risk During University Budget Cuts?*, 2023.

[Gar19] Gina A Garcia, *Defining "servingness" at Hispanic-serving institutions (HSIs): Practical implications for HSI leaders*, American Council on Education (2019).

[Jac97] Allyn Jackson, *Whatever happened to Rochester? Two years later, mathematics is getting accolades*, Notices of the American Mathematical Society **44** (1997), no. 11, 1463–1466.

[JLL96] Arthur Jaffe, Joseph Lipman, and Morton Lowengrub, *U. of Rochester Plan to Cut Mathematics Is Recipe for Disaster*, The Chronicle of Higher Education, 1996, accessed October 6, 2023, `https://www.chronicle.com/article/u-of-rochester-plan-to-cut-mathematics-is-recipe-for-disaster/`.

[KFW⁺23] Laird Kramer, Edgar Fuller, Charity Watson, Adam Castillo, Pablo Duran Oliva, and Geoff Potvin, *Establishing a new standard of care for calculus using trials with randomized student allocation*, Science **381** (2023), no. 6661, 995–998, https://www.science.org/doi/abs/10.1126/science.ade9803.

[Kno23] Liam Knox, *Grasping for a foothold on the enrollment cliff*, Inside Higher Ed, 2023, accessed October 31, 2023, https://www.insidehighered.com/news/business/revenue-strategies/2023/05/12/grasping-foothold-enrollment-cliff.

[Led23] Doug Lederman, *A New Approach to Categorizing Colleges*, Inside Higher Ed, 2023, accessed November 1, 2023, https://www.insidehighered.com/news/institutions/2023/11/01/major-overhaul-coming-key-framework-organizing-higher-ed.

[Mal20] Shirley M Malcom, *Katherine Johnson (1918–2020)*, Science **368** (2020), no. 6491, 591–591.

[Mil21] Courtland Milloy, *Bob Moses saw math as the path to equality. School systems should build upon his work.*, The Washington Post, 2021. Published July 27, 2021, accessed September 6, 2023, https://www.washingtonpost.com/local/bob-moses-algebra-math-black-students/2021/07/27/74e41f24-eef5-11eb-81d2-ffae0f931b8f_story.html.

[Moo23] Josh Moody, *16 states underfunded land-grant HBCUs by over $12b, Biden admin says*, Inside Higher Ed, 2023, accessed September 11, 2023, https://www.insidehighered.com/news/2022/02/18/university-kansas-plans-cut-42-academic-programs.

[Off23] State Higher Education Executive Officers, *State higher education finance: Fy 2022*, 2023, accessed September 6, 2023, https://shef.sheeo.org/data-downloads.

[OR12] Steve Olson and Donna Gerardi Riordan, *Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. report to the president.*, Executive Office of the President (2012).

[RRN20] Daniel L Reinholz, Chris Rasmussen, and Elena Nardi, *Time for (research on) change in mathematics departments*, International Journal of Research in Undergraduate Mathematics Education **6** (2020), 147–158.

[Spi23] Laura Spitalniak, *WVU board approves dramatic academic cuts to address $45m deficit*, Higher Ed Dive, 2023, accessed September 19, 2023, https://www.highereddive.com/news/wvu-board-approves-academic-cuts-45m-deficit/693857/.

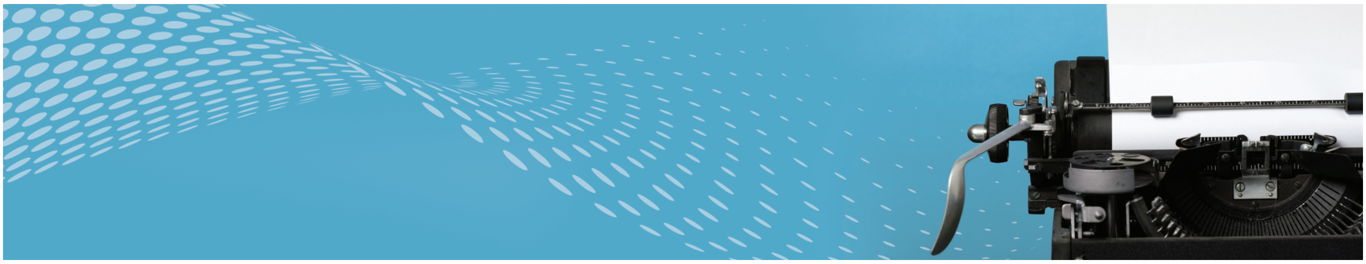
[Ton23] Mike Toney, *WVU expenses paid to aircraft charter provider have risen since Gee called for decreasing costs*, Charleston Gazette-Mail, 2023, accessed October 6, 2023, https://www.wvgazettemail.com/news/education/wvu-expenses-paid-to-aircraft-charter-provider-have-risen-since-gee-called-for-decreasing-costs/article_7a5d42eb-75b5-5216-b20b-0c8788cdc4fd.html.

[Uni23] West Virginia University, *Academic program portfolio review*, 2023. Retrieved September 13, 2023 from https://provost.wvu.edu/academic-transformation/academic-program-portfolio-review?asd.

Edgar Fuller

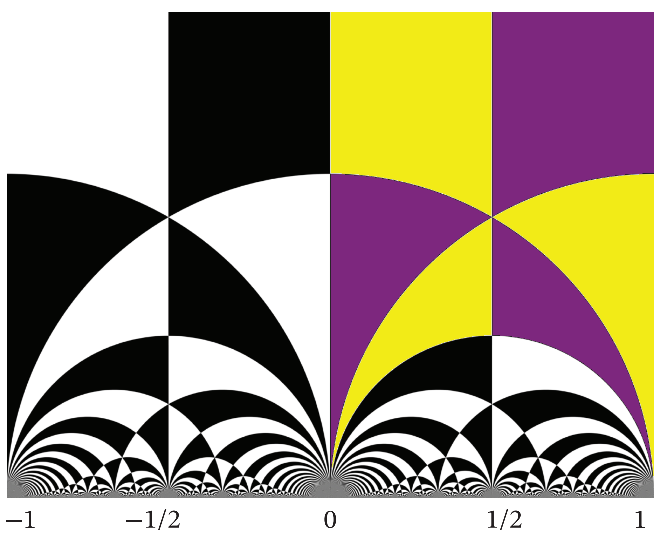
### Credits

Figures 1–4 and photo of Edgar Fuller are courtesy of Edgar Fuller.



# The Moduli Space of Acute Triangles

*John C. Baez*




In mathematics we often like to classify objects up to isomorphism. Sometimes the classification is discrete, but sometimes we have a notion of when two objects are "close." Then we can make the set of isomorphism classes into a topological space called a "moduli space." A simple example involves acute triangles. We can define two triangles to be isomorphic if they are similar. Then the moduli space of acute triangles is just the space of similarity classes of such triangles.

As a step toward explicitly describing this moduli space, first consider triangles in the complex plane with labeled vertices—that is, with a specified "first," "second," and "third" vertex. Every such triangle is similar to one with the first vertex at 0, the second at 1, and the third at some point in the upper half-plane. This triangle is acute precisely when its third vertex lies in this set:

$$T = \left\{ z \in \mathbb{C} \;\middle|\; \mathrm{Im}(z) > 0,\ 0 < \mathrm{Re}(z) < 1,\ |z - \tfrac{1}{2}| > \tfrac{1}{2} \right\}.$$

So, we say $T$ is the moduli space of acute triangles with labeled vertices. This set is colored yellow and purple above; the yellow and purple regions on top extend infinitely upward.

To get the moduli space of acute triangles with unlabeled vertices, we must mod out $T$ by the action of $S_3$ that permutes the three vertices. To act by a permutation we relabel the vertices of the corresponding triangle, apply a similarity so that the new first and second vertices lie at 0 and 1, and then record the new location of the third vertex. The six yellow and purple regions in $T$ are "fundamental domains" for this $S_3$ action: that is, they each contain exactly one point from each orbit. If we reflect a labeled triangle corresponding to a point in a yellow region across the line $\mathrm{Re}(z) = \frac{1}{2}$, we get a triangle corresponding to a point

*John C. Baez is a professor of mathematics at UC Riverside. His email address is* baez@math.ucr.edu.

*Communicated by* Notices *Associate Editor Han-Bom Moon.*



DOI: https://doi.org/10.1090/noti2932

in a purple region, and vice versa. Points on the boundary between two regions correspond to isosceles triangles. All six regions meet at the point that corresponds to an equilateral triangle.

The moduli space of acute triangles is closely related to a more famous moduli space: that of elliptic curves. An **elliptic curve** is a torus equipped with the structure of a complex manifold. But for those unfamiliar with complex manifolds, we can describe elliptic curves more concretely as follows. Given a parallelogram in the complex plane we can identify its edges and get a torus called an elliptic curve. Two parallelograms are said to give isomorphic elliptic curves if we can get one parallelogram by translating, rotating and/or dilating the other. Note that we do not include reflections here: for example, applying complex conjugation to some parallelogram can yield a parallelogram giving a nonisomorphic elliptic curve.

For any point $z \in \mathcal{H}$ we can form a parallelogram with vertices $0, 1, z$, and $z+1$. If we identify the opposite edges of this parallelogram we get an elliptic curve. We can get every elliptic curve this way, up to isomorphism. So, it is interesting to ask when two points in $\mathcal{H}$ give isomorphic elliptic curves.

To answer this question, we introduce the group $\mathrm{GL}(2,\mathbb{Z})$, consisting of invertible $2\times 2$ integer matrices, and its subgroup $\mathrm{SL}(2,\mathbb{Z})$ consisting of those matrices with determinant 1. Both these groups act on the upper half-plane

$$\mathcal{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$$

as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto \frac{az+b}{cz+d}.$$

Each of the light or dark regions shown above is a fundamental domain for the action of $\mathrm{GL}(2,\mathbb{Z})$. Elements of $\mathrm{GL}(2,\mathbb{Z})$ with determinant $-1$ map light regions to dark ones and vice versa. Elements with determinant 1 map light regions to light ones and dark ones to dark ones. People more often study the action of the subgroup $\mathrm{SL}(2,\mathbb{Z})$. The union of any light region and dark one sharing an edge forms a fundamental domain of $\mathrm{SL}(2,\mathbb{Z})$.

We can now answer our question: two points $z, z' \in \mathcal{H}$ give isomorphic elliptic curves if and only if $z' = gz$ for some $g \in \mathrm{SL}(2,\mathbb{Z})$. Thus the quotient space $\mathcal{H}/\mathrm{SL}(2,\mathbb{Z})$ is the **moduli space of elliptic curves**: points in this space correspond to isomorphism classes of elliptic curves. For details, see any good introduction to elliptic curves [2, 3].

Since $T$ is the union of six fundamental domains for $\mathrm{GL}(2,\mathbb{Z})$—the yellow and purple regions in the figure—it is the union of three fundamental domains for $\mathrm{SL}(2,\mathbb{Z})$. There is thus a map

$$p\colon T \to \mathcal{H}/\mathrm{SL}(2,\mathbb{Z})$$

from the moduli space of acute triangles with labeled vertices to the moduli space of elliptic curves, and generically this map is three-to-one. This map is not onto, but if we take the closure of $T$ inside $\mathcal{H}$ we get a larger set

$$\overline{T} = \left\{ z \in \mathbb{C} \;\middle|\; \mathrm{Im}(z) > 0,\ 0 \le \mathrm{Re}(z) \le 1,\ |z - \tfrac{1}{2}| \ge \tfrac{1}{2} \right\}$$

whose boundary consists of points corresponding to right triangles. Then $p$ extends to an onto map

$$p\colon \overline{T} \to \mathcal{H}/\mathrm{SL}(2,\mathbb{Z}).$$

The existence of this map suggests that from any acute or right triangle in the plane we can construct an elliptic curve. This is in fact true. How can we understand this more directly?

Take any acute or right triangle in the complex plane. Rotating it 180° around the midpoint of any edge we get another triangle. The union of these two triangles is a parallelogram. Identifying opposite edges of this parallelogram, we get an elliptic curve! There are three choices of how to build this parallelogram, one for each edge of the original triangle, but they give isomorphic elliptic curves. Moreover, two acute or right triangles differing by a translation, rotation or dilation give isomorphic elliptic curves. Even better, every elliptic curve is isomorphic to one arising from this construction. Thus, this construction gives a map from $\overline{T}$ onto $\mathcal{H}/\mathrm{SL}(2,\mathbb{Z})$, and with a little thought one can see that this map is $p$.

I learned about the moduli space of acute triangles from James Dolan. In fact all the moduli spaces mentioned here are better thought of as moduli "stacks": stacks give a way to understand the special role of more symmetrical objects, like isosceles and equilateral triangles. Behrend [1] has written an introduction to stacks using various moduli stacks of triangles and the moduli space of elliptic curves as examples. Though he does not describe the map $p$ or its stacky analogue, his work is a nice way to dig deeper into some of the ideas discussed here.

## References

[1] K. Behrend, *Introduction to algebraic stacks*, Moduli spaces, London Math. Soc. Lecture Note Ser., vol. 411, Cambridge Univ. Press, Cambridge, 2014, pp. 1–131, `https://secure.math.ubc.ca/~behrend/math615A/stacksintro.pdf`. MR3221293

[2] Anthony W. Knapp, *Elliptic curves*, Mathematical Notes, vol. 40, Princeton University Press, Princeton, NJ, 1992. MR1193029

[3] Neal Koblitz, *Introduction to elliptic curves and modular forms*, 2nd ed., Graduate Texts in Mathematics, vol. 97, Springer-Verlag, New York, 1993, DOI 10.1007/978-1-4612-0909-6. MR1216136

## Credits

Opener is courtesy of Roice Nelson.


# AMS Prizes and Awards

## NEW! I. Martin Isaacs Prize for Excellence in Mathematical Writing


I. Martin Isaacs

The I. Martin Isaacs Prize is awarded for excellence in writing of a research article published in a primary journal of the AMS in the past two years.

### About this Prize

The prize focuses on the attributes of excellent writing, including clarity, grace, and accessibility; the quality of the research is implied by the article's publication in *Communications of the AMS*, *Journal of the AMS*, *Mathematics of Computation*, *Memoirs*, *Proceedings of the AMS*, or *Transactions of the AMS*, and is therefore not a prize selection criterion.

Professor Isaacs is the author of several graduate-level textbooks and of about 200 research papers on finite groups and their characters, with special emphasis on groups—such as solvable groups—that have an abundance of normal subgroups. He is a Fellow of the American Mathematical Society, and received teaching awards from the University of Wisconsin and from the School of Engineering at the University of Wisconsin. He is especially proud of his 29 successful PhD students.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Submit a letter of nomination describing the candidate's accomplishments including complete bibliographic citations for the work being nominated, a CV for the nominee, and a brief citation that explains why the work is important.

Information on how to nominate can be found here: `https://www.ams.org/isaacs-prize`.

## NEW! Elias M. Stein Prize for Transformative Exposition


Elias M. Stein

The Elias M. Stein Prize for Transformative Exposition is awarded for a written work, such as a book, survey, or exposition, *in any area of mathematics* that transforms the mathematical community's understanding of the subject or reshapes the way it is taught.

### About this Prize

This prize was endowed in 2022 by students, colleagues, and friends of Elias M. Stein to honor his remarkable legacy of writing monographs and textbooks, both singly and with collaborators. Stein's research monographs, such as *Singular Integrals and Differentiability Properties of Functions* and *Harmonic Analysis*, became canonical references for generations of researchers, and textbooks such as the Stein and Shakarchi series *Princeton Lectures in Analysis* became instant classics in undergraduate and graduate classrooms. Stein is remembered for his ability to find a perspective to make a method of proof seem so natural as to be inevitable, and for his strategy of revealing the essential difficulties, and their solutions, in the simplest possible form before elaborating on more general settings. This prize seeks to recognize mathematicians at any career stage who, like Stein, have invested in writing a book or manuscript that transforms how their research community, or the next generation, understands the current state of knowledge in their area.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Submit a letter of nomination describing the candidate's accomplishments including complete bibliographic citations for the work being nominated, a CV for the nominee, and a brief citation that explains why the work is important.

Information on how to nominate can be found here: `https://www.ams.org/stein-exposition`.

# Ivo and Renata Babuška Thesis Prize

The Ivo and Renata Babuška Thesis Prize is awarded annually to the author of an outstanding PhD thesis in mathematics, interdisciplinary in nature, possibly with applications to other fields. The current prize amount is US$3,000.

## About this Prize

Ivo Babuška (1926–2023) was a Czech-American mathematician whose honors include five doctorates honoris causa, the Czechoslovak State Prize for Mathematics, the Leroy P. Steele Prize, the Birkhoff Prize, the Humboldt Award of Federal Republic of Germany, the John von Neumann Medal, the Neuron Prize Czech Republic, the ICAM Congress Medal (Newton Gauss), the Bolzano Medal, and the Honorary Medal De Scientia Et Humanitate Optime Meritis. Asteroid 36060 Babuška was named in his honor by the International Astronomical Union.

Renata Babuška (nee Mikulášek) was Ivo's wife and partner for 63 years. Renata grew up in Prague, Czechoslovakia and graduated from Charles University in 1953 with a degree in mathematical statistical engineering. Upon graduation, she was assigned to the Education Department as an administrator evaluating universities and technical schools. Two years later she became an assistant professor of mathematics at the Czech Technical University. After moving to the US, Renata worked as a data and computing management consultant for different government agencies in Washington, DC. She liked to point out that behind every successful man is a strong woman and he often said that without Renata, he would not have accomplished all that he did.

Babuška was a Distinguished Professor at the University of Maryland at College Park and then the Robert B. Trull Chair in Engineering, TICAM senior research scientist, professor of aerospace engineering and engineering mechanics, and professor of mathematics at the University of Texas, Austin. He was a Fellow of SIAM, ACM, and ICAM; a member of the US National Academy of Engineering, the Academy of Medicine, Engineering, and Sciences of Texas, and the European Academy of Sciences; and an honorary foreign member of the Czech Learned Society.

Babuška's work spanned the fields of theoretical and applied mathematics with emphasis on numerical methods, finite element methods, and computational mechanics. His interest in fostering collaboration among mathematicians, engineers, and physicists led him to establish this prize to encourage and recognize interdisciplinary work with practical applications.

The Ivo and Renata Babuška Thesis Prize is awarded in line with other AMS prizes and awards, according to governance rules and practice in effect at that time.

**Next Prize:** January 2025

**Nomination Period:** February 1–June 30

**Nomination Procedure:**
1. The prize will recognize a thesis for a PhD granted between July 1 of year -1 and June 30 of year 0 (the year of nomination and selection) and will be presented at the Joint Mathematics Meetings in January of year +1 wherever it appears.
2. The nominating institution will be a PhD-granting institution that is either (a) located in the United States of America (USA), or (b) located outside the USA and an institutional AMS member at the time of the nomination.
3. One PhD thesis may be nominated by a nominating institution.
4. The nominating institution will submit a copy of the thesis along with a letter in support of the nomination, and both will be written in English.
5. A selection committee will be appointed by the AMS president.

`https://www.ams.org/babuska`

# Mary P. Dolciani Prize for Excellence in Research

The AMS Mary P. Dolciani Prize for Excellence in Research recognizes a mathematician from a department that does not grant a PhD who has an active research program in mathematics and a distinguished record of scholarship. The primary criterion for the prize is an active research program as evidenced by a strong record of peer-reviewed publications.

Additional selection criteria may include the following:
- Evidence of a robust research program involving undergraduate students in mathematics;
- Demonstrated success in mentoring undergraduates whose work leads to peer-reviewed publication, poster presentations, or conference presentations;
- Membership in the AMS at the time of nomination and receipt of the award is preferred but not required.

## About this Prize

This prize is funded by a grant from the Mary P. Dolciani Halloran Foundation. Mary P. Dolciani Halloran (1923–1985) was a gifted mathematician, educator, and author. She devoted her life to developing excellence in mathematics education and was a leading author in the field of mathematical textbooks at the college and secondary school levels.

The prize amount is $5000, awarded every other year for five award cycles.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Nominations should include a letter of nomination, the nominee's CV, and a short citation to be used in the event that the nomination is successful.

Information on how to nominate can be found here: `https://www.ams.org/dolciani-prize`.

# Award for an Exemplary Program or Achievement in a Mathematics Department

This award recognizes a department which has distinguished itself by undertaking an unusual or particularly effective program of value to the mathematics community, internally or in relation to the rest of society. Examples might include a department that runs a notable minority outreach program, a department that has instituted an unusually effective industrial mathematics internship program, a department that has promoted mathematics so successfully that a large fraction of its university's undergraduate population majors in mathematics, or a department that has made some form of innovation in its research support to faculty and/or graduate students, or which has created a special and innovative environment for some aspect of mathematics research.

## About this Award

This award was established in 2004. For the first three awards (2006–2008), the prize amount was US$1,200. The prize was endowed by an anonymous donor in 2008, and starting with the 2009 prize, the amount is US$5,000. This US$5,000 prize is awarded annually. Departments of mathematical sciences in North America that offer at least a bachelor's degree in mathematical sciences are eligible.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** A letter of nomination may be submitted by one or more individuals. Nomination of the writer's own institution is permitted. The letter should describe the specific program(s) for which the department in being nominated as well as the achievements which make the program(s) an outstanding success, and may include any ancillary documents which support the success of the program(s). Where possible, the letter and documentation should address how these successes came about by 1) systematic, reproducible changes in programs that might be implemented by others, and/or 2) have value outside the mathematical community. The letter should not exceed two pages, with supporting documentation not to exceed an additional three pages.

Information on how to nominate can be found here: `https://www.ams.org/department-award`.

# Award for Impact on the Teaching and Learning of Mathematics

This award is given annually to a mathematician (or group of mathematicians) who has made significant contributions of lasting value to mathematics education.

Priorities of the award include recognition of:

(a) accomplished mathematicians who have worked directly with precollege teachers to enhance teachers' impact on mathematics achievement for all students, or

(b) sustainable and replicable contributions by mathematicians to improving the mathematics education of students in the first two years of college.

## About this Award

The Award for Impact on the Teaching and Learning of Mathematics was established by the AMS Committee on Education in 2013. The endowment fund that supports the award was established in 2012 by a contribution from Kenneth I. and Mary Lou Gross in honor of their daughters Laura and Karen.

The US$1,000 award is given annually.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Letters of nomination may be submitted by one or more individuals. The letter of nomination should describe the significant contributions made by the nominee(s) and provide evidence of the impact these contributions have made on the teaching and learning of mathematics. The letter of nomination should not exceed two pages, and may include supporting documentation not to exceed three additional pages. A brief curriculum vitae for each nominee should also be included. The nonwinning nominations will automatically be reconsidered, without further updating, for the awards to be presented over the next two years.

Information on how to nominate can be found here: `https://www.ams.org/impact`.

# Ciprian Foias Prize in Operator Theory

The Ciprian Foias Prize in Operator Theory is awarded for notable work in Operator Theory published during the preceding six years. The work must be published in a recognized, peer-reviewed venue.

### About this Prize

This prize was established in 2020 in memory of Ciprian Foias (1933–2020) by colleagues and friends. He was an influential scholar in operator theory and fluid mechanics, a generous mentor, and an enthusiastic advocate of the mathematical community.

The current prize amount is US$5,000, and the prize is awarded every three years.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Nominations require CV of the nominee, a letter of nomination, and a citation.

Information on how to nominate can be found here: `https://www.ams.org/foias-prize`.

# David P. Robbins Prize

The Robbins Prize is for a paper with the following characteristics: it shall report on novel research in algebra, combinatorics, or discrete mathematics and shall have a significant experimental component; and it shall be on a topic which is broadly accessible and shall provide a simple statement of the problem and clear exposition of the work. Papers published within the six calendar years preceding the year in which the prize is awarded are eligible for consideration.

### About this Prize

This prize was established in 2005 in memory of David P. Robbins by members of his family. Robbins, who died in 2003, received his PhD in 1970 from MIT. He was a longtime member of the Institute for Defense Analysis Center for Communications Research and a prolific mathematician whose work (much of it classified) was in discrete mathematics.

The current prize amount is US$5,000 and the prize is awarded every 3 years.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Submit a letter of nomination, a complete bibliographic citation for the work being nominated, and a brief citation that explains why the work is important.

Information on how to nominate can be found here: `https://www.ams.org/robbins-prize`.

# E. H. Moore Research Article Prize

The Moore Prize is awarded for an outstanding research article to have appeared in one of the AMS primary research journals (namely, the *Journal of the AMS*, *Proceedings of the AMS*, *Transactions of the AMS*, *Memoirs of the AMS*, *Mathematics of Computation*, *Electronic Journal of Conformal Geometry and Dynamics*, and *Electronic Journal of Representation Theory*) during the six calendar years ending a full year before the meeting at which the prize is awarded.

### About this Prize

The prize was established in 2002 in honor of E. H. Moore. Among other activities, Moore founded the Chicago branch of the American Mathematical Society, served as the Society's sixth president (1901–1902), delivered the Colloquium Lectures in 1906, and founded and nurtured the *Transactions of the AMS*.

The current prize amount is US$5,000, awarded every three years.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Submit a letter of nomination, a complete bibliographic citation for the work being nominated, and a brief citation that explains why the work is important.

Information on how to nominate can be found here: `https://www.ams.org/moore-prize`.

# Levi L. Conant Prize

This prize was established in 2000 in honor of Levi L. Conant to recognize the best expository paper published in either the *Notices of the AMS* or the *Bulletin of the AMS* in the preceding five years.

### About this Prize

Levi L. Conant was a mathematician and educator who spent most of his career as a faculty member at Worcester Polytechnic Institute. He was head of the mathematics department from 1908 until his death and served as interim president of WPI from 1911 to 1913. Conant was noted as an outstanding teacher and an active scholar. He published a number of articles in scientific journals and wrote four textbooks. His will provided for funds to be donated to the AMS upon the death of his wife.

Prize winners are invited to present a public lecture at Worcester Polytechnic Institute as part of their Levi L. Conant Lecture Series, which was established in 2006.

The Conant Prize is awarded annually in the amount of US$1,000.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Nominations with supporting information should be submitted online. Nominations should include a letter of nomination, a short description of the work that is the basis of the nomination a complete bibliographic citation for the article being nominated.

Information on how to nominate can be found here: `https://www.ams.org/conant-prize`.

# Mathematics Programs that Make a Difference

This Award for Mathematics Programs that Make a Difference was established in 2005 by the AMS's Committee on the Profession to compile and publish a series of profiles of programs that:

1. aim to bring more persons from underrepresented backgrounds into some portion of the pipeline beginning at the undergraduate level and leading to advanced degrees in mathematics and professional success, or retain them once in the pipeline;
2. have achieved documentable success in doing so; and
3. are potentially replicable models.

## About this Award

This award brings recognition to outstanding programs that have successfully addressed the issues of underrepresented groups in mathematics. Examples of such groups include racial and ethnic minorities, women, low-income students, and first-generation college students.

One program is selected each year by a selection committee appointed by the AMS president and is awarded US$1,000 provided by the Mark Green and Kathryn Kert Green Fund for Inclusion and Diversity.

Preference is given to programs with significant participation by underrepresented minorities. Note that programs aimed at pre-college students are eligible only if there is a significant component of the program benefiting individuals from underrepresented groups at or beyond the undergraduate level. Nomination of one's own institution or program is permitted and encouraged.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** The letter of nomination should describe the specific program being nominated and the achievements that make the program an outstanding success. It should include clear and current evidence of that success. A strong nomination typically includes a description of the program's activities and goals, a brief history of the program, evidence of its effectiveness, and statements from participants about its impact. The letter of nomination should not exceed two pages, with supporting documentation not to exceed three more pages. Up to three supporting letters may be included in addition to these five pages. Nomination of the writer's own institution or program is permitted. Nonwinning nominations will automatically be reconsidered for the award for the next two years.

Information on how to nominate can be found here: `https://www.ams.org/make-a-diff-award`.

# Oswald Veblen Prize in Geometry

The award is made for a notable research work in geometry or topology that has appeared in the last six years. The work must be published in a recognized, peer-reviewed venue.

## About this Prize

This prize was established in 1961 in memory of Professor Oswald Veblen through a fund contributed by former students and colleagues. The fund was later doubled by the widow of Professor Veblen. An anonymous donor generously augmented the fund in 2008. In 2013, in honor of her late father, John L. Synge, who knew and admired Oswald Veblen, Cathleen Synge Morawetz and her husband, Herbert, substantially increased the endowment.

The current prize amount of US$5,000 is awarded every three years.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Submit a letter of nomination, a complete bibliographic citation for the work being nominated, and a brief citation that explains why the work is important.

Information on how to nominate can be found here: `https://www.ams.org/Veblen-prize`.

# Ruth Lyttle Satter Prize in Mathematics

The Satter Prize recognizes an outstanding contribution to mathematics research by a woman in the previous six years.

### About this Prize

This prize was established in 1990 using funds donated by Joan S. Birman in memory of her sister, Ruth Lyttle Satter. Professor Birman requested that the prize be established to honor her sister's commitment to research and to encourage women in science. An anonymous benefactor added to the endowment in 2008.

The current prize amount is $5,000 and the prize is awarded every 2 years.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Submit a letter of nomination describing the candidate's accomplishments including complete bibliographic citations for the work being nominated, a CV for the nominee, and a brief citation that explains why the work is important.

Information on how to nominate can be found here: `https://www.ams.org/satter-prize`.

# Joint Prizes and Awards

## Frank and Brennie Morgan Prize for Outstanding Research in Mathematics by an Undergraduate Student

## (AMS-MAA-SIAM)

The Morgan Prize is awarded each year to an undergraduate student (or students for joint work) for outstanding research in mathematics. Any student who was enrolled as an undergraduate in December at a college or university in the United States or its possessions, Canada, or Mexico is eligible for the prize.

The prize recipient's research need not be confined to a single paper; it may be contained in several papers. However, the paper (or papers) to be considered for the prize must be completed while the student is an undergraduate. Publication of research is not required.

### About this Prize

The prize was established in 1995. It is entirely endowed by a gift from Mrs. Frank (Brennie) Morgan. It is made jointly by the American Mathematical Society, the Mathematical Association of America, and the Society for Industrial and Applied Mathematics.

The current prize amount is $1,200, awarded annually.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** To nominate a student, submit a letter of nomination, a brief description of the work that is the basis of the nomination, and complete bibliographic citations (or copies of unpublished work). All submissions for the prize must include at least one letter of support from a person, usually a faculty member, familiar with the student's research.

Information on how to nominate can be found here: `https://www.ams.org/morgan-prize`.

## JPBM Communications Award

This award is given each year to reward and encourage communicators who, on a sustained basis, bring mathematical ideas and information to non-mathematical audiences.

### About this Award

This award was established by the Joint Policy Board for Mathematics (JPBM) in 1988. JPBM is a collaborative effort of the American Mathematical Society, the Mathematical Association of America, the Society for Industrial and Applied Mathematics, and the American Statistical Association.

Up to two awards of US$2,000 are made annually. Both mathematicians and non-mathematicians are eligible.

**Next Prize:** January 2025

**Nomination Period:** open

**Nomination Procedure:** Nominations should be submitted on mathprograms.org. Note: Nominations collected before September 15th in year N will be considered for an award in year N+2.

Information on how to nominate can be found here: `https://www.ams.org/jpbm-comm-award`.

# AMS-SIAM Norbert Wiener Prize in Applied Mathematics

The Wiener Prize is awarded for an outstanding contribution to "applied mathematics in the highest and broadest sense."

## About this Prize

This prize was established in 1967 in honor of Professor Norbert Wiener and was endowed by a fund from the Department of Mathematics of the Massachusetts Institute of Technology. The endowment was further supplemented by a generous donor.

Since 2004, the US$5,000 prize has been awarded every three years. The American Mathematical Society and the Society for Industrial and Applied Mathematics award this prize jointly; the recipient must be a member of one of these societies.

**Next Prize:** January 2025

**Nomination Period:** February 1–May 31

**Nomination Procedure:** Submit a letter of nomination describing the candidate's accomplishments including complete bibliographic citations for the work being nominated, a CV for the nominee, and a brief citation that explains why the work is important.

Information on how to nominate can be found here: `https://www.ams.org/wiener-prize`.

## Credits

Photo of I. Martin Isaacs is courtesy of Yvonne Nagel.

Photo of Elias M. Stein is courtesy of William Crow/Princeton University.

# Sharing Research, Making Connections: A Tale of Two Mathematicians at JMM

## *Elaine Beebe*

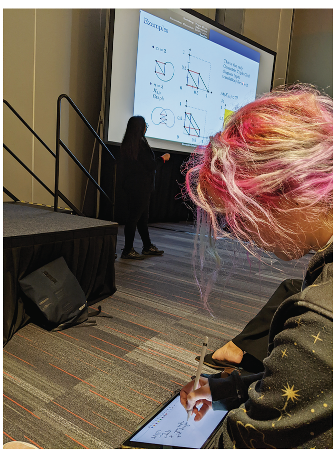


**Figure 1.** Maggie Miller takes notes while Devashi Gulati presents her talk at JMM 2024.

**Maggie Miller had just one** day to spend at JMM 2024 in San Francisco.

A first-year assistant professor at the University of Texas at Austin (UT) and a Clay Research Fellow, Miller had another math conference to go to that week. But having attended JMM for more than a decade meant she knew her way around.

On the opening morning of JMM, Miller would present her research. That afternoon, as people flooded to the awards ceremony elsewhere at JMM, Miller perched on a chair in the front row at "Searching for Triple Grid Diagrams" by Devashi Gulati.

"It's sort of about two areas of math, one of which I know a lot about and one of which I know less about, so I like that it might make the second more accessible to me," she said.

"But to be honest, part of the reason I'm looking forward to the talk is that the speaker is a PhD student, so I'm happy to see her doing something interesting. I feel like it would be more normal to say one of the plenary talks is interesting, but I honestly like the regular talks better."

Miller had seen a few shorter versions of Gulati's talk at other conferences and was intrigued. "I was surprised by a theorem she had proved with her coauthors—they proved that triple grid diagrams exist for any Lagrangian surface in CP2, which I believe she has stated as a conjecture in previous talks," Miller said. "I had asked one of her coauthors, Sarah Blackwell, about it a long time ago and had the impression that a general existence theorem wasn't likely on the horizon, so I thought that was quite good."

**Figure 2.** Devashi Gulati.

**Devashi Gulati is a PhD** candidate in mathematics at the University of Georgia in Athens. She attended BITS Pilani in Goa, India, for her master's degree in mathematics and bachelor of engineering in computer science, earning both with honors.

JMM 2024 was Gulati's first mathematical speaking engagement longer than a lightning talk. She had never attended JMM before.

Gulati explained her talk this way: "Sarah Blackwell, David Gay, and Peter Lambert-Cole defined what triple grid diagrams are... a one-dimensional higher analogue of grid diagrams, using trisections. Grid diagrams are useful since they are a combinatorial representation of smooth objects, which computers can work with. But triple grid diagrams are hard to find.

"My advisor [Lambert-Cole] and I tried to computationally search all grid diagrams, but that turned out to be a resource-intensive endeavor. We also found many, many

*Elaine Beebe is the communications and outreach content specialist at the American Mathematical Society. Her email address is* `exb@ams.org`.



DOI: https://doi.org/10.1090/noti2927

examples which were combinatorially different but the underlying surface was mathematically equivalent. So, we decided to take advantage of the context of these diagrams to try to determine when they are equivalent.

"That led to us looking at the structure of the collection of all triple grid diagrams, called the moduli space. Doing so, we realized that if we fix the underlying graph, we get a space of triple grid diagrams associated with it which represent the same underlying surface. Furthermore, we were also able to define 'moves' on triple grid diagrams."

Gulati's talk was part of the AMS-AWM Special Session for Women and Gender Minorities in Symplectic and Contact Geometry and Topology, held every other year for the past eight years. Each time the session runs at JMM, the previous occurrence's organizers find the leaders for the next installment.

"When choosing speakers for the session, the main goal is to invite people within this research network, with a particular focus on graduate students and early career mathematicians who are about to graduate and/or currently on the job market," said the aforementioned Sarah Blackwell, an NSF Postdoctoral Fellow at the University of Virginia, who organized the 2024 session with Luya Wang of Stanford University and Nicole Magill of Cornell University.

This year, "we tried to invite speakers working on a wide range of topics within symplectic geometry," Blackwell said. "As it turns out, I actually do happen to know more about Devashi's topic, because I know her personally, and part of my dissertation was closely related to what she is doing."

**Figure 3.** Maggie Miller.

**She is just 30—even** named one of *Forbes* magazine's "30 Under 30"—but Maggie Miller, winner of the Maryam Mirzakhani New Frontiers Prize, is an old hand at the JMM.

Sponsored by AMS travel grants, Miller attended the JMM as a UT undergraduate in 2013 (San Diego), 2014 (Baltimore), and 2015 (San Antonio). She participated in JMM 2020 in Denver while a graduate student at Princeton University.

For 2022's virtual JMM, as a Visiting Clay Fellow at Stanford University, Miller organized the AMS Special Session on Knot Theory in Dimension Four with Jeffrey Meier (Western Washington University) and Patrick Naylor (Princeton). She also presented research undertaken with Mark Hughes (Brigham Young University) and Seungwon Kim (Sungkyunkwan University), and had an additional research project presented.

Flash back to her first JMM 11 years ago, where Miller earned a joint poster prize and gave her first math talk ever, "Community Detection by Maximizing Partition Efficiency." The project, with Brendan Shah, an undergrad at Rochester Institute of Technology (RIT), combined math and computer science. "Our program was much slower than existing algorithms, so not at all practical—but it was an interesting problem mathematically," she said.

"I made slides and then gave several practice talks in front of just Brendan," she continued. "For me the hardest part was not speaking very quickly.

"I was really nervous the first time I spoke at JMM—even though the talk was only ten minutes long, it was my first time presenting in front of strangers. Of course, now that's something I have to do all the time. Getting to practice early on gave me a start on developing public speaking abilities, so that by the time I was giving hour-long talks I was a little more confident."

Miller said of her own talk earlier that day, "I always try to keep JMM or AMS meeting talks a bit general, but I noticed at the start of the talk a lot of students in the back who I thought were likely undergraduate students. So, I explained some of the background more carefully than usual, which was kind of interesting to adapt to with slides, but I think went okay."

**"A grid is a computer-friendly** way to code a knot . . . "

For Gulati, everything went smoothly as a presenter: no slide mishaps and an engaged audience of perhaps two dozen.

At the end of Gulati's talk, Miller raised her hand.

She asked Gulati what was technically difficult about producing a set of moves relating diagrams of Lagrangian-isotopic surfaces, or even just isotopic surfaces.

"The triple grid diagrams come from another system of diagrams (shadow diagrams) that do have a complete set of moves, so for them to not have proved the same thing for triple grid diagrams means there is probably a technical obstruction," Miller said. "Devashi implied that the conditions for a triple grid diagram to be valid are very rigid, and the natural move you might guess to do isn't actually an allowable move on these diagrams."

After her talk, Gulati said, "I am happy I was able to share the math I have been thinking about recently," confessing to having been a bit nervous, though it didn't show in her presentation.

"Attending the JMM was a great experience," Gulati added. "Usually at conferences, I end up meeting the people who work in only my field, in the USA. At the JMM, I was able to meet not only international mathematicians working in my field but also other mathematicians working in related fields. I was also able to reconnect with some other graduate students who I have met over the years at conferences, workshops.

"As doing math research can be a solitary endeavor, conferences are vital to not only refresh one's perspective but also connect with others and feel a sense of belonging in the math community. I was able to talk with others not only about math, but also their personal journey doing math."

**Since midway through graduate school**, Miller has traveled "a lot" for seminars, conferences, and research collaboration. "In-person meetings have been really influential to my career," she said. "Most of my projects are joint research, and I can usually point at a specific trip that was the impetus of any given project. I don't really enjoy brainstorming ideas over email or Zoom, and everything is much more efficient and interesting in person, for me at least."

Among her conferences, JMM is unique, Miller said. "I enjoy how broad these meetings are, not just in terms of mathematical variety, but the teaching- and career-focused activities, panels on social topics or advice, etc. It's completely different from any other conference I participate in during the year."

Miller said that she saves various application deadlines to her calendar "just in case I had a student who would want to go, and will keep doing that in the future. . . As a faculty member, I would want to send future undergraduate or graduate students to JMM to give talks when possible, just like I did."

### Credits

Figure 1 is courtesy of AMS Communications.
Figure 2 is courtesy of Devashi Gulati.
Figure 3 is courtesy of Maggie Miller.

AMS COMMUNICATION

# The Serious Whimsy of Recreational Mathematics at JMM 2024

## *Elaine Beebe*

**They swarm the Rubik's Cube** booth, opening night at JMM 2024.

All hands are on deck, swirling and swooping and spinning cubes in an organized frenzy.

Once solved, each $3 \times 3 \times 3$ cube is stacked purposefully in a large, square wooden frame, creating an elaborate mosaic of Ernő Rubik's namesake in honor of its fiftieth birthday.

More about that mosaic design later from event organizer Tomas Rokicki, who is too busy to talk now, running around, corralling "cubers."

Scanning the buzzing scene is co-organizer Robert Hearn of H3 Labs and Gathering 4 Gardner, an educational nonprofit devoted to recreational math named for writer and polymath Martin Gardner.

Hearn observes the speedy solve rates of the 384 cubes with mock consternation. "I thought it would take longer, at least an hour or two," he said. "Constructing that frame took me two days."

In a mere 25 minutes, the mosaic is completed.

**Seemingly whimsical topics such as** toys, games, and puzzles can lead to significant discoveries. Recreational mathematics has inspired findings in probability and graph theory, and the aperiodic monotile of 2023 known as "the Hat."

This thinking underscored JMM 2024's Serious Recreational Mathematics special sessions, organized by Erik Demaine of the Massachusetts Institute of Technology

*Elaine Beebe is the communications and outreach content specialist at the American Mathematical Society. Her email address is* `exb@ams.org`.



DOI: https://doi.org/10.1090/noti2943

**Figure 1.** At JMM 2024, cubers solve twisty puzzles.

(MIT), along with Rokicki and Hearn. Demaine, who met Rokicki at a "Gathering 4 Gardner," was co-advisor to Hearn's PhD from MIT.

"I got interested in recreational mathematics through my father [MIT mathematician Martin Demaine], who got interested in it from Martin Gardner's 'Mathematical Games' column in *Scientific American*," Demaine said. "In the late 1990s when I was a new PhD student, we started working in the field of computational origami, which was just getting started, and its motivation was essentially recreational: How to design cooler origami art. Later it turned into a practical field with many applications to science and engineering."

"Ever since, I've been motivated by mathematical problems that I find fun—be it origami or video games or puzzles."

To celebrate the Rubik Cube's fiftieth anniversary, Demaine's magician friend Mark Mitton suggested organizing a panoply of events in 2024. "A JMM special session seemed like a great venue to get more mathematicians

**Figure 2.** The Rubik's Cube mosaic at JMM 2024 was composed of 384 $3 \times 3 \times 3$ cubes.

excited about recreational mathematics," said Demaine. At JMM, he presented a talk called "Puzzles and Games Meet Algorithms and Complexity," while Hearn discussed "The Puzzling Origins of Compound Symmetry Groups" and Rokicki presented "Twenty Moves Suffice for Rubik's Cube."

Let's get back to that completed mosaic: a colorful cube with what appears to be an aesthetically appealing, random pattern.

Not so fast, according to Hearn.

The pattern created by the 384-cube mosaic is the so-called "superflip," where every one of the 12 edge pieces on the cube is flipped. "Tom (Rokicki) chose this pattern, I suspect, because it is one of a very small number of patterns that requires a full 20 moves to solve, as he was the first to prove," Hearn says.

In 1999, Rokicki made a bold claim that he would determine the maximum number of moves required to solve Rubik's Cube, referred to as "God's number." This required analyzing solutions for all 43,252,003,274,489,856,000 possible positions of the cube. Nearly 10 years later, "with help from some very bright people and a donation of computer time from Google," Rokicki ultimately determined God's number to be 20.

Rokicki corrects Hearn, giving credit where due. "Michael Reid was the first to show that superflip required 20 moves and it was the first position that was shown to require 20 moves. My result was that this is a maximum, that every position can be solved in 20 moves or less."

**Serious Recreational Mathematics special-session** topics ran the gamut. Donald Knuth of Stanford University spoke of "Recreational Computer Programming" while Susan Goldstine of St. Mary's College of Maryland discussed "Counting Stitches: Enumerative Problems in Knitting." Martin Demaine presented "Fun with Fonts: Algorithmic Typography." Ryuhei Uehara of the Japan Advanced Institute of Science and Technology presented his research on common shape puzzles, while Klara Mundilova, a re-


**Figure 3.** Steve Butler of Iowa State University explains the mathematics of juggling.

cent PhD graduate from MIT advised by Erik Demaine, discussed "Art-Inspired Curved-Crease Origami Analysis and Design."

"We were fortunate to be able to recruit a very high-quality selection of speakers, and I'd call the session an unqualified success," said Hearn, citing "much larger attendance than we could have imagined. It seems possible this could become a regular JMM session going forward, as the Mathematics and the Arts sessions have been for a while."

Mundilova added, "Hearing how accomplished researchers playfully apply various disciplines of mathematics to recreational areas, such as puzzles, games, art, origami, juggling, tilings, and knitting, was informative, entertaining, and very inspiring. I would definitely attend a Serious Recreational Mathematics session again."

In a session called "The Mathematics of Discrete Periodic Patterns . . . or How I Learned to Stop Worrying and Love the Throw," Steve Butler of Iowa State University used colorful round props.

"Juggling is about patterns and we can apply math to juggling by exploring these patterns," Butler said. "This can mean discovering new patterns, connecting patterns together, and finding the limits of what is possible."

In their talks, the presenters used humor to good effect. As Butler told the crowd, "Every math talk should have one proof and one joke. They should not be the same thing."

**For the three session organizers,** a pinnacle of Serious Recreational Mathematics was the Saturday morning interview with Ernő Rubik himself: via Zoom, with the Hungarian inventor's image splayed huge across screens like the Wizard of Oz.

"It was great talking with Ernő!" Rokicki said. "I thought he had some great answers and stories!"

**Figure 4.** Ernő Rubik, via Zoom, at JMM 2024.

"Really, Rubik shaped a lot of the direction of my life. To be able to ask him questions for an hour was really something special," noted Hearn, who did his PhD on games and puzzles. Erik Demaine added that they had hoped to reveal something of Rubik the person behind the cube.

The three interviewers covered a lot of ground with their questions. "Do aliens have Rubik's Cubes? What is the meaning of life? Why did the $3 \times 3 \times 3$ appear before the $2 \times 2 \times 2$? What was it like living behind the Iron Curtain as the cube spread throughout the world? and much more," Hearn said.

Mundilova attended the 8 a.m. interview session. "It was very interesting to hear about the thought processes and the historical contexts of the development and design of the Rubik's Cube," she said.

"I find it fascinating how this puzzle has captivated people across generations. My parents enjoyed solving the Rubik's Cube decades ago, despite it initially being a rare and sought-after item in Czechoslovakia. I find it remarkable that the same concept, though now in various forms, continues to engage and challenge people today," she said.

"The Rubik's Cube has been a wonderful way to get the general population to experience math without telling them that is what they are doing," Butler said.

"Playing with the Rubik's Cube makes people think about algorithms, symmetries, and group actions," he said. "As a result, the Rubik's Cube has been a wonderful tool for the math community."

But Butler chose a different highlight for his JMM experience this year.

"The best part was getting communities back together again. It felt like we were back to normal."

### Credits

Figures 1–4 are courtesy of Kyle Hurley, AMS Communications.

# AMS Updates

## Speakers Named for NZMS-AustMS-AMS Joint Meeting

Plenary speakers have been announced for the 2024 joint meeting of the New Zealand Mathematical Society (NZMS), Australian Mathematical Society (AustMS), and the AMS, to be held in Auckland, New Zealand, December 9–13.

The named lectures will be presented as follows:

- Rachael Ka'ai-Mahute (Auckland University of Technology) and Michael Miller (Victoria University Wellington) representing the Te ara o te reo Maori language revitalization project: Dr. Yunupingu Lecture
- James Saunderson (Monash University): Butcher-Kalman Lecture
- Svitlana Mayboroda (University of Minnesota and ETH Zurich): Hana-Neumann Lecture
- Persi Diaconis (Stanford University): Public Lecture

General Plenary Lectures will be delivered by Katherine Turner (Australian National University), Lara Alcock (Loughborough University), Geordie Williamson (University of Sydney), Richard Kenyon (Yale University), Priya Subramanian (University of Auckland), and Eamonn O'Brien (University of Auckland).

"Mathematics is a global enterprise, and, as such, joint international meetings are an important part of the AMS's activities," said Steve Weintraub, AMS associate secretary, Eastern section. "Our goal in this meeting is to foster cooperation between mathematicians in the US, New Zealand, and Australia."

*—AMS Communications*

DOI: https://doi.org/10.1090/noti2944

## Graduate Students: Apply Now for Fall Travel Support

On May 15, applications for AMS 2024 Fall Sectional Travel Grants will open on MathPrograms.org. Apply before July 24, 2024, for $250 in support to attend one of the four fall sectional meetings:

- September 14–15, 2024: University of Texas at San Antonio, San Antonio, TX
- October 5–6, 2024: Georgia Southern University, Savannah, GA
- October 19–20, 2024: University at Albany, State University of New York, Albany, NY
- October 26–27, 2024: University of California, Riverside, Riverside, CA

Held on weekends, these regional meetings provide valuable connection and cross-pollination with peers, potential mentors, and future colleagues at an intimate and welcoming scale.

*—AMS Programs Department*

## Deaths of AMS Members

John Brillhart, of Tucson, Arizona, died on May 21, 2022. Born on November 13, 1930, he was a member of the Society for 58 years.

Henry W. Haslach Jr., of Greenbelt, Maryland, died on January 15, 2023. Born on May 1, 1942, he was a member of the Society for 56 years.

Denis Michael Falvey, of Canada, died on September 14, 2023. Born on October 1, 1948, he was a member of the Society for 18 years.

Alessandro Figa-Talamanca, of Italy, died on November 27, 2023. Born on May 25, 1938, he was a member of the Society for 59 years.

R. B. Burckel, of Manhattan, Kansas, died on December 10, 2023. Born on December 15, 1939, he was a member of the Society for 59 years.

# Mathematics People

## Clay Research Fellows 2024 Announced

The Clay Mathematics Institute is pleased to announce that Ishan Levy and Mehtaab Sawhney have been awarded Clay Research Fellowships. Each has been appointed as a Clay Research Fellow for five years beginning July 1, 2024. Clay Research Fellowships are awarded on the basis of the exceptional quality of candidates' research and their promise to become mathematical leaders.

Levy will receive his PhD from the Massachusetts Institute of Technology in 2024 under the supervision of Michael Hopkins. "Levy is known for his deep and ingenious contributions to homotopy theory. His new techniques in algebraic K-theory have led to solutions of many old problems. . . . He is most renowned for his work on Ravenel's 'Telescope Conjecture.'"

Sawhney will receive his PhD from the Massachusetts Institute of Technology in 2024 under the supervision of Yufei Zhao. "While still a graduate student, Sawhney has achieved a stunning number of breakthroughs on fundamental problems across extremal combinatorics, probability theory, and theoretical computer science. . . . His remarkable body of work has already transformed swathes of combinatorics."

*—Clay Mathematics Institute*

## Brydges Wins Heineman Prize for Mathematical Physics

The American Institute of Physics (AIP) and the American Physical Society (APS) are pleased to announce David Brydges as the recipient of the 2024 Dannie Heineman Prize for Mathematical Physics, "for achievements in the fields of constructive quantum field theory and rigorous statistical mechanics, especially the introduction of new techniques including random walk representation in spin systems, the lace expansion, and mathematically rigorous implementations of the renormalization group."

Born in Chester, UK, Brydges recalls exploring the world around him from an early age, mesmerized by technology and the natural world. He earned his doctorate in mathematics from the University of Michigan. He is currently a professor emeritus of mathematics at the University of British Columbia.

Named after Dannie N. Heineman, an engineer, business executive, and philanthropic sponsor of the sciences, the prize was established in 1959 by the Heineman Foundation for Research, Education, Charitable and Scientific Purposes, Inc. This annual award of $10,000 recognizes significant contributions within the field of mathematical physics and was to have been presented at the APS March Meeting in Minneapolis.

*—American Institute of Physics*

## *Journal of Complexity* Lauds 2023 Best Paper

The 2023 Best Paper Award of the *Journal of Complexity* has been awarded to "Sampling numbers of smoothness classes via $\ell^1$-minimization," by Thomas Jahn, Tino Ullrich, and Felix Voigtlaender, published in Volume 79, December 23, Article 101786.

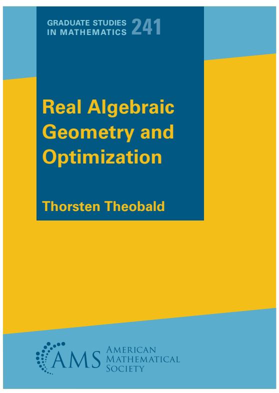
The $4,000 prize will be divided between the winners. Each author will also receive a plaque to be presented at the Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing Conference (MCQMC), August 18–23, 2024, Waterloo, Canada.

—Journal of Complexity

DOI: https://doi.org/10.1090/noti2945

# New Books Offered by the AMS

## Algebra and Algebraic Geometry




### Real Algebraic Geometry and Optimization

**Thorsten Theobald**, *Goethe University Frankfurt, Frankfurt am Main, Germany*

This book provides a comprehensive and user-friendly exploration of the tremendous recent developments that reveal the connections between real algebraic geometry and optimization, two subjects that were usually taught separately until the beginning of the 21st century. Real algebraic geometry studies the solutions of polynomial equations and polynomial inequalities over the real numbers. Real algebraic problems arise in many applications, including science and engineering, computer vision, robotics, and game theory. Optimization is concerned with minimizing or maximizing a given objective function over a feasible set. Presenting key ideas from classical and modern concepts in real algebraic geometry, this book develops related convex optimization techniques for polynomial optimization. The connection to optimization invites a computational view on real algebraic geometry and opens doors to applications.

Intended as an introduction for students of mathematics or related fields at an advanced undergraduate or graduate level, this book serves as a valuable resource for researchers and practitioners. Each chapter is complemented by a collection of beneficial exercises, notes on references, and further reading. As a prerequisite, only some undergraduate algebra is required.

*This item will also be of interest to those working in applications and geometry and topology.*

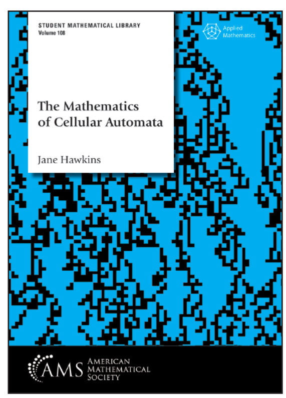
**Graduate Studies in Mathematics**, Volume 241
June 2024, approximately 271 pages, Hardcover, ISBN: 978-1-4704-7431-7, LC 2023052870, 2020 *Mathematics Subject Classification*: 14Pxx, 90Cxx, 68W30, 12D10, 52Axx, List US$135, **AMS members US$108**, **MAA members US$121.50**, Order code GSM/241

`bookstore.ams.org/gsm-241`

**Graduate Studies in Mathematics**, Volume 241
June 2024, Softcover, ISBN: 978-1-4704-7636-6, LC 2023052870, 2020 *Mathematics Subject Classification*: 14Pxx, 90Cxx, 68W30, 12D10, 52Axx, List US$89, **AMS members US$71.20**, **MAA members US$80.10**, Order code GSM/241.S

`bookstore.ams.org/gsm-241-s`

## Analysis




### The Mathematics of Cellular Automata

**Jane Hawkins**, *University of North Carolina at Chapel Hill, North Carolina*

This textbook offers a rigorous mathematical introduction to cellular automata (CA). Numerous colorful graphics illustrate the many intriguing phenomena, inviting undergraduates to step into the rich field of symbolic dynamics.

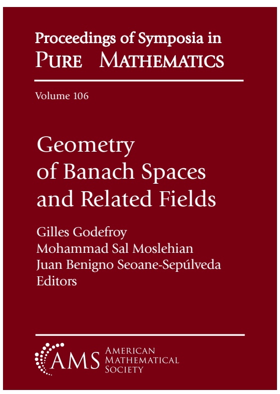
Beginning with a brief history, the first half of the book establishes the mathematical foundations of cellular automata. After recapping the essentials from advanced calculus, the chapters that follow introduce symbolic spaces, equicontinuity, and attractors. More advanced topics include the Garden of Eden theorem and Conway's Game of Life, and a chapter on stochastic CA showcases a model of virus spread. Exercises and labs end each chapter, covering a range of applications, both mathematical and physical.

Designed for undergraduates studying mathematics and related areas, the text provides ample opportunities for end-of-semester projects or further study. Computer use for the labs is largely optional, providing flexibility for different preferences and resources. Knowledge of advanced calculus and linear algebra is essential, while a course in real analysis would be ideal.

*This item will also be of interest to those working in geometry and topology.*

**Student Mathematical Library**, Volume 108
May 2024, 228 pages, Softcover, ISBN: 978-1-4704-7537-6, LC 2023053399, 2020 *Mathematics Subject Classification*: 37B15, 37B10, 68Q80, 37Exx, List US$59, **AMS Institutional member US$47.20**, **MAA members US$47.20**, All Individuals US$47.20, Order code STML/108

`bookstore.ams.org/stml-108`




## Geometry of Banach Spaces and Related Fields

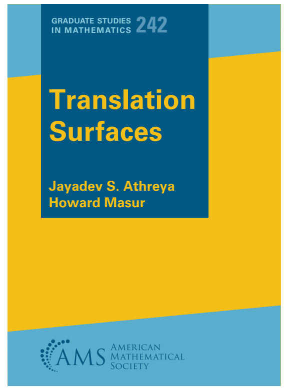
**Gilles Godefroy**, *Institute de Mathématiques de Jussieu, Paris, France*, **Mohammad Sal Moslehian**, *Ferdowsi University of Mashhad, Iran*, and **Juan Benigno Seoane-Sepúlveda**, *Universidad Complutense de Madrid, Spain*, Editors

This book provides a comprehensive presentation of recent approaches to and results about properties of various classes of functional spaces, such as Banach spaces, uniformly convex spaces, function spaces, and Banach algebras. Each of the 12 articles in this book gives a broad overview of current subjects and presents open problems. Each article includes an extensive bibliography.

This book is dedicated to Professor Per. H. Enflo, who made significant contributions to functional analysis and operator theory.

**Proceedings of Symposia in Pure Mathematics**, Volume 106
May 2024, 346 pages, Softcover, ISBN: 978-1-4704-7570-3, LC 2023053394, 2020 *Mathematics Subject Classification*: 43A22, 46B04, 46B15, 46B20, 46E15, 46E30, 46E40, 46H25, 46L52, 47B49, List US$139, **AMS members US$111.20**, **MAA members US$125.10**, Order code PSPUM/106

`bookstore.ams.org/pspum-106`




## Translation Surfaces

**Jayadev S. Athreya**, *University of Washington, Seattle, WA*, and **Howard Masur**, *University of Chicago, IL*

This textbook offers an accessible introduction to translation surfaces. Building on modest prerequisites, the authors focus on the fundamentals behind big ideas in the field: ergodic properties of translation flows, counting problems for saddle connections, and associated renormalization techniques. Proofs that go beyond the introductory nature of the book are deftly omitted, allowing readers to develop essential tools and motivation before delving into the literature.

Beginning with the fundamental example of the flat torus, the book goes on to establish the three equivalent definitions of translation surface. An introduction to the moduli space of translation surfaces follows, leading into a study of the dynamics and ergodic theory associated to a translation surface. Counting problems and group actions come to the fore in the latter chapters, giving a broad overview of progress in the 40 years since the ergodicity of the Teichmüller geodesic flow was proven. Exercises are included throughout, inviting readers to actively explore and extend the theory along the way.

*Translation Surfaces* invites readers into this exciting area, providing an accessible entry point from the perspectives of dynamics, ergodicity, and measure theory. Suitable for a one- or two-semester graduate course, it assumes a background in complex analysis, measure theory, and manifolds, while some familiarity with Riemann surfaces and ergodic theory would be beneficial.

*This item will also be of interest to those working in geometry and topology.*

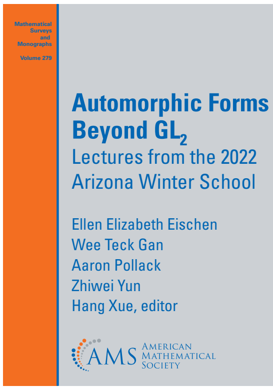
**Graduate Studies in Mathematics**, Volume 242
July 2024, 179 pages, Hardcover, ISBN: 978-1-4704-7655-7, LC 2023053390, 2020 *Mathematics Subject Classification*: 32G15, 30F60, 37A10, 37A25, 37–02, List US$135, **AMS members US$108**, **MAA members US$121.50**, Order code GSM/242

`bookstore.ams.org/gsm-242`

**Graduate Studies in Mathematics**, Volume 242
July 2024, Softcover, ISBN: 978-1-4704-7677-9, LC 2023053390, 2020 *Mathematics Subject Classification*: 32G15, 30F60, 37A10, 37A25, 37–02, List US$89, **AMS members US$71.20**, **MAA members US$80.10**, Order code GSM/242.S

`bookstore.ams.org/gsm-242-s`

# Number Theory




### Automorphic Forms Beyond $GL_2$

Lectures from the 2022 Arizona Winter School

**Hang Xue**, *University of Arizona, Tuscon, AZ*, Editor

**Ellen Elizabeth Eischen**, *University of Oregon, Eugene, OR*, **Wee Teck Gan**, *National University of Singapore, Republic of Singapore*, **Aaron Pollack**, *University of California San Diego, La Jolla, CA*, and **Zhiwei Yun**, *Massachusetts Institute of Technology, Cambridge, MA*

The Langlands program has been a very active and central field in mathematics ever since its conception over 50 years ago. It connects number theory, representation theory and arithmetic geometry, and other fields in a profound way. There are nevertheless very few expository accounts beyond the GL(2) case. This book features expository accounts of several topics on automorphic forms on higher rank groups, including rationality questions on unitary group, theta lifts and their applications to Arthur's conjectures, quaternionic modular forms, and automorphic forms over functions fields and their applications to inverse Galois problems. It is based on the lecture notes prepared for the twenty-fifth Arizona Winter School on "Automorphic Forms beyond GL(2)", held March 5–9, 2022, at the University of Arizona in Tucson. The speakers were Ellen Eischen, Wee Teck Gan, Aaron Pollack, and Zhiwei Yun.

The exposition of the book is in a style accessible to students entering the field. Advanced graduate students as well as researchers will find this a valuable introduction to various important and very active research areas.

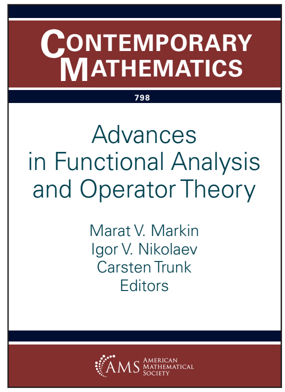
*This item will also be of interest to those working in algebra and algebraic geometry.*

**Mathematical Surveys and Monographs**, Volume 279
April 2024, 187 pages, Softcover, ISBN: 978-1-4704-7492-8, LC 2023053456, 2020 *Mathematics Subject Classification*: 11F70, 11F27, 11F67, 14D24, 20G41, 14F08, List US$135, **AMS members US$108**, **MAA members US$121.50**, Order code SURV/279

bookstore.ams.org/surv-279

# New in Contemporary Mathematics

## Analysis




### Advances in Functional Analysis and Operator Theory

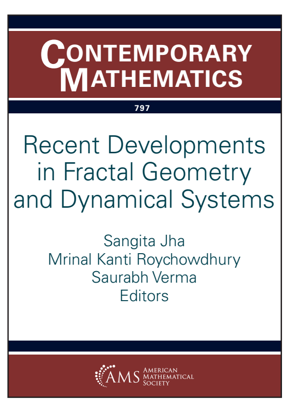
**Marat V. Markin**, *California State University, Fresno, CA*, **Igor V. Nikolaev**, *St. John's University, New York, NY*, and **Carsten Trunk**, *Technische Universität Ilmenau, Germany*, Editors

This volume contains the proceedings of the AMS-EMS-SMF Special Session on Advances in Functional Analysis and Operator Theory, held July 18–22, 2022, at the Université de Grenoble-Alpes, Grenoble, France.

The papers reflect the modern interplay between differential equations, functional analysis, operator algebras, and their applications from the dynamics to quantum groups to number theory. Among the topics discussed are the Sturm-Liouville and boundary value problems, axioms of quantum mechanics, $C^*$-algebras and symbolic dynamics, von Neumann algebras and low-dimensional topology, quantum permutation groups, the Jordan algebras, and the Kadison–Singer transforms.

**Contemporary Mathematics**, Volume 798
May 2024, approximately 238 pages, Softcover, ISBN: 978-1-4704-7305-1, LC 2023042586, 2020 *Mathematics Subject Classification*: 11G10, 34A09, 34B24, 37A55, 46L05, 46L10, 46L70, 46L85, List US$129, **AMS members US$103.20**, **MAA members US$116.10**, Order code CONM/798

bookstore.ams.org/conm-798




### Recent Developments in Fractal Geometry and Dynamical Systems

**Sangita Jha**, *National Institute of Technology, Rourkela, India*, **Mrinal Kanti Roychowdhury**, *University of Texas at Rio Grande Valley, Edinburg, TX*, and **Saurabh Verma**, *Indian Institute of Information Technology Allahabad, Prayagraj, India*, Editors

This volume contains the proceedings of the virtual AMS Special Session on Fractal Geometry and Dynamical Systems, held from May 14–15, 2022.

The content covers a wide range of topics. It includes nonautonomous dynamics of complex polynomials, theory and applications of polymorphisms, topological and geometric problems related to dynamical systems, and also covers fractal dimensions, including the Hausdorff dimension of fractal interpolation functions. Furthermore, the book contains a discussion of self-similar measures as well as the theory of IFS measures associated with Bratteli diagrams. This book is suitable for graduate students interested in fractal theory, researchers interested in fractal geometry and dynamical systems, and anyone interested in the application of fractals in science and engineering. This book also offers a valuable resource for researchers working on applications of fractals in different fields.

*This item will also be of interest to those working in applications.*

**Contemporary Mathematics**, Volume 797
May 2024, approximately 254 pages, Softcover, ISBN: 978-1-4704-7216-0, 2020 *Mathematics Subject Classification*: 28A33, 28A50, 28A80, 35B41, 37D05, 37F10, 37F12, 37F20, List US$129, **AMS members US$103.20**, **MAA members US$116.10**, Order code CONM/797

bookstore.ams.org/conm-797

# New in Memoirs of the AMS

## Algebra and Algebraic Geometry

### Multiplicity-free Representations of Algebraic Groups

**Martin W. Liebeck**, *Imperial College, London, United Kingdom*, **Gary M. Seitz**, *University of Oregon, Eugene, OR*, and **Donna M. Testerman**, *École Polytechnique Fédérale de Lausanne, Switzerland*

**Memoirs of the American Mathematical Society**, Volume 294, Number 1466
March 2024, 268 pages, Softcover, ISBN: 978-1-4704-6905-4, 2020 *Mathematics Subject Classification*: 20G05, 20G20, List US$85, **AMS members US$68**, **MAA members US$76.50**, Order code MEMO/294/1466

bookstore.ams.org/memo-294-1466

## Geometry and Topology

### A Multiplicative Tate Spectral Sequence for Compact Lie Group Actions

**Alice Hedenlund**, *University of Oslo, Norway*, and **John Rognes**, *University of Oslo, Norway*

**Memoirs of the American Mathematical Society**, Volume 294, Number 1468
March 2024, 134 pages, Softcover, ISBN: 978-1-4704-6878-1, 2020 *Mathematics Subject Classification*: 55T25, 55P91, 16E30, List US$85, **AMS members US$68**, **MAA members US$76.50**, Order code MEMO/294/1468

bookstore.ams.org/memo-294-1468

## Probability and Statistics

### Optimal Feedback for Stochastic Linear Quadratic Control and Backward Stochastic Riccati Equations in Infinite Dimensions

**Qi Lü**, *Sichuan University, Chengdu, People's Republic of China*, and **Xu Zhang**, *Sichuan University, Chengdu, People's Republic of China*

*This item will also be of interest to those working in differential equations.*

**Memoirs of the American Mathematical Society**, Volume 294, Number 1467
March 2024, 107 pages, Softcover, ISBN: 978-1-4704-6875-0, 2020 *Mathematics Subject Classification*: 60H15; 93E20, 60H25, List US$85, **AMS members US$68**, **MAA members US$76.50**, Order code MEMO/294/1467

bookstore.ams.org/memo-294-1467

### Kinetic Theory for the Low-Density Lorentz Gas

**Jens Marklof**, *University of Bristol, United Kingdom*, and **Andreas Strömbergsson**, *Uppsala University, Sweden*

*This item will also be of interest to those working in mathematical physics.*

**Memoirs of the American Mathematical Society**, Volume 294, Number 1464
March 2024, 136 pages, Softcover, ISBN: 978-1-4704-6869-9, 2020 *Mathematics Subject Classification*: 82C40, 60G55, 35Q20, List US$85, **AMS members US$68**, **MAA members US$76.50**, Order code MEMO/294/1464

bookstore.ams.org/memo-294-1464

### Groups, Graphs, and Hypergraphs: Average Sizes of Kernels of Generic Matrices with Support Constraints

**Tobias Rossmann**, *University of Galway, Ireland*, and **Christopher Voll**, *Universität Bielefeld, Germany*

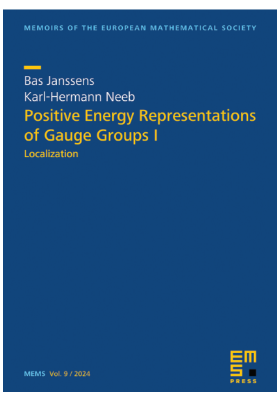
*This item will also be of interest to those working in discrete mathematics and combinatorics.*

**Memoirs of the American Mathematical Society**, Volume 294, Number 1465
March 2024, 120 pages, Softcover, ISBN: 978-1-4704-6868-2, 2020 *Mathematics Subject Classification*: 11M41, 20D15, 20E45, 15B33, 05A15, 05C50, 14M25, 11S80, List US$85, **AMS members US$68**, **MAA members US$76.50**, Order code MEMO/294/1465

`bookstore.ams.org/memo-294-1465`

# New AMS-Distributed Publications

## Algebra and Algebraic Geometry




### Positive Energy Representations of Gauge Groups I
#### Localization

**Bas Janssens**, *Delft University of Technology, The Netherlands*, and **Karl-Hermann Neeb**, *Friedrich-Alexander University, Erlangen-Nuremberg, Germany*

This is the first in a series of papers on projective positive energy representations of gauge groups. Let $\Xi \to M$ be a principal fiber bundle, and let $\Gamma_C(M,\mathrm{Ad}(\Xi))$ be the group of compactly supported (local) gauge transformations. If $P$ is a group of "space-time symmetries" acting on $\Xi \to M$, then a projective unitary representation of $\Gamma_C(M,\mathrm{Ad}(\Xi)) \rtimes P$ is of *positive energy* if every "timelike generator" $p_0 \in p$ gives rise to a Hamiltonian $H(p_0)$ whose spectrum is bounded from below. The authors' main result shows that in the absence of fixed points for the cone of timelike generators, the projective positive energy representations of the connected component $\Gamma_C(M,\mathrm{AD}(\Xi))_0$ come from 1-dimensional $P$-orbits. For compact $M$, this yields a complete classification of the projective positive energy representations in terms of lowest weight representations of affine Kac-Moody algebras. For noncompact $M$, it yields a classification under further restrictions on the space of ground states.

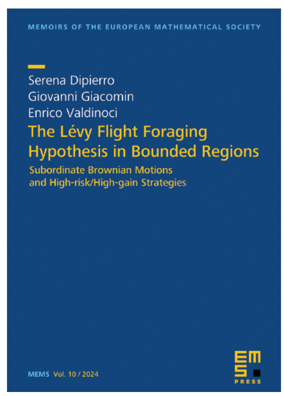
In the second part of this series, the authors consider larger groups of gauge transformations, which also contain global transformations. The present results are used to localize the positive energy representations at (conformal) infinity.

A publication of the European Mathematical Society (EMS). Distributed within the Americas by the American Mathematical Society.

**Memoirs of the European Mathematical Society**, Volume 9
February 2024, 156 pages, Softcover, ISBN: 978-3-98547-067-9, 2020 *Mathematics Subject Classification*: 22E66; 17B15, 17B56, 17B65, 17B66, 17B67, 17B81, 22E60, 22E65, 22E67, List US$75, **AMS members US$60**, Order code EMSMEM/9

`bookstore.ams.org/emsmem-9`

## Differential Equations




### The Lévy Flight Foraging Hypothesis in Bounded Regions
#### Subordinate Brownian Motions and High-risk/ High-gain Strategies

**Serena Dipierro**, *The University of Western Australia, Australia*, **Giovanni Giacomin**, *The University of Western Australia, Australia*, and **Enrico Valdinoci**, *The University of Western Australia, Australia*

The authors investigate the problem of the Lévy flight foraging hypothesis in an ecological niche described by a bounded region of space, with either absorbing or reflecting boundary conditions. To this end, they consider a forager diffusing according to a fractional heat equation in a bounded domain, and they define several efficiency functionals whose optimality is discussed in relation to the fractional exponent $s \in (0,1)$ of the diffusive equation. Such an equation is taken to be the spectral fractional heat equation (with Dirichlet or Neumann boundary conditions).

The authors analyze the biological scenarios in which a target is close to the forager or far from it. In particular, for all the efficiency functionals considered here, they show that if the target is close enough to the forager, then the most rewarding search strategy will be in a small neighborhood of $s = 0$. Interestingly, we show that $s = 0$ is a global pessimizer for some of the efficiency functionals. From this, together with the aforementioned optimality results, the authors deduce that the most rewarding strategy can be unsafe or unreliable in practice, given its proximity with the pessimizing exponent; thus, the forager may opt for a less performant, but safer, hunting method.

However, the biological literature has already collected several pieces of evidence of foragers diffusing with very low Lévy exponents, often in relation with a high energetic content of the prey. It is thereby suggestive to relate these patterns, which are induced by distributions with a very fat tail, with a high-risk/high-gain strategy, in which the forager adopts a potentially very profitable, but also potentially completely unrewarding, strategy due to the high value of the possible outcome.

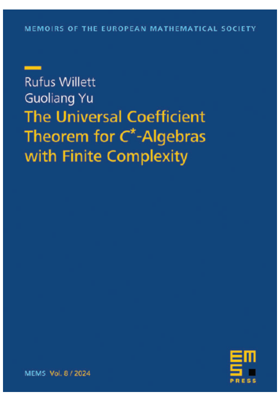
*This item will also be of interest to those working in probability and statistics and applications.*

A publication of the European Mathematical Society (EMS). Distributed within the Americas by the American Mathematical Society.

**Memoirs of the European Mathematical Society**, Volume 10 February 2024, 99 pages, Softcover, ISBN: 978-3-98547-068-6, 2020 *Mathematics Subject Classification*: 92–10; 35Q92, 60G07, 35Q93, 60G22, 35R11, List US$75, **AMS members US$60**, Order code EMSMEM/10

bookstore.ams.org/emsmem-10

# Logic and Foundations




## The Universal Coefficient Theorem for $C^*$-Algebras with Finite Complexity

**Rufus Willett**, *University of Hawaii at Manoa, HI*, and **Guoliang Yu**, *Texas A & M University, College Station, TX*

A $C^*$-algebra satisfies the Universal Coefficient Theorem (UCT) of Rosenberg and Schochet if it is equivalent in Kasparov's $K$-theory to a commutative $C^*$-algebra. This book is motivated by the problem of establishing the range of validity of the UCT, and in particular, whether the UCT holds for all nuclear $C^*$-algebras.

The authors introduce the idea of a $C^*$-algebra that "decomposes" over a class $\mathcal{C}$ of $C^*$-algebras. Roughly, this means that locally there are approximately central elements that approximately cut the $C^*$-algebra into two $C^*$-sub-algebras from $C$ that have well-behaved intersection. The authors show that if a $C^*$-algebra decomposes over the class of nuclear, UCT $C^*$-algebras, then it satisfies the UCT. The argument is based on a Mayer-Vietoris principle in the framework of controlled $K$-theory; the latter was introduced by the authors in an earlier work. Nuclearity is used via Kasparov's Hilbert module version of Voiculescu's theorem, and Haagerup's theorem that nuclear $C^*$-algebras are amenable.

The authors say that a $C^*$-algebra has finite complexity if it is in the smallest class of $C^*$-algebras containing the finite-dimensional $C^*$-algebras, and closed under decomposability; their main result implies that all $C^*$-algebras in this class satisfy the UCT. The class of $C^*$-algebras with finite complexity is large, and comes with an ordinal-number invariant measuring the complexity level. They conjecture that a $C^*$-algebra of finite nuclear dimension and real rank zero has finite complexity; this (and several other related conjectures) would imply the UCT for all separable nuclear $C^*$-algebras. The authors also give new local formulations of the UCT, and some other necessary and sufficient conditions for the UCT to hold for all nuclear $C^*$-algebras.

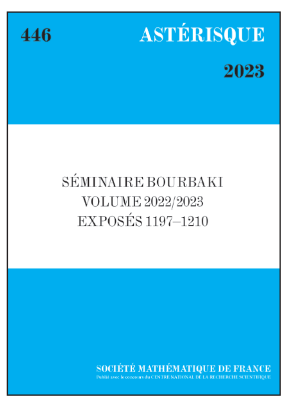
*This item will also be of interest to those working in analysis.*

A publication of the European Mathematical Society (EMS). Distributed within the Americas by the American Mathematical Society.

**Memoirs of the European Mathematical Society**, Volume 8 February 2024, 108 pages, Softcover, ISBN: 978-3-98547-066-2, 2020 *Mathematics Subject Classification*: 19K35; 19K33, 46L05, 46L85, 46L80, List US$75, **AMS members US$60**, Order code EMSMEM/8

bookstore.ams.org/emsmem-8

# Number Theory




## Séminaire Bourbaki

### Volume 2022/2023
### Exposés 1197–1210

This 74th volume of the Bourbaki Seminar gathers the texts of the fourteen lectures delivered during the year 2022/2023: unbounded denominators conjecture, validity of the kinetic theory of gases, pointwise ergodic theory, twisted Lang-Weil theorem, direct factor conjecture, random permutations and Ramanujan graphs, von Neumann algebras and quantum correlations, structure of the group of homeomorphisms of the 2-dimensional sphere, pointwise convergence for the Schrödinger equation, exponential growth in hyperbolic groups, strong forcing axioms and the continuum problem, non-uniqueness of Leray solutions to the Navier-Stokes system, tensor categories in positive characteristic, and rotation invariance for planar percolation.

*This item will also be of interest to those working in algebra and algebraic geometry and differential equations.*

A publication of the Société Mathématique de France, Marseilles (SMF), distributed by the AMS in the US, Canada, and Mexico. Orders from other countries should be sent to the SMF. Members of the SMF receive a 30% discount from list.

**Astérisque**, Number 446

January 2024, 520 pages, Softcover, ISBN: 978-2-85629-984-5, 2020 *Mathematics Subject Classification*: 11F11, 11F30, 14G99, 20H05, 30D35, 30F35, 33C05, 82C22, 35Q20, 35Q70, 82B40, 37A46, 14G17, 14C17, 12L12, 13C14, 13H10, 13D22, 13D02, 14G45, 60B20, 15B52, 46L54, 05C48, 46L06, 46L10, 68Q10, 81P05, 81P45, 37B55, 53D40, 37A05, 37C85, 57S25, 42B10, 42B37, 81U30, 20F67, 20E08, 57K32, 20F69, 03E35, 03E50, 03E57, 03C10, 00A30, 03A05, 35Q30, 76D03, 76D05, 35P15, 18M05, 18M25, 20C20, 60K35, List US$122, **AMS members US$97.60**, Order code AST/446

bookstore.ams.org/ast-446


MC of A
Mathematical Council of the Americas
AMS
AMERICAN MATHEMATICAL SOCIETY
IMSA
The Mathematical Congress of the Americas 2025
July 21-25, 2025
InterContinental Hotel, Miami
MCA 2025
MCA 25 is authorized by the Mathematical Council of the Americas and organized through the Institute of the Mathematical Sciences of the Americas (IMSA) at the University of Miami
https://www.mca2025.org

# Meetings & Conferences of the AMS

## May Table of Contents

The Meetings and Conferences section of the *Notices* gives information on all AMS meetings and conferences approved by press time for this issue. Please refer to the page numbers cited on this page for more detailed information on each event. *Paid meeting registration is required to submit an abstract to a sectional meeting.*

Invited Speakers and Special Sessions are listed as soon as they are approved by the cognizant program committee; the codes listed are needed for electronic abstract submission. For some meetings the list may be incomplete. Information in this issue may be dated.

The most up-to-date meeting and conference information can be found online at `www.ams.org/meetings`.

**Important Information About AMS Meetings:** Potential organizers, speakers, and hosts should refer to `https://www.ams.org/meetings/meetings-general` for general information regarding participation in AMS meetings and conferences.

**Abstracts:** Speakers should submit abstracts on the easy-to-use interactive Web form. No knowledge of LaTeX is necessary to submit an electronic form, although those who use LaTeX may submit abstracts with such coding, and all math displays and similarly coded material (such as accent marks in text) must be typeset in LaTeX. Visit `www.ams.org/cgi-bin/abstracts/abstract.pl`. Questions about abstracts may be sent to `abs-info@ams.org`. Close attention should be paid to specified deadlines in this issue. Unfortunately, late abstracts cannot be accommodated.

### Associate Secretaries of the AMS

**Central Section:** Betsy Stovall, University of Wisconsin–Madison, 480 Lincoln Drive, Madison, WI 53706; email: `stovall@math.wisc.edu`; telephone: (608) 262-2933.

**Eastern Section:** Steven H. Weintraub, Department of Mathematics, Lehigh University, Bethlehem, PA 18015-3174; email: `steve.weintraub@lehigh.edu`; telephone: (610) 758-3717.

**Southeastern Section:** Brian D. Boe, Department of Mathematics, University of Georgia, 220 D W Brooks Drive, Athens, GA 30602-7403; email: `brian@math.uga.edu`; telephone: (706) 542-2547.

**Western Section:** Michelle Manes, University of Hawaii, Department of Mathematics, 2565 McCarthy Mall, Keller 401A, Honolulu, HI 96822; email: `mamanes@hawaii.edu`; telephone: (808) 956-4679.

## Meetings in this Issue



*The AMS strives to ensure that participants in its activities enjoy a welcoming environment. Please see our full Policy on a Welcoming Environment at* `https://www.ams.org/welcoming-environment-policy`.

# Meetings & Conferences of the AMS

IMPORTANT information regarding meetings programs: AMS Sectional Meeting programs do not appear in the print version of the *Notices*. However, comprehensive and continually updated meeting and program information with links to the abstract for each talk can be found on the AMS website. See `https://www.ams.org/meetings`.

Final programs for Sectional Meetings will be archived on the AMS website accessible from the stated URL.

**New: Sectional Meetings Require Registration to Submit Abstracts.** In an effort to spread the cost of the sectional meetings more equitably among all who attend and hence help keep registration fees low, starting with the 2020 fall sectional meetings, you must be registered for a sectional meeting in order to submit an abstract for that meeting. You will be prompted to register on the Abstracts Submission Page. In the event that your abstract is not accepted or you have to cancel your participation in the program due to unforeseen circumstances, your registration fee will be reimbursed.

## Milwaukee, Wisconsin

*University of Wisconsin-Milwaukee*

### April 20–21, 2024

*Saturday – Sunday*

### Meeting #1195

Central Section
Associate Secretary for the AMS: Betsy Stovall

Program first available on AMS website: To be announced
Issue of *Abstracts*: Volume 45, Issue 2

### Deadlines

For organizers: Expired
For abstracts: Expired

*The scientific information listed below may be dated. For the latest information, see* `https://www.ams.org/amsmtgs/sectional.html`.

### Invited Addresses

**Mihaela Ifrim**, University of Wisconsin-Madison, *The small data global well-posedness conjecture for 1D defocusing dispersive flows*.

**Lin Lin**, University of California, Berkeley, *Quantum advantage in scientific computation?*.

**Kevin Schreve**, Louisiana State University, *Homological growth of groups and aspherical manifolds*.

### Special Sessions

*If you are volunteering to speak in a Special Session, you should send your abstract as early as possible via the abstract submission form found at* `https://www.ams.org/cgi-bin/abstracts/abstract.pl`.

*Algebraic methods in graph theory and applications I*, **Tung T. Nguyen**, Western University, **Sunil K. Chebolu**, Illinois State University, and **Jan Minac**, Western University.

*Algorithms, Number Theory, and Cryptography I*, **Jonathan P. Sorenson**, Butler University, **Eric Bach**, University of Wisconsin at Madison, and **Jonathan Webster**, Butler University.

*Applications of Algebra and Geometry I*, **Thomas Yahl**, University of Wisconsin - Madison, and **Jose Israel Rodriguez**, University of Wisconsin Madison.

*Applications of Numerical Algebraic Geometry I*, **Emma R Cobian**, University of Notre Dame.

*Artificial Intelligence in Mathematics I*, **Tony Shaska**, University of Michigan, **Alessandro Arsie**, The University of Toledo, **Elira Curri**, Oakland University, Rochester Hills, MI, 48126, and **Mee Seong Im**, United States Naval Academy.

*Automorphisms of Riemann Surfaces and Related Topics I*, **Aaron D. Wootton**, University of Portland, **Jennifer Paulhus**, Grinnell College, **Sean Allen Broughton**, Rose-Hulman Institute of Technology (emeritus), and **Tony Shaska**, Oakland University.

*Cluster algebras, Hall algebras and representation theory I*, **Xueqing Chen**, University of Wisconsin-Whitewater, and **Yiqiang Li**, SUNY At Buffalo.

*Combinatorial and geometric themes in representation theory I*, **Jeb F. Willenbring**, UW-Milwaukee, and **Pamela E. Harris**, University of Wisconsin, Milwaukee.

*Complex Dynamics and Related Areas I*, **James Waterman**, Stony Brook University, and **Alastair N Fletcher**, Northern Illinois University.

*Computability Theory I*, **Matthew Harrison-Trainor**, University of Illinois Chicago, and **Steffen Lempp**, University of Wisconsin-Madison.

*Connections between Commutative Algebra and Algebraic Combinatorics I*, **Alessandra Costantini**, Oklahoma State University, **Matthew James Weaver**, University of Notre Dame, and **Alexander T Yong**, University of Illinois at Urbana-Champaign.

*Developments in hyperbolic-like geometry and dynamics I*, **Jonah Gaster**, University of Wisconsin-Milwaukee, **Andrew Zimmer**, University of Wisconsin-Madison, and **Chenxi Wu**, University of Wisconsin At Madison.

*Geometric group theory I*, **G Christopher Hruska**, University of Wisconsin-Milwaukee, and **Emily Stark**, Wesleyan University.

*Geometric Methods in Representation Theory I*, **Daniele Rosso**, Indiana University Northwest, and **Joshua Mundinger**, University of Wisconsin - Madison.

*Harmonic Analysis and Incidence Geometry I*, **Sarah E Tammen** and **Terence L. J Harris**, UW Madison, and **Shengwen Gan**, University of Wisconsin - Madison.

*Mathematical aspects of cryptography and cybersecurity I*, **Lubjana Beshaj**, Army Cyber Institute.

*Model Theory I*, **Uri Andrews**, University of Wisconsin-Madison, and **James Freitag**, University of Illinois Chicago.

*New research and open problems in combinatorics I*, **Pamela Estephania Harris**, University of Wisconsin, Milwaukee, **Erik Insko**, Central College, and **Mohamed Omar**, York University.

*Nonlinear waves I*, **Mihaela Ifrim**, University of Wisconsin-Madison, and **Daniel I Tataru**, UC Berkeley.

*Panorama of Holomorphic Dynamics I*, **Suzanne Lynch Boyd**, University of Wisconsin Milwaukee, and **Rodrigo A. Perez** and **Roland Roeder**, Indiana University - Purdue University Indianapolis.

*Posets in algebraic and geometric combinatorics I*, **Martha Yip**, University of Kentucky, and **Rafael S. González D'León**, Loyola University Chicago.

*Ramification in Algebraic and Arithmetic Geometry I*, **Charlotte Ure**, Illinois State University, and **Nick Rekuski**, Wayne State University.

*Recent Advances in Nonlinear PDEs and Their Applications I*, **Xiang Wan**, Loyola University Chicago, **Rasika Mahawattege**, University of Maryland, Baltimore County, and **Madhumita Roy**, Graduate Student, University of Memphis.

*Recent Advances in Numerical PDE Solvers by Deep Learning I*, **Dexuan Xie**, University of Wisconsin-Milwaukee, and **Zhen Chao**, University of Michigan-Ann Arbor.

*Recent Developments in Harmonic Analysis I*, **Naga Manasa Vempati**, Louisiana State University, **Nathan A. Wagner**, Brown University, and **Bingyang Hu**, Auburn University.

*Recent trends in nonlinear PDE I*, **Fernando Charro** and **Catherine Lebiedzik**, Wayne State University, and **Md Nurul Raihen**, Fontbonne University.

*Stochastic Control and Related Fields: A Special Session in Honor of Professor Stockbridge's 70th Birthday I*, **Chao Zhu**, University of Wisconsin-Milwaukee, and **MoonJung Cho**, U.S. Bureau of Labor Statistics.

*The Algebras and Special Functions around Association Schemes I*, **Paul M Terwilliger**, U. Wisconsin-Madison, **Sarah R Bockting-Conrad**, DePaul University, and **Jae-Ho Lee**, University of North Florida.

# San Francisco, California

*San Francisco State University*

### May 4–5, 2024

*Saturday – Sunday*

### Meeting #1196

Western Section
Associate Secretary for the AMS: Michelle Ann Manes

Program first available on AMS website: To be announced
Issue of *Abstracts*: Volume 45, Issue 3

### Deadlines

For organizers: Expired
For abstracts: Expired

*The scientific information listed below may be dated. For the latest information, see* `https://www.ams.org/amsmtgs/sectional.html`.

## Invited Addresses

**Julia Yael Plavnik**, Indiana University, *Title to be announced*.

**Mandi A. Schaeffer Fry**, University of Denver, *Counting with blocks and hide-and-seek with character tables: Brauer's problems and beyond*.

## Special Sessions

*If you are volunteering to speak in a Special Session, you should send your abstract as early as possible via the abstract submission form found at* `https://www.ams.org/cgi-bin/abstracts/abstract.pl`.

*Commutative and Noncommutative Algebra, Together at Last*, **Pablo S. Ocal**, University of California, Los Angeles, **Benjamin Briggs**, University of Copenhagen, and **Janina C Letz**, Bielefeld University.

*Diagrammatic Algebras in Representation Theory and Beyond*, **Mee Seong Im**, United States Naval Academy, **Liron Speyer**, Okinawa Institute of Science and Technology, **Arik Wilbert**, University of Georgia, and **Jieru Zhu**, University of Queensland.

*Extremal Combinatorics and Connections*, **Sam Spiro**, Rutgers University, and **Van Magnan**, University of Montana.

*Geometry and Topology of Quantum Phases of Matter*, **Ralph Martin Kaufmann**, Purdue University, and **Markus J Pflaum**, University of Colorado.

*Geometry, Integrability, Symmetry and Physics*, **Birgit Kaufmann** and **Sasha Tsymbaliuk**, Purdue University.

*Groups and Representations (associated with Invited Address by Mandi Schaeffer Fry)*, **Nathaniel Thiem**, University of Colorado, **Mandi A. Schaeffer Fry**, University of Denver, and **Klaus Lux**, University of Arizona.

*Homological Methods in Commutative Algebra & Algebraic Geometry*, **Ritvik Ramkumar**, Cornell University, **Michael Perlman**, University of Minnesota, and **Aleksandra C Sobieska**, University of Wisconsin-Madison.

*Inverse Problems*, **Hanna E. Makaruk**, Los Alamos National Laboratory, Los Alamos, NM, and **Robert M. Owczarek**, University of New Mexico.

*Mathematical Fluid Dynamics*, **Igor Kukavica** and **Juhi Jang**, University of Southern California, and **Wojciech S. Ozanski**, Florida State University.

*Mathematical Modeling of Complex Ecological and Social Systems*, **Daniel Brendan Cooney**, University of Illinois Urbana-Champaign, **Mari Kawakatsu**, University of Pennsylvania, and **Chadi M Saad-Roy**, University of California, Berkeley.

*Partial Differential Equations and Convexity*, **Ben Weinkove**, Northwestern University, **Stefan Steinerberger**, University of Washington, Seattle, and **Albert Chau**, University of British Columbia.

*Partial Differential Equations of Quantum Physics*, **Israel Michael Sigal**, University of Toronto, and **Stephen Gustafson**, University of British Columbia.

*Random Structures, Computation, and Statistical Inference*, **Lutz Warnke**, University of California, San Diego, and **Ilias Zadik**, Yale University.

*Recent Advances in Differential Geometry*, **Lihan Wang**, California State University, Long Beach, **Zhiqin Lu**, UC Irvine, and **Shoo Seto** and **Bogdan D. Suceavă**, California State University, Fullerton.

*Recent Developments in Commutative Algebra*, **Arvind Kumar**, **Louiza Fouli**, and **Michael Robert DiPasquale**, New Mexico State University.

*Representations of Lie Algebras and Lie Superalgebras*, **Dimitar Grantcharov**, University of Texas At Arlington, **Daniel Nakano**, University of Georgia, and **Vera Serganova**, UC Berkeley.

*Research in Combinatorics by Early Career Mathematicians*, **Nicholas Mayers**, North Carolina State University, and **Laura Colmenarejo**, NCSU.

*Special Session in Celebration of Bruce Reznick's Retirement*, **Katie Anders**, University of Texas at Tyler, **Simone Sisneros-Thiry**, California State University- East Bay, and **Dana Neidmann**, Centre College.

*Tensor Categories and Noncommutative Algebras, I (associated with invited address by Julia Plavnik)*, **Ellen E Kirkman**, Wake Forest University, and **Julia Yael Plavnik**, Indiana University, Bloomington.

### Contributed Paper Sessions

*AMS Contributed Paper Session*, **Michelle Ann Manes**, University of Hawaii.

# Palermo, Italy

### July 23–26, 2024

*Tuesday – Friday*
Associate Secretary for the AMS: Brian D. Boe
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced
For abstracts: To be announced

# San Antonio, Texas

*University of Texas, San Antonio*

### September 14–15, 2024

*Saturday – Sunday*

### Meeting #1198

Central Section
Associate Secretary for the AMS: Betsy Stovall
Program first available on AMS website: To be announced
Issue of *Abstracts*: Volume 45, Issue 3

### Deadlines

For organizers: Expired
For abstracts: July 23, 2024

*The scientific information listed below may be dated. For the latest information, see* `https://www.ams.org/amsmtgs/sectional.html`.

### Invited Addresses

**James A M Alvarez**, The University of Texas at Arlington, *Title To Be Announced*.

**Jason R Schweinsberg**, University of California San Diego, *Title To Be Announced*.

**Anne Shiu**, Texas A&M University, *Dynamics of Biochemical Reaction Networks*.

### Special Sessions

*If you are volunteering to speak in a Special Session, you should send your abstract as early as possible via the abstract submission form found at* `https://www.ams.org/cgi-bin/abstracts/abstract.pl`.

*Additive Number Theory and Modular Forms I* (Code: SS 17A), **Debanjana Kundu**, University of Texas - Rio Grande Valley, and **Brandt Kronholm**, University of Texas Rio Grande Valley.

*Advances in Coding Theory and Cryptography I* (Code: SS 27A), **Henry Chimal-Dzul**, University of Notre Dame, and **Jingbo Liu**, Texas A&M University-San Antonio.

*Advances in Differential Equations: Theory, Methods, and Applications I* (Code: SS 10A), **Faranak Rabiei**, Texas A & M University Kingsville, **Aden Omar Ahmed**, Texas A&M University-Kingsville, and **Dongwook Kim**, Texas A & M University Kingsville.

*Advances in Mathematical and Numerical Analysis of Partial Differential Equations for Application-Oriented Computations I* (Code: SS 25A), **Bruce A Wade**, University of Louisiana at Lafayette, **Qin Sheng**, Baylor University, **Abdul Q.M. Khaliq**, Middle Tennessee State University, **Jaeun Ku**, Oklahoma State University, and **Xiang-Sheng Wang** and **Yangwen Zhang**, University of Louisiana at Lafayette.

*Applications of Algebraic Geometry I* (Code: SS 23A), **Frank Sottile**, Texas A&M University, **Alperen Ergur**, University of Texas at San Antonio, and **Anne Joyce Shiu**, Texas A&M University.

*Applications of analysis, topology and set theory to model theory I* (Code: SS 33A), **Eduardo Dueñez** and **Jose N Iovino**, The University of Texas at San Antonio.

*Applications of Probability in Biology I* (Code: SS 8A), **Jason R Schweinsberg**, University of California San Diego.

*A Showcase of Algebraic Geometry at Undergraduate Institutions I* (Code: SS 22A), **David Swinarski**, Fordham University, **Julie Rana**, Lawrence University, and **Han-Bom Moon**, Fordham University.

*Commutative algebra and connections to combinatorics I* (Code: SS 28A), **Michael Robert DiPasquale**, **Louiza Fouli**, and **Arvind Kumar**, New Mexico State University.

*Differential Geometry I* (Code: SS 1A), **Alvaro Pampano**, Texas Tech University, **Bogdan D. Suceava**, California State University Fullerton, and **Magdalena Daniela Toda**, Texas Tech University and NSF.

*Dynamical systems: Statistical properties, spectral theory, and fractal geometry I* (Code: SS 9A), **Mrinal Kanti Roychowdhury**, School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, and **William R Ott**, University of Houston.

*Enumerative Combinatorics I* (Code: SS 16A), **Brian K. Miceli**, Trinity University, and **Lara Pudwell**, Valparaiso University.

*Geometric Group Theory and Low-Dimensional Topology I* (Code: SS 37A), **George Domat** and **Khanh Le**, Rice University, **Jing Tao**, University of Oklahoma, and **Christopher Jay Leininger**, Rice University.

*Graph Theory I* (Code: SS 15A), **Youngho Yoo** and **Chun-Hung Liu**, Texas A&M University.

*Harmonic Analysis, Geometric Measure Theory and PDE I* (Code: SS 14A), **Dorina I. Mitrea** and **Marius Mitrea**, Baylor University.

*Homological and combinatorial methods in noncommutative algebra I* (Code: SS 5A), **Amrei Oswald** and **Be"eri Greenfeld**, University of Washington.

*Homological Commutative Algebra I* (Code: SS 13A), **Luigi Ferraro**, University of Texas Rio Grande Valley, and **Alexis Hardesty**, Texas Woman's University.

*Inquiry Oriented Learning in the Mathematics Classroom I* (Code: SS 34A), **Carolyn Luna**, University of Texas At San Antonio, and **Jennifer Austin**, University of Texas at Austin.

*L-functions and Automorphic Forms I* (Code: SS 24A), **Lea Beneish**, University of North Texas, and **Melissa Emory**, Oklahoma State University.

*Link invariants and surfaces in 4-manifolds I* (Code: SS 21A), **Michael Willis** and **Sherry Gong**, Texas A&M University.

*Machine Learning, Data Science and Related Fields I* (Code: SS 3A), **Hansapani Rodrigo**, The University of Texas Rio Grande Valley, and **Lakshmi Roychowdhury** and **Mrinal Kanti Roychowdhury**, University of Texas Rio Grande Valley.

*Mathematical Modeling at the Interface of Ecology, Epidemiology, and Human Behavior I* (Code: SS 12A), **Tamer Oraby**, University of Texas - Rio Grande Valley, **Lale Asik**, University of the Incarnate Word, **Ummugul Bulut**, ubulut@uiwtx.edu, and **Md Rafiul Islam**, University of the Incarnate Word.

*Mathematical Physics and Numerical Methods I* (Code: SS 39A), **Vu Hoang** and **Jose Morales**, University of Texas at San Antonio.

*Mathematics of Infectious Disease Emergence, Spread, and Control I* (Code: SS 38A), **Zhuolin Qu**, University of Texas at San Antonio, and **Michael A. Robert**, Virginia Tech.

*Mathematics: The gateway to Social Justice I* (Code: SS 31A), **Juan B. Gutiérrez**, University of Texas at San Antonio, **James Broda**, Washington and Lee University, **Funda Gultepe**, University of Toledo, **Ron Buckmire**, Occidental College, **Matthew Salomone**, Bridgewater State University, **Joseph Edward Hibdon**, Northeastern Illinois University, and **Terrance Pendleton**, Drake University.

*Methods & Applications of Data-driven Manufacturing I* (Code: SS 19A), **Kristen Lee Hallas**, The University of Texas Rio Grande Valley, and **Benjamin Peters** and **Jianzhi Li**, University of Texas Rio Grande Valley.

*Modeling and analysis in biological and epidemiological systems I* (Code: SS 18A), **Michael Lindstrom**, The University of Texas Rio Grande Valley, and **Erwin Suazo** and **Zhaosheng Feng**, University of Texas Rio Grande Valley.

*Non-Archimedean, Algebraic, Tropical Geometry and applications I* (Code: SS 30A), **Jackson S. Morrow**, University of North Texas, and **Farbod Shokrieh**, University of Washington.

*Noncommutative Geometry and Analysis I* (Code: SS 20A), **Zhizhang Xie**, **Guoliang Yu**, **Bo Zhu**, and **Simone Cecchini**, Texas A&M University.

*Operator algebras, quantum information and computation I* (Code: SS 36A), **Jose A Morales Escalante**, University of Texas at San Antonio, and **Marius Junge**, University of Illinois, Urbana and Champaign.

*Periodicity in Quantum Systems I* (Code: SS 11A), **Long Li**, Rice University, **Wencai Liu**, Texas A&M University, and **Tal Malinovitch**, Rice Universityt.

*Quasi-periodic and Disordered Systems I* (Code: SS 7A), **Alberto Takase**, Rice University, **Omar Hurtado**, University of California, Irvine, and **Matthew H Faust**, Texas A&M University.

*Recent developments on local and nonlocal PDEs I* (Code: SS 6A), **Fernando Charro**, Wayne State University, and **Thialita Nascimento**, Iowa State University.

*Recent studies in topics related to ion channel problems I* (Code: SS 29A), **Mingji Zhang**, New Mexico Institute of Mining and Technology, and **Saulo Orizaga**, New Mexico Tech.

*Recent trends in differential equations applied to biological processes I* (Code: SS 32A), **Rachidi B. Salako**, University of Nevada, Las Vegas, and **Markjoe O. Uba** and **Maria Amarakristi Onyido**, Northern Illinois University.

*Research in Post-Secondary Teaching and Learning of Mathematics I* (Code: SS 40A), **James A M Alvarez**, The University of Texas at Arlington, and **Paul Christian Dawkins**, Texas State University.

*Spectral Theory of Schrödinger Operators and Related Topics I* (Code: SS 26A), **Christoph Fischbacher**, **Fritz Gesztesy**, and **Jon Harrison**, Baylor University.

*The many scales of mathematical analysis of fluid I* (Code: SS 4A), **Xin Liu**, Texas A&M University, **Quyuan Lin**, Clemson University, and **Cheng Yu**, University of Florida.

*Theoretical and Numerical Aspects of Nonlinear Dispersive Wave Equations I* (Code: SS 35A), **Baofeng Feng**, University of Texas Rio Grande Valley, and **Geng Chen** and **Yannan Shen**, University of Kansas.

*Topics in Convexity I* (Code: SS 2A), **Zokhrab Mustafaev**, University of Houston-Clear Lake.

# Savannah, Georgia

*Georgia Southern University*

### October 5–6, 2024

*Saturday – Sunday*

### Meeting #1199

Southeastern Section
Associate Secretary for the AMS: Brian D. Boe

Program first available on AMS website: Not applicable
Issue of *Abstracts*: Volume 45, Issue 4

### Deadlines

For organizers: Expired
For abstracts: August 13, 2024

*The scientific information listed below may be dated. For the latest information, see* `https://www.ams.org/amsmtgs/sectional.html`.

### Invited Addresses

**Peter Bubenik**, University of Florida, *To Be Announced*.
**Akos Magyar**, University of Georgia, *To Be Announced*.
**Sarah Peluse**, Princeton/IAS, *To Be Announced*.

### Special Sessions

*If you are volunteering to speak in a Special Session, you should send your abstract as early as possible via the abstract submission form found at* `https://www.ams.org/cgi-bin/abstracts/abstract.pl`.

*Advanced Topics in Graph Theory and Combinatorics.*, **Songling Shan**, Auburn University, and **Zi-Xia Song**, University of Central Florida.

*Advances in applied algebraic geometry*, **Kisun Lee** and **Michael Byrd**, Clemson University.

*Applicable Analysis of Multi-physics Partial Differential Equations Systems*, **George Avalos**, University of Nebraska-Lincoln, and **Justin Thomas Webster**, University of Maryland, Baltimore County.

*Biological Systems Modeling and Analysis: recent progress and current challenges*, **Dawit Denu**, Georgia Southern University.

*Commutative Algebra*, **Saeed Nasseh**, **Tricia Muldoon Brown**, and **Alina C. Iacob**, Georgia Southern University.

*Convexity, Probability, and Asymptotic Geometric Analysis*, **Galyna Livshyts**, Georgia Institute of Technology, **Steven Hoehner**, Longwood University, and **Stephanie Mui**, Georgia Institute of Technology.

*Deterministic and Stochastic PDEs: Theoretical and Numerical Analyses*, **Pelin Guven Geredeli**, Clemson University, and **Xiang Wan**, Loyola University Chicago.

*Dynamical Systems and Control Systems with Applications*, **Yan Wu**, Georgia Southern University, and **Liancheng Wang**, Kennesaw State University.

*Extremal and structural graph theory.*, **Ruth Luo**, University of South Carolina, and **Zhiyu Wang**, Georgia Institute of Technology.

*Fluids, Waves, and Free Boundaries.*, **David M. Ambrose**, Drexel University.

*Geometric Maximal Operators and Related Topics.*, **Paul Hagelstein**, Baylor University, and **Alex Stokolos**, Georgia Southern University.

*Harmonic analysis, fractals, and related topics in memory of Ka-Sing Lau and Robert Strichartz*, **Sze-Man Ngai**, Georgia Southern University, and **Alexander Teplyaev**, University of Connecticut.

*Modules over Commutative Rings*, **Laura Ghezzi**, New York City College of Technology and The Graduate Center-Cuny, and **Joseph P Brennan**, University of Central Florida.

*Recent advances in Molecular based Computational and Mathematical Bioscience*, **Shan Zhao**, University of Alabama, and **Zhan Chen**, Georgia Southern University.

*Recent Advances of PDEs in Modern Mathematical Physics: Theory and Applications*, **Yuanzhen Shao**, The University of Alabama, and **Yi Hu** and **Shijun Zheng**, Georgia Southern University.

*Recent Progress in Numerical Methods for PDEs*, **Xuejian Li** and **Leo Rebholz**, Clemson University.

*Topics in commutative algebra and algebraic geometry*, **Prashanth Sridhar**, Charles University, Prague, and **Michael Brown**, Auburn University.

*Topological Data Analysis, Theory and Applications*, **Peter Bubenik** and **Kevin P. Knudson**, University of Florida.

# Albany, New York

*University at Albany*

### October 19–20, 2024

*Saturday – Sunday*

### Meeting #1200

Eastern Section
Associate Secretary for the AMS: Steven H. Weintraub

Program first available on AMS website: Not applicable
Issue of *Abstracts*: Volume 45, Issue 4

### Deadlines

For organizers: Expired
For abstracts: August 27, 2024

*The scientific information listed below may be dated. For the latest information, see* `https://www.ams.org/amsmtgs/sectional.html`.

### Invited Addresses

**Jennifer Balakrishnan**, Boston University, *Title to be announced*.
**Jose Perea**, Northeastern University, *Title to be announced*.
**Richard Rimanyi**, UNC, *Title to be Announced*.

# Riverside, California

*University of California, Riverside*

### October 26–27, 2024

*Saturday – Sunday*

### Meeting #1201

Western Section
Associate Secretary for the AMS: Michelle Ann Manes

Program first available on AMS website: Not applicable
Issue of *Abstracts*: Volume 45, Issue 4

### Deadlines

For organizers: Expired
For abstracts: September 3, 2024

# Auckland, New Zealand

### December 9–13, 2024

*Monday – Friday*
Associate Secretary for the AMS: Steven H. Weintraub
Program first available on AMS website: To be announced

Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced
For abstracts: To be announced

# Seattle, Washington

*Washington State Convention Center and the Sheraton Seattle Hotel*

### January 8–11, 2025

*Wednesday – Saturday*

### Meeting #1203

Associate Secretary for the AMS: Brian D. Boe
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: Expired

For abstracts: September 10, 2024

# Clemson, South Carolina

*Clemson University*

### March 8–9, 2025

*Saturday – Sunday*
Southeastern Section
Associate Secretary for the AMS: Brian Boe, University of Georgia
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced

For abstracts: To be announced

*The scientific information listed below may be dated. For the latest information, see* `https://www.ams.org/amsmtgs/sectional.html`.

### Invited Addresses

**Sarah Hart**, Birkbeck, University of London and Gresham College, *Title to be announced* (Einstein Public Lecture).

# Lawrence, Kansas

*University of Kansas*

### March 29–30, 2025

*Saturday – Sunday*
Central Section
Associate Secretary for the AMS: Betsy Stovall, University of Wisconsin-Madison
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced

For abstracts: To be announced

# Hartford, Connecticut

*Hosted by University of Connecticut; taking place at the Connecticut Convention Center and Hartford Marriott Downtown*

### April 5–6, 2025

*Saturday – Sunday*
Eastern Section
Associate Secretary for the AMS: Steven H. Weintraub
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced
For abstracts: To be announced

# San Luis Obispo, California

*Cal Poly San Luis Obispo*

### May 3–4, 2025

*Saturday – Sunday*
Western Section
Associate Secretary for the AMS: Michelle Ann Manes, American Institute of Mathematics
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced
For abstracts: To be announced

# St. Louis, Missouri

*St. Louis University*

### October 18–19, 2025

*Saturday – Sunday*
Central Section
Associate Secretary for the AMS: Betsy Stovall, University of Wisconsin-Madison
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced
For abstracts: To be announced

# Denver, Colorado

*University of Denver*

### December 6–7, 2025

*Saturday – Sunday*
Western Section
Associate Secretary for the AMS: Michelle Ann Manes, American Institute of Mathematics
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced
For abstracts: To be announced

# Washington, District of Columbia

*Walter E. Washington Convention Center and Marriott Marquis Washington DC*

### January 4–7, 2026

*Sunday – Wednesday*
Associate Secretary for the AMS: Betsy Stovall
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced
For abstracts: To be announced

# Fargo, North Dakota

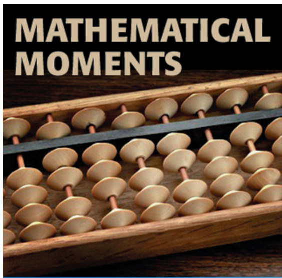
*North Dakota State University*

### April 18–19, 2026

*Saturday – Sunday*
Central Section
Associate Secretary for the AMS: Betsy Stovall, University of Wisconsin-Madison
Program first available on AMS website: To be announced
Issue of *Abstracts*: To be announced

### Deadlines

For organizers: To be announced
For abstracts: To be announced




# Keeping the Lights On

When you flip a switch to turn on a light, where does that energy come from? In a traditional power grid, electricity is generated at large power plants and then transmitted long distances. But now, individual homes and businesses with solar panels can generate some or all of their own power and even send energy into the rest of the grid. Modifying the grid so that power can flow in both directions depends on mathematics. With linear programming and operations research, engineers design efficient and reliable systems that account for constraints like the electricity demand at each location, the costs of solar installation and distribution, and the energy produced under different weather conditions. Similar mathematics helps create "microgrids"—small local systems that can operate independent of the main grid.

A microgrid can keep the power flowing when a natural disaster knocks the main grid offline. For example, after Japan's devastating earthquake and tsunami in 2011, a microgrid powered in part by solar panels provided energy to a university and hospital in Sendai. A recently completed microgrid in Chicago's historically Black Bronzeville area—the US's first neighborhood-scale microgrid—demonstrates that such technology can increase the resilience of the communities most impacted by natural disasters. Researchers continue to develop their methods, often turning equations into computer simulations to visualize new microgrid designs. A related area of study is solar forecasting. Mathematical models that predict levels of sunlight days in advance will allow grid operators to better plan the flow of power from the Sun to your home, school, and community.

**For More Information**: "Solar Systems Integration Basics," Solar Energy Technologies Office, US Department of Energy.

Watch an interview with an expert!

MM/163

AMERICAN MATHEMATICAL SOCIETY
*Advancing research. Creating connections.*

*The **Mathematical Moments** program promotes appreciation and understanding of the role mathematics plays in science, nature, technology, and human culture.*

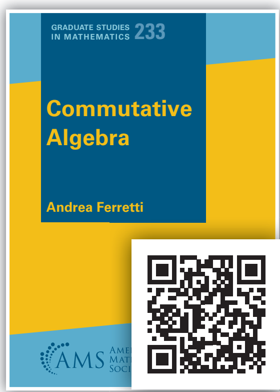
**www.ams.org/mathmoments**

American Mathematical Society
Distribution Center

35 Monticello Place,
Pawtucket, RI 02861 USA

# Two new volumes for the study of commutative algebra

Andrea Feretti offers an introductory volume perfect for use in a course, followed by a second volume for further study of homological methods.

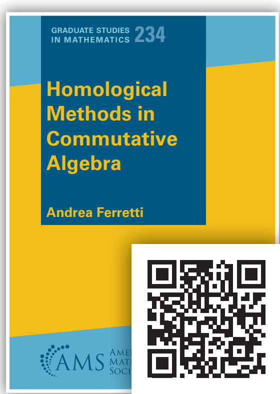



## Commutative Algebra

**Andrea Ferretti**

This book provides an introduction to classical methods in commutative algebra and their applications to number theory, algebraic geometry, and computational algebra. The use of number theory as a motivating theme throughout the book provides a rich and interesting context for the material covered. In addition, many results are reinterpreted from a geometric perspective, providing further insight and motivation for the study of commutative algebra.


**Graduate Studies in Mathematics**, Volume 233; 2023; 373 pages; Softcover; ISBN: 978-1-4704-7434-8; List US$89; AMS members US$71.20; MAA members US$80.10; Order code GSM/233.S


## Homological Methods in Commutative Algebra

**Andrea Ferretti**

This book develops the machinery of homological algebra and its applications to commutative rings and modules. It assumes familiarity with basic commutative algebra, for example, as covered in the author's book, *Commutative Algebra* (**Graduate Studies in Mathematics, Volume 233**).


**Graduate Studies in Mathematics**, Volume 234; 2023; 411 pages; Hardcover; ISBN: 978-1-4704-7128-6; List US$135; AMS members US$108; MAA members US$121.50; Order code GSM/234


*"In my opinion, Volume 1 is an excellent choice for a one-semester introductory course on commutative algebra.*

*Together, this two-volume set provides an engaging and friendly introduction to the subject, and is a welcome addition to the literature."*

*—Pramod Achar, for the* Notices of the AMS

AMS AMERICAN MATHEMATICAL SOCIETY

*Advancing research. Creating connections.*