INTERACTIONS OF COMPUTATIONAL COMPLEXITY THEORY AND MATHEMATICS

AVI WIGDERSON

ABSTRACT

[This paper is a (modified, self contained) chapter in my recent book on computational complexity theory [176], called *Mathematics and Computation*, available online at https://www.math.ias.edu/avi/book].

We survey some concrete interaction areas between computational complexity theory and different fields of mathematics. We hope to demonstrate here that hardly any area of modern mathematics is untouched by the computational connection (which in some cases is completely natural and in others may seem quite surprising). In my view, the breadth, depth, beauty, and novelty of these connections is inspiring, and speaks to a great potential of future interactions (which indeed, are quickly expanding). We aim for *variety*. We give short, simple descriptions (without proofs or much technical detail) of ideas, motivations, results, and connections; this will hopefully entice the reader to dig deeper. Each vignette focuses only on a single topic within a large mathematical field, and is meant to be illustrative rather that comprehensive. We cover the following:

- Number Theory: *Primality testing*
- Combinatorial Geometry: *Point-line incidences*
- Operator Theory: *The Kadison–Singer problem*
- Metric Geometry: *Distortion of embeddings*
- Group Theory: *Generation and random generation*
- Statistical Physics: *Monte Carlo Markov chains*
- Analysis and Probability: *Noise stability*
- Lattice Theory: *Short vectors*
- Invariant Theory: *Group actions on matrix tuples (and beyond)*

MATHEMATICS SUBJECT CLASSIFICATION 2020

68Q01

KEYWORDS

Computational complexity

ICM INTERNATIONAL CONGRESS OF MATHEMATICIANS 2022 JULY 6—14

Proc. Int. Cong. Math. 2022, Vol. 2, pp. 1392–1432
DOI 10.4171/ICM2022/190

1. INTRODUCTION

The Theory of Computation (ToC) lays out the mathematical foundations of computer science. I am often asked if ToC is a branch of Mathematics, or of Computer Science. The answer is easy: it is clearly both (and in fact, much more). Ever since Turing's 1936 definition of the *Turing machine*, we have had a formal mathematical model of computation that enables the rigorous mathematical study of computational tasks, algorithms to solve them, and the resources these require. At the same time, the simple description of the Turing machine allowed its simple logical structure to be implemented in hardware, and its universal applicability fueled the rapid development of computer technology, which now dominates our life.

Computation was part mathematics from its origins, and motivated many of its developments. Algorithmic questions have occupied mathematicians throughout history (as elaborated in the introduction to the book [176]), and this naturally grew considerably when computers arrived. However, the advent of *computational complexity theory* over the past few decades has greatly expanded and deepened these connections. The study of new diverse models generated and studied in complexity theory broadened the nature of mathematical problems it encountered and formulated, and the mathematical areas and tools which bear upon these problems. This expansion has led to numerous new interactions that enrich both disciplines. This survey tells the stories of some of these interactions with different mathematical fields, illustrating their diversity.

We note in passing that a similar explosion of connections and interactions is underway between ToC and practically *all* sciences. These stem from computational aspects of diverse natural processes, which beg for algorithmic modeling and analysis. As with mathematics, these interactions of ToC with the sciences enrich both sides, expose *computation* as a central notion of intellectual thought, and highlight its study as an independent discipline, whose mission and goals expand way beyond those emanating from its parent fields of Math and CS. But this is the subject of a different survey (which I partly provide in the last chapter of [176]).

Back to the interactions of computational complexity theory and different areas of math. I have chosen to focus on essentially one problem or development within each mathematical field. Typically, this touches only a small subarea, and does not do justice to a wealth of other connections. Thus each vignette should be viewed as a demonstration of a larger body of work and even bigger potential. Indeed, while in some areas the collaborations are quite well established, in others they are just budding, with lots of exciting problems waiting to be solved and theories to be developed. Furthermore, the connections to algorithms and complexity (which I explain in each) are quite natural in some areas, but quite surprising in others. While the descriptions of each topic are relatively short, they include background and intuition, as well as further reading material. Indeed, I hope these vignettes will tempt the reader to explore further.

We note that new connections are discovered at a rapid pace. A strong case in point is the recent complexity-theoretic breakthrough of $\mathcal{MIP}^* = \mathcal{RE}$ [100], establishing the sur-

prising power of quantum, multiprover interactive proof systems. This paper had already discussed several surprising applications resolving key conjectures in different mathematical areas, including operator algebras, quantum information theory, and group theory, and now more implications of the techniques and results are being pursued.

The sections below can be read in any order. The selection of fields and foci was affected by my personal taste and limited knowledge. More connections to other fields like Combinatorics, Optimization, Logic, Topology, Coding Theory, and Information Theory appear in parts of the book [176].

2. NUMBER THEORY

As mentioned, the need to efficiently compute mathematical objects has been central to mathematicians and scientists throughout history, and, of course, the earliest subject is arithmetic. Perhaps the most radical demonstration is the place value system we use to represent integers, which is in place for millenia precisely due to the fact that it supports extremely efficient manipulation of arithmetic operations. The next computational challenge in arithmetic, since antiquity, was accessing the multiplicative structure of integers represented this way.

Here is an except from C. F. Gauss' appeal[1] to the mathematics community of his time (in article 329 of *Disquisitiones Arithmeticae* (1801)), regarding the computational complexity of *testing primality* and *integer factorization*. The importance Gauss assigns to this computational challenge, his frustration of the state-of-the-art, and his imploring the mathematical community to resolve it shine through!

The problem of distinguishing prime numbers from composite numbers, and of resolving the latter into their prime factors is known to be one of the most important and useful in arithmetic. It has engaged the industry and wisdom of ancient and modern geometers to such an extent that it would be superfluous to discuss the problem at length. Nevertheless, we must confess that all methods that have been proposed thus far are either restricted to very special cases or are so laborious and difficult that even for numbers that do not exceed the limits of tables constructed by estimable men, they try the patience of even the practiced calculator. And these methods do not apply at all to larger numbers … the dignity of the science itself seems to require that every possible means be explored for the solution of a problem so elegant and so celebrated.

We briefly recount the state-of-the-art of these two basic algorithmic problems in number theory. A remarkable response to Gauss' first question, *efficiently deciding primality*, was found in 2002 by Agrawal, Kayal, and Saxena [8]. The use of symbolic polynomials for this problem is completely novel. Here is their elegant characterization of prime numbers.

Theorem 2.1 ([8]). *An integer $N \geq 2$ is prime if and only if*

1 Which is, of course, in Latin. I copied this English translation from a wonderful survey of Granville [82] on the subject matter of this section.

- *N is not a perfect power,*
- *N does not have any prime factor $\leq (\log N)^4$,*
- *For every $r, a < (\log N)^4$, we have the following equivalence of polynomials over $\mathbb{Z}_N[X]$:*

$$(X + a)^N \equiv X^N + a \bmod (X^r - 1).$$

It is not hard to see that this characterization gives rise to a simple algorithm for testing primality that is deterministic, and runs in time that is *polynomial* in the binary description length of N. Previous deterministic algorithms either assumed the generalize Riemann hypothesis [133] or required slightly superpolynomial time [5]. The AKS deterministic algorithm came after a sequence of efficient *probabilistic* algorithms [4,80,150,163], some elementary and some requiring sophisticated use and development of number-theoretic techniques. These probabilistic and deterministic algorithms were partly motivated by and are important to the field of cryptography.

What is not so well-known, even for those who did read the beautiful, ingenious proof in [8], is that AKS developed their deterministic algorithm by carefully "derandomizing" a previous probabilistic algorithm for primality of [7] (which uses polynomials). We note that *derandomization*, the conversion of probabilistic algorithms into deterministic ones, is by now a major area in computational complexity with a rich theory, and many other similar successes as well as challenges. The stunning possibility that *every* efficient probabilistic algorithm has a deterministic counterpart is one of the major problems of computational complexity, and there is strong evidence supporting it (see [94]). Much more on this can be found in the randomness chapters of [176].

Gauss' second challenge, of whether efficiently factoring integers is possible, remains open. But this very challenge has enriched computer science, both practical and theoretical, in several major ways. Indeed, the assumed hardness of factoring is the main guarantee of security in almost all cryptographic and e-commerce systems around the world (showing that difficult problems can be useful!). More generally, cryptography is an avid consumer of number theoretic notions, including elliptic curves, Weil pairings, and more, which are critical to a variety of cryptographic primitives and applications. These developments shatter Hardy's view of number theory as a completely useless intellectual endeavor.

There are several problems on integers whose natural definitions depend on factorization, but can, nevertheless, be solved efficiently, bypassing the seeming need to factor. Perhaps the earliest algorithm ever formally described is Euclid's algorithm for computing the GCD (greatest common divisor) of two given integers[2] m and n. Another famous such algorithm is for computing the Legendre–Jacobi symbol $(\frac{m}{n})$ via Gauss' law of quadratic reciprocity.

A fast algorithm for factoring may come out of left-field with the new development of quantum computing, the study of computers based on quantum-mechanical principles,

2 It extends to polynomials, and allows for an efficient way of computing multiplicative inverses in quotient rings of $\mathbb{Z}$ and $\mathbb{F}[x]$.

which we discussed in the quantum chapter of the book [176]. Shor has shown in [158] that such computers are capable of factoring integers in polynomial time. This result led governments, companies, and academia to invest billions in developing technologies which will enable building large-scale quantum computers, and the jury is still out on the feasibility of this project. There is no known theoretical impediment for doing so, but one possible reason for failure of this project is the existence of yet-undiscovered principles of quantum mechanics.

Other central computational problems include solving polynomial equations in finite fields, for which one of the earliest efficient (probabilistic) algorithms was developed by Berlekamp [32] (it remains a great challenge to derandomize this algorithm!). Many other examples can be found in the Algorithmic Number Theory book [26].

3. COMBINATORIAL GEOMETRY

What is the smallest area of a planar region which contains a unit length segment in *every* direction? This is the Kakeya needle problem (and such sets are called *Kakeya sets*), which was solved surprisingly by Besicovich [33] who showed that this area can be arbitrarily close to zero! Slight variation on his method produces a Kakeya set of Lebesque measure zero. It makes sense to replace "area" (namely, Lesbegue measure) by the more robust measures, such as the Hausdorff and Minkowski dimensions. This changes the picture: Davies [48] proved that a Kakeya set in the plane must have full dimension ($= 2$) in both measures, despite being so sparse in Lebesgue measure.

It is natural to extend this problem to higher dimensions. However, obtaining analogous results (namely, that the Hausdorff and Minkowski dimensions are full) turns out to be extremely difficult. Despite the seemingly recreational flavor, this problem has significant importance in a number of mathematical areas (Fourier analysis, wave equations, analytic number theory, and randomness extraction), and has been attacked through a considerable diversity of mathematical ideas (see [169]).

The following finite field analogue of the above Euclidean problem was suggested by Wolff [177]. Let $\mathbb{F}$ denote a finite field of size q. A set $K \subseteq \mathbb{F}^n$ is called Kakeya if it contains a line in every direction. More precisely, for every direction $b \in \mathbb{F}^n$ there is a point $a \in \mathbb{F}^n$ such that the line $\{a + bt : t \in \mathbb{F}\}$ is contained in K. As above, we would like to show that any such K must be large (think of the dimension n as a large constant, and the field size q as going to infinity).

Conjecture 3.1. Let $K \subseteq \mathbb{F}^n$ be a Kakeya set. Then $|K| \geq C_n q^n$, where C_n is a constant depending only on the dimension n.

The best exponent of q in such a lower bound intuitively corresponds to the Hausdorff and Minkowski dimensions in the Euclidean setting. Using sophisticated techniques from arithmetic combinatorics, Bourgain, Tao, and others improved the trivial bound of $n/2$ to about $4n/7$.

Curiously, the exact same conjecture arose, completely independently, within ToC, from the work [122] on *randomness extractors*, an area which studies the "purification" of "weak random sources" (see, e.g., the survey [171] on this important notion). In [122] Wolff's conjecture takes a probabilistic form, asking about the (min)-entropy of a random point on a random line in a Kakeya set. With this motivation, Dvir [61] brilliantly proved the Wolff conjecture (sometimes called the Finite Field Kakeya conjecture), using the (algebraic-geometric) "polynomial method" (which is inspired by techniques in decoding algebraic error-correcting codes). Many other applications of this technique to other geometric problems quickly followed, including the Guth–Katz [87] resolution of the famous Erdős distance problem, as well as for optimal randomness extraction and more (some are listed in Dvir's survey [62]).

Subsequent work determined the exact value of the constant C_n above (up to a factor of 2) [63].

Theorem 3.2 ([63]). *Let $K \subseteq \mathbb{F}^n$ be a Kakeya set. Then $|K| \geq (q/2)^n$. On the other hand, there exist Kakeya sets of size $\leq 2 \cdot (q/2)^n$.*

Many other problems regarding incidences of points and lines (and higher-dimensional geometric objects) have been the source of much activity and collaboration between geometers, algebraists, combinatorialists, and computer scientists. The motivation for these questions in the computer science side come from various sources, e.g., problems on local correction of errors [27] and derandomization [64, 105]. Other incidence theorems, e.g., Szemerédi–Trotter [168] and its finite field version of Bourgain–Katz–Tao [37] have been used, e.g., in randomness extraction [28] and compressed sensing [85].

4. OPERATOR THEORY

The following basic mathematical problem of Kadison and Singer from 1959 [102] was intended to formalize a basic question of Dirac concerning the "universality" of measurements in quantum mechanics. We need a few definitions. Consider $B(\mathcal{H})$, the algebra of continuous linear operators on a Hilbert space $\mathcal{H}$. Define a *state* to be a linear functional f on $B(\mathcal{H})$, normalized to $f(I) = 1$, which takes nonnegative values on positive semidefinite operators. The states form a convex set, and a state is called *pure* if it is not a convex combination of other states. Finally, let D be the subalgebra of $B(\mathcal{H})$ consisting of all *diagonal* operators (after fixing some basis).

Kadison and Singer asked if every pure state on D has a *unique* extension to $B(\mathcal{H})$. This problem on infinite-dimensional operators found a host of equivalent formulations in finite dimensions, with motivations and intuitions from operator theory, discrepancy theory, Banach space theory, signal processing, and probability. All of them were solved affirmatively in recent work of Marcus, Spielman, and Srivastava [127] (which also surveys the many related conjectures). Here is one statement they prove, which implies the others.

Theorem 4.1 ([127]). *For every $\varepsilon > 0$, there is an integer $k = k(\varepsilon)$ so that the following holds. Fix any n and any $n \times n$ matrix A with zeros on the diagonal and of spectral norm 1. Then there is a partition of $\{1, 2, \ldots, n\}$ into k subsets, $S_1, S_2, \ldots, S_k$, so that each of the principal minors A_i (namely A restricted to rows and columns in S_i) has spectral norm at most ε.*

This statement clearly implies that one of the minors has linear size, at least n/k. This consequence is known as the *Restricted Invertibility* Theorem of Bourgain and Tzafriri [38], itself an important result in operator theory.

How did computer scientists get interested in this problem? Without getting into too many details, here is a sketchy description of the meandering path which led to this spectacular result.

A central computational problem, at the heart of numerous applications, is solving a linear system of equations. While Gaussian elimination does the job quite efficiently (the number of arithmetic operations is about n^3 for $n \times n$ matrices), for large n this is still inefficient. Thus faster methods are sought, hopefully nearly linear in the number of nonzero entries of the given matrix. For *Laplacian*[3] linear systems (arising in many graph theory applications, such as computing electrical flows and random walks), Spielman and Teng [165] achieved precisely that! A major notion they introduced was *spectral sparsifiers* of matrices (or equivalently, weighted graphs).

A sparsifier of a given matrix is another matrix, with far fewer (indeed, linear) nonzero entries, which, nevertheless, has essentially the same (normalized) spectrum as the original (it is not even obvious that such a sparse matrix exists). We note that a very special case of sparsifiers of complete graphs are by definition *expander graphs*[4] (see much more about this central concept of expanders in [93, 176]). The algorithmic applications led to a quest for optimal constructions of sparsifiers for arbitrary Laplacian matrices (in terms of trade-off between sparsity and approximation), and these were beautifully achieved in [29] (who also provided a deterministic polynomial time algorithm to construct such sparsifiers). This in turn has led [164] to a new proof, with better analysis, of the Restricted Invertibility theorem mentioned above, making the connection to the Kadison–Singer problem.

However, the solution to Kadison–Singer seemed to require another detour. The same team [126] first resolved a bold conjecture of Bilu and Linial [34] on the spectrum of "signings" of matrices.[5] This conjecture was part of a plan for a *simple*, iterative construction of Ramanujan graphs, the best[6] possible expander graphs. Ramanujan graphs were introduced and constructed in [124, 128], but rely on deep results in number theory and alge-

---

3 Simply, symmetric PSD matrices with zero row sum.

4 All nontrivial eigenvalues of the complete graph (or constant matrix) are 0, and an expander is a sparse graph in which all nontrivial eigenvalues are tiny.

5 Simply, this beautiful conjecture states that for *every* d-regular graph, there exist $\{-1, 1\}$ signs of the edges which make all eigenvalues of the resulting signed adjacency matrix lie in the "Ramanujan interval" $[-2\sqrt{d-1}, 2\sqrt{d-1}]$.

6 With respect to the spectral gap. This is one of a few important expansion parameters to optimize.

braic geometry (believed by some to be essential for *any* such construction). Bilu and Linial sought instead an elementary construction, and made progress on their conjecture, showing how their iterative approach gives yet another way to construct "close to" Ramanujan expanders.

To prove the Bilu–Linial conjecture (and indeed produce bipartite Ramanujan graphs of every possible degree—something the algebraic constructions could not provide), [126] developed a theory of *interlacing polynomials* that turned out to be the key technical tool for resolving Kadison–Singer in [127]. In both cases, the novel view is to think of these conjectures probabilistically, and analyze the norm of a random operator by analyzing the average characteristic polynomial. That this method makes sense and actually works is deep and mysterious. Moreover, it provides a new kind of existence proofs for which no efficient algorithm (even probabilistic) of finding the desired objects is known. The analysis makes heavy use of the theory of *real stable* polynomials, and the inductive process underlying it is reminiscent (and inspired by) Gurvits' [86] remarkable proof of the van der Waerden conjecture and its generalizations.[7]

5. METRIC GEOMETRY

How close one metric space is to another is captured by the notion of *distortion*, measuring how distorted distances of one become when embedded into the other. More precisely,

Definition 5.1. Let (X, d) and (X', d') be two metric spaces. An embedding $f : X \to X'$ has distortion $\leq c$ if for every pair of points $x, y \in X$ we have

$$d(x, y) \leq d'\big(f(x), f(y)\big) \leq c \cdot d(x, y).$$

When X is finite and of size n, we allow $c = c(n)$ to depend on n.

Understanding the best embeddings between various metric and normed spaces has been a long endeavor in Banach space theory and metric geometry. An example of one major result in this area is Bourgain's embedding theorem [36].

Theorem 5.2 ([36]). *Every metric space of size n can be embedded into Euclidean space L_2 with distortion $O(\log n)$.*

The first connection between these structural questions and computational complexity was made in the important paper of Linial, London, and Rabinovich [120]. They asked for efficient algorithms for actually finding embeddings of low distortion, and noticed that for some such problems it is natural to use semidefinite programming. They applied

7 This is yet another example of a structural result (on doubly stochastic matrices) whose proof was partly motivated by algorithmic ideas. The connection is the use of hyperbolic polynomials in optimization (more specifically, as barrier functions in interior point methods.

this geometric connection to get old and new results for algorithmic problems on graphs (in particular, the sparsest cut problem we will soon discuss. Another motivation they discuss (which quickly developed into a major direction in approximation algorithms) is that some computations (e.g., finding nearest neighbors) are more efficient in some spaces than others, and so *efficient*, low-distortion embedding may provide useful reductions from harder to easier space. They describe such an efficient algorithm implementing Bourgain's Theorem 5.2 above, and also prove that his bound is best possible (the metric proving it is simply the distance between points in any constant-degree *expander* graph[8]).

The next shift in the evolution of this field, and in the level of interactions between geometers and ToC researchers, came from trying to prove "hardness of approximation" results. One example is the Goemans–Linial conjecture [78, 119], studying the sparsest cut problem, about the relation between L_1 and the "negative type" metric space L_2^2 (a general class of metrics which arise naturally in several contexts). Roughly, these are metrics on $\mathbb{R}^n$ in which Euclidean distances are squared. More precisely, a metric (X, d) is of negative type (namely, in L_2^2), if $(X, \sqrt{d})$, is isometric (has no distortion) to a subset of L_2.

Conjecture 5.3. Every L_2^2 metric can be embedded into L_1 with constant distortion.

This conjecture was proved false by Khot and Vishnoi [110]:

Theorem 5.4 ([110]). *For every n, there are n-point subsets of L_2^2 for which every embedding to L_1 requires distortion $\Omega(\log\log n)^{1/6}$.*

Far more interesting than the result itself is its origin. Khot and Vishnoi were trying to prove that the (weighted) "sparsest cut" problem is hard to approximate. They managed to do so under a computational assumption, known as the *Unique Games* conjecture of Khot [107] via a so-called *PCP*-reduction (see also [108, 176]). The elimination of this computational assumption is the magical part that demonstrates the power and versatility of reductions between computational problems. They apply their PCP reduction to a *particular*, carefully chosen unique games instance, which cannot be well approximated by a certain semidefinite program. The outcome was an instance of the sparsest cut problem which the same reduction ensures is hard to approximate by a semidefinite program. As discussed above, that outcome instance could be understood as a metric space, and the hardness of approximation translates to the required distortion bound!

The exact distortion of embedding L_2^2 into L_1 has been determined precisely to be $\sqrt{\log n}$ (up to lower order factors) in two beautiful sequences of works developing new algorithmic and geometric tools; we mention only the final word for each, as these papers contain a detailed history. On the upper bound side, the efficient algorithm approximating nonuniform sparsest cut to a factor $\sqrt{\log n}\log\log n$, which yields the same distortion bound, was obtained by Arora, Lee, and Naor [20] via a combination of the so-called "chaining argument" of [21] and the "measured descent" embedding method of [112]. A lower bound

8 The presence of such graphs in different sections illustrate how fundamental they are in diverse mathematical areas, and the same holds for algorithms and complexity theory.

of $\sqrt{\log n}$ on the distortion was very recently proved by Naor and Young [144] using a new isoperimetric inequality on the Heisenberg group.

Another powerful connection between such questions and ToC is through (again) expander graphs. A basic example is that the graph metric of any constant-degree expander proves that Bourgain's embedding theorem above is optimal! Much more sophisticated examples arise from trying to understand (and perhaps disprove) the Novikov and the Baum–Connes conjectures (see [104]). This program relies on another, much weaker notion of *coarse* embedding.

Definition 5.5. (X, d) has a coarse embedding into (X', d') if there is a map $f : X \to X'$ and two increasing, unbounded real functions α, β such that for every two points $x, y \in X$,

$$\alpha\big(d(x,y)\big) \le d'\big(f(x), f(y)\big) \le \beta\big(d(x,y)\big).$$

Gromov [83] was the first to construct a metric (the word metric of a group) which cannot be coarsely embedded into a Hilbert space. His construction uses an infinite family of *Cayley* expanders (graphs defined by groups). This result was greatly generalized by Lafforgue [114] and Mendel–Naor [130], who constructed graph metrics that cannot be coarsely embedded into any *uniformly convex* space. It is interesting that while Lafforgue's method is algebraic, the Mendel–Naor construction follows the combinatorial *zigzag* construction of expanders [155] from computational complexity.

Many other interaction projects regarding metric embeddings and distortion we did not touch on include their use in numerous algorithmic and data structure problems like clustering, distance oracles the k-server problem, as well as the fundamental interplay between distortion and *dimension reduction* relevant to both geometry and CS, where so many basic problems are open.

6. GROUP THEORY

Group theorists, much like number theorists, have been intrinsically interested in computational problems since the origin of the field. For example, the *word problem* (given a word in the generators of some group, does it evaluate to the trivial element?) is so fundamental to understanding any group one studies, that as soon as language was created to formally discuss the computational complexity of this problem, hosts of results followed trying to pinpoint that complexity. These include decidability and undecidability results once Turing set up the theory of computation and provided the first undecidable problems, and these were followed with $\mathcal{NP}$-completeness results and efficient algorithms once $\mathcal{P}$ and $\mathcal{NP}$ were introduced around 1970. Needless to say, these *algorithmic results* inform of *structural* complexity of the groups at hand. And the word problem is but the first example. Another demonstration is the beautiful interplay between algorithmic and structural advances over decades, on the *graph isomorphism problem*, recently leading to breakthrough of Babai [24]! A huge body of work is devoted to finding efficient algorithms for computing commutator subgroups, Sylow subgroups, centralizers, bases, representations, characters, and a host

of other important substructures of a group from some natural description of it. Excellent textbooks include [92, 157].

Here we focus on two related problems, the *generation* and *random generation* problems, and new conceptual notions borrowed from computational complexity which are essential for studying them. Before defining them formally (below), let us consider an example. Assume I hand you 10 invertible matrices, say 100×100 in size, over the field of size 3. Can you tell me if they generate another such given matrix? Can you even produce convincing evidence of this before we both perish? How about generating a random matrix in the subgroup spanned by these generators? The problem, of course, is that this subgroup will have size far larger than the number of atoms in the known universe, so its elements cannot be listed, and typical words generating elements in the group may need to be prohibitively long. Indeed, even the extremely special cases, for elements in $\mathbb{Z}_p^*$ (namely one, 1×1 matrix), the first question is related to the *discrete logarithm* problem, and for $\mathbb{Z}_{p \cdot q}^*$ it is related to the *integer factoring* problem, both currently requiring exponential time to solve (as a function of the description length).

Let us consider any finite group G and let $n \approx \log |G|$ be roughly the length of a description of an element of G. Assume we are given k elements in G, $S = \{s_1, s_2, \ldots, s_k\}$. It would be ideal if the procedures we describe would work in time polynomial in n and k (which prohibits enumerating the elements of G, whose size is exponential in n).

The *generation problem* asks if a given element $g \in G$ is generated by S. How does one prove such a fact? A standard certificate for a positive answer is a *word* in the elements of S (and their inverses) which evaluates to g. However, even if G is cyclic, the shortest such word may be exponential in n. An alternative, computationally motivated description, is to give a *program* for g. Its definition shows that the term "program" suits it perfectly, as it has the same structure as usual computer programs, only that instead of applying some standard Boolean or arithmetic operations, we use the group operations of multiplication and inverse.

Definition 6.1. A *program* (over S) is a finite sequence of elements $g_1, g_2, \ldots, g_m$, where every element g_i is either in S, or is the inverse of a previous g_j, or is the product of previous g_j, g_ℓ. We say that it computes g simply if $g = g_m$.

In the cyclic case, programs afford exponential savings over words in description length, as a program allows us to write large powers by repeatedly squaring elements. What is remarkable is that such savings are possible for *every* group. This discovery of Babai and Szemerédi [25] says that every element of every group has an extremely succinct description in terms of any set of elements generating it.

Theorem 6.2 ([25]). *For every group G, if a subset of elements S generates another element g, then there is a program of length at most $n^2 \approx (\log |G|)^2$ which computes g from S.*

It is interesting to note that the proof uses a structure which is very combinatorial and counterintuitive for group theorists, namely that of a *cube*, which we will see again later. For a sequence $(h_1, h_2, \ldots, h_t)$ of elements from G, the cube $C(h_1, h_2, \ldots, h_t)$ is the (multi)set of 2^t elements $\{h_1^{\varepsilon_1}, h_2^{\varepsilon_2}, \ldots, h_t^{\varepsilon_t}\}$, with $\varepsilon_i \in \{0, 1\}$. Another important feature of the proof

is that it works in a very general setting of "black-box" groups—it never needs an explicit description of the host group, only the ability to multiply elements and take their inverses. This is a very important paradigm for arguing about groups, and will be used again below.

How does one prove that an element g is *not* generated by S? It is possible that there is no short "classical" proof! This question motivated Babai to define Arthur–Merlin games—a new notion of probabilistic, interactive proofs (simultaneously with Goldwasser, Micali, and Rackoff [81], who proposed a similar notion for cryptographic reasons), and showed how nonmembership can be certified in this new framework. The impact of the definition of interactive proofs on the theory of computation has been immense, and is discussed in, e.g., in the books [19, 79, 176].

Returning to the generation problem, let us now consider the problem of *random generation*. Here we are given S, and would like a randomized procedure which will quickly output an (almost) uniform distribution on the subgroup H of G generated by S. This problem, besides its natural appeal, is often faced by computational group theorists, being a subroutine in many group-theoretic algorithms. In practice often heuristics are used, like the famous "product replacement algorithm" and its variants, which often work well in practice (see, e.g., the recent [22] and references). We will discuss here provable bounds.

It is clear that sufficiently long random words in the elements of S and its inverses will do the job, but just as with certificates, sufficiently long is often prohibitively long. In a beautiful paper, Babai [23] describes a certain process generating a random program which computes a nearly-uniform element of H, and runs in time $n^5 \approx (\log |G|)^5$ steps. It again uses cubes, and works in the full generality of black-box groups. This paper was followed by even faster algorithms with simpler analysis by Cooperman and by Dixon [45, 58], and the state-of-the-art is an algorithm whose number of steps is remarkably the same as the length of proofs of generation above—in other words, randomness roughly achieves the efficiency of nondeterminism for this problem. Summarizing:

Theorem 6.3 ([23, 45, 58]). *For every group G, there is a* probabilistic *program of length* $\mathrm{poly}(n) \approx \mathrm{poly}(\log |G|)$ *that, given* any *generating set S for G, produces with high probability a (nearly) uniformly random element of G.*

7. STATISTICAL PHYSICS

The field of statistical physics is huge, and we focus here mainly on connections of statistical mechanics with the theory of computation. Numerous mathematical models exist of various physical and chemical systems, designed to understand basic properties of different materials and the dynamics of basic processes. These include such familiar models as Ising, Potts, monomer–dimer, spin-glass, percolation, etc. A typical example explaining the connection of such mathematical models to physics and chemistry, and the basic problems studied is the seminal paper of Heilmann and Lieb [90].

Many of the problems studied can be viewed in the following general setting. We have a huge (exponential) space of objects called Ω (these objects may be viewed as the

different configurations of a system). Each object is assigned a nonnegative weight (which may be viewed as the "energy" of that state). Scaling these weights gives rise to a probability distribution (often called the Gibbs distribution) on Ω, and to study its properties (phase transitions, critical temperatures, free energy, etc.) one attempts to generate samples from this distribution. Note that if the description of a state takes n bits, then brute-force listing of all probabilities in question is exponentially prohibitive. Thus efficiency of the sampling procedure is essential to this study.

As Ω may be highly unstructured, the most common approach to this sampling problem is known as "Monte Carlo Markov Chain" (or "MCMC") method. The idea is to build a graph on the objects of Ω, with a pair of objects connected by an edge if they are similar in some sense (e.g., sequences which differ only in a few coordinates). Next, one starts from any object, and performs a biased random walk on this graph for some time, and the object reached is the sample produced. In many settings it is not hard to set up the random walk (often called Glauber dynamics or the Metropolis algorithm) so that the *limiting* distribution of the Markov chain is indeed the desired distribution. The main question in this approach is *when* to stop the walk and output a sample; *when* are we close enough to the limit? In other words, how long does it take the chain to converge to the limit? In most cases, these decisions were taken on intuitive, heuristic grounds, without rigorous analysis of convergence time. The exceptions where rigorous bounds were known were typically structured, e.g., where the chain was a Cayley graph of a group (e.g., [11, 56]).

This state of affairs has changed considerably since the interaction in the past couple of decades with the theory of computation. Before describing it, let us see where computational problems even arise in this field. The two major sources are *optimization* and *counting*. That the setting above suits many instances of optimization problems is easy to see. Think of Ω as the set of solutions to a given optimization problem (e.g., the values of certain parameters designed to satisfy a set of constraints), and the weights representing the quality of a solution (e.g., the number of constraints satisfied). So, picking at random from the associated distribution favors high-quality solutions. The counting connection is more subtle. Here Ω represents a set of combinatorial objects one wants to count or approximate (e.g., the set of perfect matchings in a graph, or satisfying assignments to a set of constraints). It turns out that for very general situations of this type, sampling an object (approximately) at random is tightly connected to counting their number; it often allows a recursive procedure to approximate the size of the set [99]. An additional observation is that viewing a finite set as a fine discretization of a continuous object (e.g., fine lattice points in a convex set) allows one to compute volumes and more generally integrate functions over such domains.

Around 1990, rigorous techniques were introduced [12, 39, 65, 161] to analyze the convergence rates of such general Markov chains arising from different approximation algorithms. They establish *conductance* bounds on the Markov chains, mainly via *canonical paths* or *coupling* arguments (a survey of this early work is [96]). Collaborative work was soon able to formally justify the physical intuition behind some of the suggested heuristics for many models, and, moreover, drew physicists to suggest such ingenious chains for optimization problems. The field drew in probabilists and geometers as well, and by now is

highly active and diverse. We mention two results to illustrate rigorous convergence bounds for important problems of this type.

Theorem 7.1 ([97]). *The permanent of any nonnegative $n \times n$ matrix can be approximated, to any multiplicative factor $(1+\varepsilon)$, in polynomial time in n/ε.*

The importance of this approximation algorithm stems from the seminal result of Valiant [173] about the permanent polynomial (that notorious sibling of the determinant polynomial, that looks identical except that the permanent has no signs; for more see [159, 176]). Valiant proved that the permanent is *universal*, capturing (via efficient reductions) essentially all natural counting problems, including those arising in the statistical physics models and optimization and counting problems above. So, unlike determinant, computing the permanent *exactly* is extremely difficult (harder than $\mathcal{NP}$-complete).

Theorem 7.2 ([65]). *The volume of any convex set in n dimensions can be approximated, to any multiplicative factor $(1+\varepsilon)$, in polynomial time in n/ε.*

The volume, besides its intrinsic interest, captures as well natural counting problems, e.g., the number of linear extensions of a given partially ordered set. The analysis of this algorithm, as well as its many subsequent improvements, has used and developed purely structural results of independent interest in differential and convex geometry. It also led to generalizations, like efficiently sampling from any log-concave distribution (see the survey [174]).

Another consequence of this collaboration was a deeper understanding of the relation between *spacial* properties (such as phase transitions, and long-range correlations between distant sites in the Gibbs distribution) and *temporal* properties (such as speed of convergence of the sampling or approximately counting algorithms, like Glauber dynamics). This connection (surveyed, e.g., in [66]) was established by physicists for spin systems since the 1970s. The breakthrough work of Weitz [175] on the *hard core* model gave an *deterministic* algorithm which is efficient up to the phase transition, and this was complemented by a hardness result of Sly [162] beyond the phase transition. These phase transition of computational complexity, at the same point as the phase transition of the Gibbs distribution are striking, and the generality of this phenomenon is still investigated.

More generally, the close similarity between statistical physics models and optimization problems, especially on random instances, is benefitting both sides. Let us mention a few exciting developments. It has unraveled the fine geometric structure of the space of solutions at the phase transition, pinpointing it, e.g., for k-SAT in [1]. At the same time, physics intuition based on such ideas as renormalization, annealing, and replica symmetry breaking has led to new algorithms for optimization problems, some of them now rigorously analyzed, e.g., as in [98]. Others, like one of the fastest (yet unproven) heuristics for such problems as Boolean Satisfiability (which is $\mathcal{NP}$-complete in general) are based on the physics method of "survey propagation" of [131]. Finally, new algorithmic techniques for similar physics and optimization problems, originate from an unexpected source, the *Lovasz Local Lemma* (LLL). The LLL is a probabilistic proof technique for the existence rare events in a proba-

bility space. Its efficient versions, formulating it algorithmically as a *directed, nonreversible* Markov chains, starting with the works of Moser [136,137], have led to approximate counting and sampling versions for such events (see, e.g., [84]). A completely different, *deterministic* algorithm of Moitra [135] for the LLL regime (of rare events) promises many more applications: it works even when the solution space (and hence the natural Markov chain) is not connected!

We conclude this story with the recent breakthrough connection between high-dimensional expanders and the analysis of MCMC of Anari, Liu, Gharan, and Vinzant [18]. The theory of high-dimensional expanders (generalizing that of expander graphs to higher dimensional complexes – see the survey [123]), which easily merits a separate vignette, has been rapidly developing in the past decade within combinatorics and complexity theory, following deep roots in the theory of Bruhat–Tits buildings, connections with several areas of math, and new applications. Anari et al. realized that the local-to-global principle underlying high-dimensional expansion can be used for an inductive analysis of the convergence rate of many families of Markov chain-based algorithms. Their first application resolves a 30-year old conjecture, proving

Theorem 7.3 ([18]). *The number of bases of any matroid on n elements can be approximated, to any multiplicative factor* $(1+\varepsilon)$*, in polynomial time in* n/ε.

The revolutionary impact and future potential of this connection (and further ideas), in just a couple of years, to problems of approximate counting and random sampling in optimization and statistical physics, can be appreciated, e.g., from these papers and the references therein [16,17,121].

8. ANALYSIS AND PROBABILITY

This section gives a taste of a growing number of families of inequalities—large deviation inequalities, isoperimetric inequalities, etc.—that have been generalized beyond their classical origins due to a variety of motivations in the theory of computing and discrete mathematics. Further, the applications sometimes call for *stability* versions of these inequalities, namely an understanding of the structures which make an inequality nearly sharp. Here too these motivations pushed for generalizations of classical results and many new ones. Most of the material below, and much more on the motivations, applications, and developments in this exciting area of the analysis of Boolean functions, can be found in the book [145] by O'Donnell.

The following story can be told from several angles. One is the *noise sensitivity* of functions. We restrict ourselves to the Boolean cube endowed with the uniform probability measure, but many of the questions and results extend to arbitrary product probability spaces. Let $f : \{-1,1\}^n \to \mathbb{R}$, which we assume is balanced, namely $E[f] = 0$. When the image of f is $\{-1,1\}$, we can think of f as a voting scheme, translating the binary votes of n individuals into a binary outcome. One natural desire from such a voting scheme may be *noise stability*—that typically very similar inputs (vote vectors) will yield the same outcome.

While natural in this social science setting, such questions also arise in statistical physics settings, where natural functions such as bond percolation turn out to be extremely sensitive to noise [31]. Let us formally define noise stability.

Definition 8.1. Let $\rho \in [0,1]$ be a correlation parameter. We say two vectors $x, y \in \{-1,1\}^n$ are ρ-correlated if they are distributed as follows. The vector x is drawn uniformly at random, and y is obtained from x by flipping each bit x_i independently with probability $(1-\rho)/2$. Note that for every i the correlation $E[x_i y_i] = \rho$. The *noise sensitivity* of f at ρ, $S_\rho(f)$, is simply defined as the correlation of the outputs, $E[f(x)f(y)]$.

It is not hard to see that the function maximizing noise stability is any *dictatorship* function, e.g., $f(x) = x_1$, for which $S_\rho(f) = \rho$. But another natural social scientific concern is the *influence* of players in voting schemes [30], which prohibits such solutions (in democratic environments). The influence of a single voter[9] is the probability with which it can change the outcome given that all other votes are uniformly random (so, in a dictatorship it is 1 for the dictator and 0 for all others). A fair voting scheme should have no voter with high influence. As we define influence for real-valued functions, we will use the (conditional) *variance* to measure a player's potential effect given all other (random) votes.

Definition 8.2. A function $f : \{-1,1\}^n \to \mathbb{R}$ has influence τ if for every i, $\mathrm{Var}[x_i | x_{-i}] \le \tau$ for all i (where x_{-i} denotes the vector x without the ith coordinate).

For example, the majority function has influence $O(1/\sqrt{n})$. The question of how small the influence of a balanced function can be is extremely interesting, and leads to a highly relevant inequality for our story (both in content and techniques). As it turns out, ultimate fairness (influence $1/n$ per player) is impossible— the authors of [103] show that every function has a player with nonproportional influence, at least $\Omega(\log n/n)$. At any rate, one can ask which of the functions with *small* influence is most stable, and it is natural to guess that majority should be the best.[10]

The conjecture that this is the case, called the *Majority is Stablest* conjecture, arose from a completely different and surprising angle—the field of optimization, specifically "hardness of approximation." A remarkable paper [109] has shown that this conjecture implies[11] the optimality of a certain natural algorithm for approximating the *maximum cut* of a graph (i.e., the partition of vertices that maximizes the number of edges between them).[12] This connection is highly nontrivial, but by now we have many examples showing how the analysis of certain (semidefinite programming-based) approximation algorithms for a vari-

---

9 This seminal paper [30] also studies the influences of coalitions of players, extremely natural in game theory, which arises in and contributes to other areas of computational complexity (including circuit complexity, learning and pseudorandomness), and raises other analytic questions which we will not discuss here.

10 This noise sensitivity tends, as n grows, to $S_\rho(\mathrm{Majority}_n) = \frac{2}{\pi}\arcsin\rho$.

11 Assuming another, complexity-theoretic, conjecture called the "Unique Games" conjecture of [107] (discussed already in the metric geometry section above; see also [108,176]).

12 Maximum Cut is a basic optimization problem whose exact complexity is $\mathcal{NP}$-complete.

ety of optimization problems raise many new isoperimetric questions,[13] greatly enriching this field.

The Majority is Stablest conjecture was proved in a strong form by [138] shortly after it was posed. Here is a formal statement (which actually works for bounded functions).

Theorem 8.3 ([138]). *For every (positive correlation parameter) $\rho \geq 0$ and $\varepsilon > 0$, there exists (an influence bound) $\tau = \tau(\rho, \varepsilon)$ such that for every n and every $f : \{-1, 1\}^n \to [-1, 1]$ of influence at most τ, $S_\rho(f) \leq S_\rho(\text{Majority}_n) + \varepsilon$.*

The proof reveals another angle on the story—large deviation inequalities and invariance principles. To see the connection, recall the Berry–Esseen theorem [70], generalizing the standard central limit theorem to *weighted* sums of independent random signs. In this theorem, influences arise very naturally. Consider $\sum_{i=1}^n c_i x_i$. If we normalize the weights c_i to satisfy $\sum_i c_i^2 = 1$, then c_i is the influence of the ith voter, and $\tau = \max_i |c_i|$. The quality of this central limit theorem deteriorates linearly with the influence τ. Lindeberg's proof of Berry–Esseen uses an invariance principle, showing that for linear functions, the cumulative probability distribution $\Pr[\sum_{i=1}^n c_i x_i \leq t]$ (for every t) is unchanged (up to τ), *regardless* of the distribution of the variables x_i, as long as they are independent and have expectation 0 and variance 1. Thus, in particular, they can be taken to be standard Gaussian, which trivializes the problem, as the weighted sum is a Gaussian as well!

To prove their theorem, [138] first observed that also in the noise stability problem, the Gaussian case is simple. If the x_i, y_i are standard Gaussians with correlation ρ, the stability problem reduces to a classical result of Borell [35]: that noise stability is maximized by any hyperplane through the origin. Note that here the rotational symmetry of multidimensional Gaussians, which also aids the proof, does not distinguish "dictator" functions from majority—both are such hyperplanes. Given this theorem, an invariance principle whose quality depends on τ would do the job. They next show that it is sufficient to prove the principle only for *low degree* multilinear polynomials (as the effect of noise decays with the degree). Finally, they prove this nonlinear extension of Berry–Esseen for such polynomials, a form of which we state below. They also use their invariance principle to prove other conjectures, and since the publication of their paper, quite a number of further generalizations and applications were found.

Theorem 8.4 ([138]). *Let x_i be any n independent random variables with mean 0, variance 1, and bounded 3rd moments. Let g_i be n independent standard Gaussians. Let Q be any degree d multilinear n-variate polynomial of influence τ. Then for any t,*

$$\left|\Pr\left[Q(x) \leq t\right] - \Pr\left[Q(g) \leq t\right]\right| \leq O(d\,\tau^{1/d}).$$

We now only seem to be switching gears… To conclude this section, let me give one more, very different demonstration of the surprising questions (and answers) regarding

13 Many over continuous domains, like the unit cube or Gaussian space, where the connection between noise stability and isoperimetry may be even clearer.

noise stability and isoperimetry, arising from the very same computational considerations of optimization of hardness of approximation. Here is the question: *What is the smallest surface area of a (volume 1) body which tiles $\mathbb{R}^d$ periodically along the integer lattice $\mathbb{Z}^d$?* Namely, we seek a d-dimensional volume 1 subset $B \subseteq \mathbb{R}^d$ such that $B + \mathbb{Z}^d = \mathbb{R}^d$, such that its boundary has minimal $(d-1)$-dimensional volume.[14] Let us denote this infimum by $s(d)$. The curious reader can stop here a bit and test your intuition, what do you expect the answer to be, asymptotically in d?

Such questions originate from the late 19th century study by Thomson (later Lord Kelvin) of *foams* in 3 dimensions [170], further studied, generalized, and applied in mathematics, physics, chemistry, material science, and even architecture. However, for this very basic question, where periodicity is defined by the simplest integer lattice, it seems that, for large d, the trivial upper and lower bounds on $s(d)$ were not improved on for over a century. The trivial upper bound on $s(d)$ is provided by the unit cube, which has surface area $2d$. The trivial lower bound on $s(d)$ comes from ignoring the tiling restriction, and considering only the volume – here the unit volume ball has the smallest surface area, $\sqrt{2\pi e d}$. Where in this quadratic range does $s(d)$ lie? In particular, can there be *"spherical cubes,"* with $s(d) = O(\sqrt{d})$?

The last question became a central issue for complexity theorists when [69] related it directly to the important Unique Games conjecture, and optimal inapproximability proofs of combinatorial problems (in particular, the maximum cut problem) discussed above. The nontrivial connection, which the paper elaborates and motivates, goes through attempts to find the tightest version of Raz' [151] celebrated parallel repetition theorem.[15] A limit on how "strong" a parallel repetition theorem can get was again provided by Raz [152]. Extending his techniques [111] to the geometric setting, resolved the question above, proving that "spherical cubes" do exist!

Theorem 8.5 ([111]). *For all d, $s(d) \leq \sqrt{4\pi d}$.*

A simple proof, and various extensions of this result were given subsequently in [14]. We note that all known proofs are probabilistic. Giving an explicit construction, that might better illustrate how a "spherical cube" (even with much worse parameters) looks like, seems a challenging problem.

9. LATTICE THEORY

Lattices in Euclidean space are among the most "universal" objects in mathematics, in that besides being natural (e.g., arising in crystalline structures) and worthy of study in their own right, they capture a variety of problems in different fields such as number theory,

14 Note that the volume of B ensures that the interiors of $B + v$ and $B + u$ are disjoint for any two distinct integer vectors $u, v \in \mathbb{Z}^d$, so this gives a tiling.

15 A fundamental information theoretic inequality of central importance to "amplification" of Probabilistically Checkable Proofs (PCPs).

analysis, approximation theory, Lie algebras, convex geometry, and more. Many of the basic results in lattice theory, as we shall see, are *existential* (namely supply no efficient means for obtaining the objects whose existence is proved), which in some cases has limited progress on these applications.

This section tells the story of one algorithm, of Lenstra, Lenstra, and Lovász [117], often called the LLL algorithm, and some of its implications on these classical applications as well as modern ones in cryptography, optimization, number theory, symbolic algebra, and more. But we had better define a lattice[16] first.

Let $B = \{b_1, b_2, \ldots, b_n\}$ be a basis of $\mathbb{R}^n$. Then the *lattice* $L(B)$ denotes the set (indeed, Abelian group) of all *integer* linear combinations of these vectors, i.e., $L(B) = \{\sum_i z_i b_i : z_i \in \mathbb{Z}\}$; B is also called a basis of the lattice. Naturally, a given lattice can have many different bases, e.g., the standard integer lattice in the plane, generated by $\{(0,1),(1,0)\}$, is equally well generated by $\{(999,1),(1000,1)\}$. A basic invariant associated with a lattice L is its determinant $d(L)$, which is the absolute value of $\det(B)$ for any basis B of L (this is also the volume of the fundamental parallelpiped of the lattice). For simplicity and without loss of generality, we will assume that B is normalized so that we only consider lattices L of $d(L) = 1$.

The most basic result about lattices, namely that they must contain *short* vectors (in any norm) was proved by Minkowski (who initiated Lattice Theory, and with it, the Geometry of Numbers) [134].

Theorem 9.1 ([134]). *Consider an arbitrary convex set K in $\mathbb{R}^n$ which is centrally symmetric*[17] *and has volume $> 2^n$. Then, every lattice L (of determinant 1) has a nonzero point in K.*

This innocent theorem, which has a simple, but *existential* (pigeonhole) proof, turns out to have numerous fundamental applications in geometry, algebra, and number theory. Among famous examples this theorem yields with appropriate choice of norms and lattices, results like Dirichlet's Diophantine approximation theorem and Lagrange's four-squares theorem, and (with much more work) the finiteness of class numbers of number fields (see, e.g., [148]).

From now on we will focus on short vectors in the (most natural) Euclidean norm. A direct corollary of Minkowski's theorem when applying it to the cube $K = [-1,1]^n$ yields:

Corollary 9.2. *Every lattice L of determinant 1 has a nonzero point of Euclidean norm at most $\sqrt{n}$.*

Digressing a bit, we note that very recently, a century after Minkowski, a strong converse of the above corollary[18] conjectured by Dadush (see [47]) for *computational* motivation has been proved in [154]. This converse has many structural consequences, on the covering

16 We only define full-rank lattices here, which suffice for this exposition.

17 Namely, $x \in K$ implies that also $-x \in K$. Such sets are precisely balls of arbitrary norms.

18 Which has to be precisely formulated.

radius of lattices, arithmetic combinatorics, Brownian motion, and others. We will not elaborate here on this new interaction of computational complexity and optimization with lattice theory and convex geometry. The papers above beautifully motivate these connections and applications, and the history of ideas and technical work needed for this complex proof.

Returning to Minkowski's corollary for the Euclidean norm, the proof is still existential, and the obvious algorithm for finding such a short vector requires exponential time in n. The breakthrough paper [117] describe the LLL algorithm, an efficient, polynomial-time algorithm, which approximates the length of the shortest vector in any n-dimensional lattice by a 2^n factor.

Theorem 9.3 ([117]). *There is a polynomial time algorithm, which, given any lattice L, produces a vector in L of Euclidean length at most 2^n factor longer than the shortest vector in L.*

This exponential bound may seem excessive at first, but the number and diversity of applications is staggering. First, in many problems, the dimension n is a small constant (so the actual input length arises from the bit-size of the given basis). This leads, for instance, to Lenstra's algorithm for (exactly solving) Integer Programming [118] in constant dimensions. It also leads to Odlyzko and Riele's refutation [146] of Mertens' conjecture about cancellations in the Möbius function, and to the long list of number-theoretic examples in [160]. But it turns out that even when n is arbitrarily large, many problems can be solved in poly(n)-time as well. Here is a list of examples of old and new problems representing this variety, some going back to the original paper [117]. In all, it suffices that real number inputs are approximated to poly(n) digits in dimension n.

- *Diophantine approximation*. While the best possible approximation of one real number by rationals with bounded denominator is readily solved by its (efficiently computable) continued fraction expansion, no such procedure is known for *simultaneous* approximation. Formally, given a *set* of real numbers, say $\{r_1, r_2, \ldots, r_n\}$, a bound Q and $\varepsilon > 0$, find integers $q \leq Q$ and $p_1, \ldots, p_n$ such that all $|r_i - p_i/q| \leq \varepsilon$. Existentially (using Minkowski), the Dirichlet "box-principle" shows that $\varepsilon < Q^{1/n}$ is possible. Using LLL, one efficiently obtains $\varepsilon < 2^{n^2} Q^{1/n}$ which is meaningful for Q described by poly(n) many bits.

- *Minimal polynomials of algebraic numbers*. Here we are given a single real number r and a degree bound n, and are asked if there is a polynomial $g(x)$ with integer coefficients, of degree at most n of which r is a root (and also to produce such a polynomial g if it exists). Indeed, this is a special case of the problem above with $r_i = r^i$. While the algorithm only outputs g for which $g(r) \approx 0$, it is often easy to check that it actually vanishes. Note that by varying n we can find the minimal such polynomial.

- *Polynomial factorization over rationals*. Here the input is an integer polynomial h of degree n, and we want to factor it over $\mathbb{Q}$. The high level idea is to first find an

(approximate) root r of h (e.g., using Newton's method), feed it to the problem above, which will return a minimal g having r as a root, and thus divides h. We stress that this algorithm produces the exact factorization, not an approximate one!

- *Small integer relations between reals*. Given reals $r_1, r_2, \ldots, r_n$, and a bound Q, determine if there exist integers $|z_i| < Q$ such that $\sum_i z_i r_i = 0$ (and if so, find these integers). As a famous example, LLL can find an integer relation among $\arctan(1) \approx 0.785398$, $\arctan(1/5) \approx 0.197395$, and $\arctan(1/239) \approx 0.004184$, yielding Machin's formula

$$\arctan(1) - 4\arctan(1/5) + \arctan(1/239) = 0.$$

- *Cryptanalysis*. Note that a very special case of the problem above (in which the coefficients z_i must be Boolean) is the "Knapsack problem," a famous $\mathcal{NP}$-complete problem. The point here is that in the early days of cryptography, some systems were based on the assumed "average case" hardness of Knapsack. Many such systems were broken by using LLL, e.g., [115]. LLL was also used to break some versions of the RSA cryptosystem (with "small public exponents").

It is perhaps a fitting epilogue to the last item that lattices cannot only destroy cryptosystems, but also create them. The problem of efficiently approximating short vectors up to polynomial (as opposed to exponential, as LLL produces) factors is believed to be computationally hard. Here are some major consequences of this assumption. First, Ajtai showed in a remarkable paper [9] that such hardness is preserved "on average", over a cleverly-chosen distribution of random lattices. This led to a new public-key encryption scheme by Ajtai and Dwork [10] based on this hardness, which is arguably the only one known that can potentially sustain quantum attacks (Shor's efficient quantum algorithms can factor integers and compute discrete logarithms [158]). In another breakthrough work of Gentry [77], this hardness assumption is used to devise *fully homomorphic* encryption, a scheme which allows not only to encrypt data, but to perform arbitrary computations directly with encrypted data. See more in this excellent survey [147].

10. INVARIANT THEORY (AND MORE)

This section is somewhat longer than the rest. One reason is that much of it has been a primary research interest of mine in recent years, and indeed is the subject of my ICM lecture. Another reason is that the connections revealed here are considerably richer. On the one hand, several different areas within the theory of computation play a role, including algebraic complexity theory, derandomization, and optimization. On the other, while invariant theory is central in these developments, connections and implications are revealed to and between other mathematical areas, including noncommutative algebra, analysis, representation theory, quantum information theory, statistics, and operator theory. We will explore some of these here.

Invariant theory, born in an 1845 paper of Cayley [43], is major branch of algebra, with important natural connections to algebraic geometry and representation theory, but also to many other areas of mathematics. We will see some here, as well as some new connections with computational complexity, leading to new questions and results in this field. We note that computational efficiency was always important in invariant theory, which is rife with ingenious algorithms (starting with Cayley's *Omega process*), as is evident from the books [46, 49, 167].

Invariants are familiar enough, from examples like the following:

- In high school physics we learn that energy and momentum are conserved (namely, are *invariants*) in the dynamics of general physical systems.
- In chemical reactions the number of atoms of each element is preserved as one mixture of molecules is transformed to yield another (e.g., as combining sodium hydroxide (NaOH) and hydrochloric acid (HCl) yields the common salt sodium chloride (NaCl) and water (H_2O)).
- In geometry, a classical puzzle asks when can a plane polygon be "cut and pasted" along straight lines to another polygon. Here the obvious invariant, *area*, is the only one![19] However, in generalizing this puzzle to 3-dimensional polyhedra, it turns out that besides the obvious invariant, *volume*, there is another invariant, discovered by Dehn.[20]

More generally, questions about the "topological equivalence" of two topological objects (e.g., knots), whether two groups are isomorphic, or whether two points are in the same orbit of a dynamical system, etc., all give rise to similar questions and treatment. A canonical way to give negative answers to such questions is through *invariants*, namely quantities preserved under some action on an underlying space.

We will focus on invariants of *linear groups* acting linearly on *vector spaces*. Let us present some notation. Fix a field $\mathbb{F}$ (while problems are interesting in every field, results mostly work for infinite fields only, and sometimes just for characteristic zero or algebraically closed ones). We will also suppress some technicalities. Let G be a group, and V a representation of G, namely an $\mathbb{F}$-vector space on which G acts: for every $g, h \in G$ and $v \in V$, we have $gv \in V$ and $g(hv) = (gh)v$.

The *orbit* under G of a vector (or point) $v \in V$, denoted Gv, is the set of all other points that v can be moved to by this action, namely $\{gv : g \in G\}$. Understanding the orbits of a group objects is a central task of this field. A basic question capturing many of the examples above is, given two points $u, v \in V$, do they lie in the same G-orbit, namely if $u \in Gv$. A related basic question, which is even more natural in algebraic geometry (when

19 And so, every two polygons of the same area can be cut to produce an *identical* (multi)sets of triangles.

20 So there are pairs of 3-dimensional polyhedra of the same volume, which cannot be cut to identical (multi)sets of tetrahedra.

the field $\mathbb{F}$ is algebraically closed of characteristic zero) is whether the *closures*[21] of the two orbits intersect, namely if some point in V can be approximated arbitrarily well by points in both Gu and Gv. We will return to specific incarnations of these questions.

When G acts on V, it also acts on $\mathbb{F}[V]$, the polynomial functions on V, also called the *coordinate ring* of V. In our setting V will have finite dimension (say m), and so $\mathbb{F}[V]$ is simply $\mathbb{F}[x_1, x_2, \ldots, x_m] = \mathbb{F}[X]$, the polynomial ring over $\mathbb{F}$ in m variables. We will denote by gp the action of a group element $g \in G$ on a polynomial $p \in \mathbb{F}[V]$.

A polynomial $p(X) \in \mathbb{F}[X]$ is *invariant* if it is unchanged by this action, namely for every $g \in G$ we have $gp = p$. All invariant polynomials clearly form a subring of $\mathbb{F}[X]$, denoted $\mathbb{F}[X]^G$, called the *ring of invariants* of this action. Understanding the invariants of group actions is the main subject of Invariant Theory. A fundamental result of Hilbert **[91]** shows that in our linear setting,[22] *all* invariant rings will be *finitely generated* as an algebra.[23] Finding the "simplest" such generating set of invariants is our main concern here.

Two familiar examples of perfect solutions to this problem follow:

- In the first, $G = S_m$, the symmetric group on m letters, is acting on the set of m formal variables X (and hence the vector space they generate) by simply permuting them. Then the ring of invariants is simply all symmetric polynomials, and a (minimal) set of generating invariants is the first m *elementary* symmetric polynomials in X.

- In the second, $G = \mathrm{SL}_n(\mathbb{F})$, the special linear group of matrices with determinant 1, is acting on the vector space $M_n(\mathbb{F})$ of $n \times n$ matrices (so $m = n^2$), simply by left matrix multiplication. In this case all polynomial invariants are generated by a single polynomial, the determinant of this m-variable matrix X.

In these two cases, which really supply a complete understanding of the invariant ring $\mathbb{F}[X]^G$, the generating sets are *good* in several senses. There are *few* generating invariants, they all have *low* degree, and they are *easy* to compute[24]—all these quantities are bounded by a polynomial in m, the dimension of the vector space.[25] In such good cases, one has efficient algorithms for the basic problems regarding orbits of group actions. For example, a fundamental duality theorem of Geometric Invariant Theory **[143]** (see Theorem A.1.1), shows how generating sets of the invariant ring can be used for the orbit closure intersection problem.

---

21 One can take closure in either the Euclidean or the Zariski topology (the equivalence in this setting was proved by Mumford **[142]**).

22 The full generality under which this result holds is actions of *reductive* groups, which we will not define here, but includes all examples we discuss.

23 This means that there is a finite set of polynomials $\{q_1, q_2, \ldots, q_t\}$ in $\mathbb{F}[X]^G$ so that for every polynomial $p \in \mathbb{F}[X]^G$ there is a t-variate polynomial r over $\mathbb{F}$ so that $p = r(q_1, q_2, \ldots, q_t)$.

24 For example, have *small* arithmetic circuits or formulae.

25 There are additional desirable structural qualities of generating sets that we will not discuss, e.g., completely understanding algebraic relations between these polynomials (called *syzygies*).

Theorem 10.1 ([143]). *For an algebraically closed field $\mathbb{F}$ of characteristic 0, the following are equivalent for any two $u, v \in V$ and generating set P of the invariant ring $\mathbb{F}[X]^G$:*

- *The orbit closures of u and v intersect.*
- *For every polynomial $p \in P$, $p(v) = p(u)$.*

10.1. Geometric complexity theory

We now briefly explain one direction from which computational complexity became interested in these algebraic problems, in work that has generated many new questions and collaboration between the fields. First, some quick background on the main problem of arithmetic complexity theory (see that chapter in [176] for definitions and more discussion). In [173], Valiant defined arithmetic analogs $\mathcal{VP}$ and $\mathcal{VNP}$ of the complexity classes $\mathcal{P}$ and $\mathcal{NP}$, respectively, and conjectured that these two arithmetic classes are different. He further proved (via surprising completeness results) that to separate these classes it is sufficient to prove that the *permanent* polynomial on $n \times n$ matrices does not project to the *determinant* polynomial on $m \times m$ matrices for any $m = \text{poly}(n)$. Note that this is a pure and concrete algebraic formulation of a central computational conjecture.

In a series of papers, Mulmuley and Sohoni introduced *Geometric Complexity Theory* (GCT) to tackle this major open problem.[26] This program is surveyed by Mulmuley here [139,140], as well as in Landsberg's book [116]. Very concisely, the GCT program starts off as follows. First, a simple "padding" of the $n \times n$ permanent polynomial makes it have degree m and act on the entries of an $m \times m$ matrix. Consider the linear group SL_{m^2} action on all entries of such $m \times m$ matrices. This action extends to polynomials in those variables, and so in particular the two we care about: determinant and modified permanent. *The main connection is that the permanent projects to the determinant (in Valiant's sense) if and only if the orbit closures of these two polynomials intersect*. Establishing that they do not intersect (for $m = \text{poly}(n)$) naturally leads to questions about finding representation-theoretic obstructions to such intersection (and hence, to the required computational lower bound). This is where things get very complicated, and describing them is beyond the scope of this survey. We note that to date, the tools of algebraic geometry and representation theory were not sufficient even to improve the quadratic bound on m of Mignon and Ressayre [132]. Indeed, some recent developments show severe limitations to the original GCT approach (and perhaps guiding it in more fruitful directions); see [42] and its historical account. Nevertheless, this line of attack (among others in computational complexity) has lead to many new questions in computational commutative algebra and to growing collaborations between algebraists and complexity theorists—we will describe some of these now.

To do so, we will focus on two natural actions of linear groups on *tuples* of matrices, simultaneous conjugation and the left–right action. Both are special cases of *quiver*

26 Origins of using invariant theory to argue computational difficulty via similar techniques go back to Strassen [166].

representations (see [55, 74]).[27] For these two group actions, we will discuss, respectively in Sections 10.2, 10.3, the classical questions and results on the rings of invariants, and recent advances motivated by computational considerations. Section 10.4 will be devoted to significant extensions of the algorithmic technique developed for the left–right action. And in Section 10.5 we will close full circle and discuss yet another central problem on matrix tuples, namely the Symbolic Determinant Identity Testing (SDIT) problem, which ties together many aspects we have seen and suggest further interesting challenges in the interface of computational complexity with invariant theory and algebraic geometry.

10.2. Simultaneous conjugation

Consider the following action of $\mathrm{SL}_n(\mathbb{F})$ on d-tuples of $n \times n$ matrices. We have $m = dn^2$ variables arranged as d $n \times n$ matrices $X = (X_1, X_2, \ldots, X_d)$. The action of a matrix $Z \in \mathrm{SL}_n(\mathbb{F})$ on this tuple is by simultaneous conjugation, by transforming it to the tuple $(Z^{-1}X_1Z, Z^{-1}X_2Z, \ldots, Z^{-1}X_dZ)$. Now, the general question above, for this action, is which polynomials in the variables X are invariant under this action?

The work of Procesi, Formanek, Razmyslov, and Donkin [60, 72, 149, 153] provides a good set (in most aspects discussed above) of generating invariants (over algebraically closed fields of characteristic zero). The generators are simply the traces of products of length at most n^2 of the given matrices,[28] namely the set

$$\left\{\mathrm{Tr}(X_{i_1}X_{i_2}\cdots X_{i_t}) : t \le n^2, i_j \in [d]\right\}.$$

These polynomials are explicit, have small degree, and are easily computable. The one shortcoming is the *exponential* size of this generating set. For example, using it to decide the intersection of orbit closures will only lead to an exponential time algorithm.

By Hilbert's so-called "Noether's normalization lemma" [91],[29] we know that the size of this set of generating invariants can, in principle (as the proof is existential), be reduced to $dn^2 + 1$. Indeed, when the group action is on a vector space of dimension m, taking $m + 1$ "random" linear combinations of any finite generating set will result (with probability 1) in a small generating set. However, as we start with an exponential number of generators above, this procedure is both inefficient and also not explicit (it is not clear how to make it deterministic). One can get an explicit generating set of minimal size deterministically using the Gröbner basis algorithm (see [129] for the best known complexity bounds) but this will take doubly exponential time in n.

---

27 We will not elaborate on the theory of quiver representations here, but only remark that reductions and completeness occur in this study as well! The left–right quiver is *complete* in a well defined sense (see [50, SECTION 5]). Informally, this means understanding its (semi)invariants implies the same understanding of the (semi)invariants of *all* acyclic quivers.

28 Convince yourself that such polynomials are indeed invariant.

29 We remark that this is the same foundational paper which proved the *finite basis* and *Nullstellensatz* theorems. It is interesting that Hilbert's initial motivation to formulate and prove these cornerstones of commutative algebra was the search for invariants of linear actions.

The works above [71,141] reduce this complexity to polynomial time! This happened in two stages. First, Mulmuley [141] gave a probabilistic polynomial time algorithm, by cleverly using the structure of the exponentially many invariants above (using which one can obtain sufficiently random linear combinations using only polynomially many random bits and in polynomial time). He then argues that using conditional derandomization results (discussed in the chapter on randomness in [176]), one can derive a deterministic polynomial time algorithm under natural computational hardness assumptions. Shortly afterwards, Forbes and Shpilka [71] derandomized a variant of Mulmuley's algorithm *without* any unproven assumption, yielding an unconditional deterministic polynomial time algorithm for the problem! Their algorithm uses the derandomization methodology: very roughly speaking, they first notice that Mulmuley's probabilistic algorithm can be implemented by a very restricted computational model (a certain read-once branching program), and then use an efficient pseudorandom generator for this computational model. Here is one important algorithmic corollary (which can be extended to other quivers).

Theorem 10.2 ([71,141]). *There is a deterministic polynomial time algorithm to solve the following problem: given two tuples of* rational *matrices* $(A_1, A_2, \ldots, A_d), (B_1, B_2, \ldots, B_d)$, *determine if the closure of their orbits under simultaneous conjugation intersect.*

It is interesting to remark that if we only consider the orbits themselves (as opposed to their closure), namely ask if there is $Z \in \mathrm{SL}_n(\mathbb{F})$ such that for all $i \in [d]$ we have $Z^{-1} A_i Z = B_i$, this becomes the *module isomorphism* problem over $\mathbb{F}$. For this important problem, there is a deterministic algorithm (of a very different nature than above, using other algebraic tools) that can solve the problem over any field $\mathbb{F}$ using only a polynomial number of arithmetic operations over $\mathbb{F}$ [40].

10.3. Left–right action

Consider now the following action of two copies, $\mathrm{SL}_n(\mathbb{F}) \times \mathrm{SL}_n(\mathbb{F})$ on d-tuples of $n \times n$ matrices. We still have $m = dn^2$ variables arranged as d $n \times n$ matrices $X = (X_1, X_2, \ldots, X_d)$. The action of a pair of matrices $(Z, W) \in \mathrm{SL}_n(\mathbb{F}) \times \mathrm{SL}_n(\mathbb{F})$ on this tuple is by left–right action, transforming it to the tuple $(Z^{-1} X_1 W, Z^{-1} X_2 W, \ldots, Z^{-1} X_d W)$. Again, for this action, which polynomials in the variables X are invariant under this action? Despite the superficial similarity to the to simultaneous conjugation, the invariants here have entirely different structure, and bounding their size required different arguments.

The works of [6, 54, 59, 156] provides an infinite set of generating invariants. The generators (again, over algebraically closed fields) are determinants of linear forms of the d matrices, with *matrix* coefficients C_i of arbitrary dimension. Namely the following set generates all invariants:

$$\left\{\det(C_1 \otimes X_1 + C_2 \otimes X_2 + \cdots + C_d \otimes X_d) : C_i \in M_k(\mathbb{F}), k \in \mathbb{N}\right\}.$$

These generators, while concisely described, fall short on most goodness aspects above, and we now discuss improvements. First, by Hilbert's finite generation, we know

in particular that some finite bound k on the dimension of the matrix coefficients C_i exist. A quest to find explicit bounds on k ensued. A quadratic upper bound $k \leq n^2$ was obtained by Derksen and Makam [50] after a long sequence of improvements described there. Still, there is an exponential number[30] of possible matrix coefficients of this size exist. However, it is easy to see that picking the C_i at random leads to a *probabilistic* polynomial time algorithm for the orbit closure intersection for this left–right action. A sequence of developments which we describe below and in the next subsection, eventually led to a *deterministic* polytime algorithm for this problem over the complex numbers by Allen-Zhu, Garg, Li, Oliveira, and Wigderson [13]. A different, simpler algorithm which works for all fields was later found by Derksen and Makam [51]).

Theorem 10.3 ([13,51]). *There is a deterministic polynomial time algorithm to solve the following problem: given two tuples of matrices* $(A_1, A_2, \ldots, A_d), (B_1, B_2, \ldots, B_d)$*, determine if the closure of their orbits under the left–right action intersect.*

In the remainder we discuss an important special case of this problem, namely when all $B_i = 0$, for which *deterministic* polynomial time algorithms were found first, which were key to the general result above. While this problem is in commutative algebra, this algorithm surprisingly has implications in analysis, noncommutative algebra, computational complexity, quantum information theory, and other areas. We will mention some of these, but let us start by defining the problem.

For an action of a linear group G on a vector space V, define the *nullcone* of the action to be the set of all points $v \in V$ such that the closure of the orbit Gv contains 0. The points in the nullcone are sometimes called *unstable*. The nullcone is of fundamental importance in invariant theory! Some examples of nullcones for actions we have discussed are the following. For the action of $\mathrm{SL}_n(\mathbb{C})$ on $M_n(\mathbb{C})$ by left multiplication, it is the set of *singular* matrices. For the action of $\mathrm{SL}_n(\mathbb{C})$ on $M_n(\mathbb{C})$ by conjugation, it is the set of *nilpotent* matrices. As you would guess (and follows from Theorem 10.1), the nullcone is precisely the set of points in V which vanish under all invariant polynomials. Thus if we have a good generating set, one can use them to efficiently test membership in the nullcone. However, we are not in this situation for the left–right action. Despite that, deterministic polynomial-time algorithms were obtained, independently, by Garg, Gurvits, Oliveira, and Wigderson [75] (which is analytic in nature) over the complex numbers, and by Ivanyos, Qiao, and Subrahmanyam [95] (which is algebraic in nature) and works for all fields. These two algorithms have different properties, and use in different ways the upper bounds on the dimension of matrix coefficients in the invariants.[31]

Theorem 10.4 ([75,95]). *There is a deterministic polynomial time algorithm that, on a given tuple of matrices* $(A_1, A_2, \ldots, A_d)$ *in* $M_n(\mathbb{F})$*, determines if it is in the nullcone of the left–right action.*

---

30 Well, a possibly infinite number, but it can be reduced to exponential.

31 Yet a third algorithm, quite different than the two above, was very recently developed by Hamada and Hirai [88].

We will focus in what follows on the first algorithm. We discuss its broad extensions in the next subsection. Here we discuss some of its diverse consequences to basic problems in different fields (reflecting the many different mathematical objects that can be represented by matrix tuples). All the precise definitions of the notions below, as well as the proofs, interconnections and the meandering story leading to it can be found in [75,76].

Theorem 10.5 ([75,76]). *There are deterministic polynomial-time algorithms to solve the following problems:*

- (Analysis) *The feasibility problem for Brascamp–Lieb inequalities, and more generally, computing the optimal constant for each one.*
- (Noncommutative algebra) *The word problem over the free skew field (of rational functions in noncommuting variables).*
- (Quantum information theory) *Testing if a completely positive quantum operator is rank-decreasing.*
- (Arithmetic complexity) *Approximating the commutative rank of a symbolic matrix to within a factor of two.*[32]

We note that this algorithm also inspired purely structural results, both in the areas mentioned above, but also in others. In frame theory, it led to the complete resolution of the central Paulsen problem [89,113]. In statistics, it has led to complete understanding of when Maximum Likelihood Estimates (MLE) exist, and when they are unique, first for matrix random models [15,52] and then for tensor random models [53].

10.4. Nullcones, moment polytopes, geodesic convexity, and noncommutative optimization

Reflecting on the algorithm of [75] from the previous section marked several features which merited further investigation. For one, it is an analytic/numerical algorithm, very different that the typical algebraic/symbolic algorithms so common for problems of invariant theory and algebraic geometry, and in the applications above. This algorithm is a special case of a general heuristic called *alternate minimization*, common in optimization, statistics, and machine learning, where the input evolves via a sequence of local, greedy steps. In general, convergence of such algorithms, let alone fast convergence, may not be guaranteed or is hard to establish, whereas here it always converges, and in polynomial time! The analysis uses the fact that the evolution above happens along the orbit of the input by the left–right group action, and tracks a measure called *capacity* which this evolution minimizes. And fast convergence to a unique optimum occurs despite the fact that both the domain (a pair of continuous linear groups) and the optimized function are patently nonconvex.

Understanding the power of such continuous optimization algorithms for a larger and larger classes of nullcone problems (capturing other problems, in discrete optimization,

32 Computing this rank exactly is the central PIT problem, discussed at the last subsection.

quantum information, representation theory, and other areas) progressed in a series of papers, culminating in a general theory, that applies in principle to *any* linear (reductive) group action [41]. The paper contains a detailed account of the history, background, theory, and applications, and we relate below just the highlights, partly explaining the mysteries above.

First, nullcone problems may be viewed as optimization problems for general group actions, where the *capacity* being minimized is simply the minimum norm of any element in the orbit of the input. This viewpoint then benefits from a beautiful noncommutative duality theory (the Kempf–Ness theorem [106], which greatly expends linear programming duality in the commutative case). Underlying this theorem are notions of *geodesic convexity* (extending the Euclidean one) and *moment maps* (extending the Euclidean gradient). Thus, the seeming nonconvexity of these problems mentioned above only stems from the wrong representation: viewed with the correct metric on the Riemannian manifold which is the acting group makes both the domain and optimization goal (geodesically) convex, explaining convergence to a unique optimal point, which determines membership in the nullcone.

Using these tools, it turns out that the most basic tools of convex optimization in Euclidean space extend to the far more general setting of Riemannian manifolds that arise from the symmetries of noncommutative groups. The paper develops "geodesic" first- and second-order algorithms in this setting, and analyzes their performance in general. Proving convergence bounds requires making quantitative the duality theory above, which uses significant algebraic and analytic machinery. However, the bounds themselves depend in an elegant way on few natural geometric "smoothness" parameters (arising from the given group action), in analogy with the Euclidean (commutative) case.

These algorithms can actually be modified to solve a significant generalization of the nullcone membership problem, namely computing membership in so-called *moment polytopes*, implicitly defined polyhedral bodies associated with any linear group action. These capture a variety of "scaling problems," such as marginal problems in classical and quantum information theory, as well as basic combinatorial optimization problem such as the matroid intersection problem.

10.5. Symbolic determinants, varieties, and circuit lower bounds

We now return to a another basic computational question on matrix tuples, the *Symbolic Determinant Identity Testing* (SDIT) problem (of interest over any field $\mathbb{F}$): Given a tuple of $n \times n$ matrices $(A_1, A_2, \ldots, A_d)$, determine if the symbolic determinant $\det(\sum_i x_i A_i)$ vanishes as a polynomial in the variables x_i. This problem has a several different formulations, and has arisen independently in different fields. We mention a few.

One equivalent formulation comes from considering the linear space $\{\sum_i c_i A_i : c_i \in \mathbb{F}\}$ arising from all possible evaluations of the variables x_i. Then SDIT asks if this matrix space contains only singular matrices. In algebraic geometry, it arises in close connection with certain sheaves on projective space [68]. In topology, it arises naturally in connection to linearly independent vector fields on spheres, which led to the development of the Adams operations on topological K-theory [2, 3]. In invariant theory, they were used

by Dieudonné [57] to classify the symmetries of the determinant, recovering a result of Frobenius [73].

Another equivalent formulation brings up beautiful connections between the cases of commuting and noncommuting variables x_i. Consider the symbolic matrix $A(x) = \sum_i x_i A_i$. In this terminology SDIT becomes the question of the invertibility of this symbolic matrix $A(x)$ over the field $\mathbb{F}(x)$ of rational functions in the (commuting) variables x_i. In his seminal work on the (noncommutative) free skew field, Cohn [44] proved that the elements of this field can be described as inverses of such symbolic matrices in *noncommuting* variables x_i. Thus the noncommutative analog of SDIT is the word problem for this field.[33] Another connection mentioned in Section 10.3 above is that the noncommutative SDIT problem is equivalent to the nullcone problem for the left–right action! Recalling the generating invariants for this group action from Section 10.3, one observes that (commutative) SDIT is the question of vanishing of the lowest possible invariant, $k = 1$.

SDIT played a crucial role in algorithms and computational complexity. It was initially raised by Edmonds [67] in the context of combinatorial optimization. Another interpretation of Valiant's completeness result is that SDIT captures the general Polynomial Identity Testing (PIT) problem (see the survey [159]). Noting that SDIT has a simple fast probabilistic algorithm over large fields (namely, assign random values to the variables and evaluate the resulting numeric determinant), finding an efficient deterministic algorithm became one of the most basic *derandomization* challenges, which has been under attack now for half a century. The difficulty (and importance) of finding such a deterministic algorithm was clarified (bigtime) by the following remarkable result of Kabanets and Impagliazzo [101].

Theorem 10.6 ([101]). *If there is a deterministic polynomial time algorithm for SDIT, then either $\mathcal{VP} \neq \mathcal{VNP}$, or $\mathcal{NEXP}$ has no polynomial size Boolean circuits.*

In simpler words, such a derandomization will result in a major breakthrough in computational complexity, providing explicit lower bounds either in arithmetic or Boolean complexity, each in the ballpark of proving $\mathcal{P} \neq \mathcal{NP}$. Even the logical nature of this theorem statement demands attention: it states that an efficient algorithm for one problem (SDIT) will mean that host of other natural problems have no efficient algorithm!

On the other hand, this theorem suggests a concrete algorithmic attack on these lower bound questions (and in particular, $\mathcal{VP}$ vs. $\mathcal{VNP}$) discussed in Section 10.1—simply design a deterministic algorithm for SDIT. The past decades have seen much progress on designing such algorithms for a variety of special cases of SDIT (and the more general PIT), which is far too large to survey here. We conclude here with the possibility of finding such an algorithm via the new algorithmic techniques described in Section 10.4 above. This is explored in [125], and we only summarize what is currently known.

First, let us note that the set of singular matrices is an algebraic variety. Thus SDIT is a special case of a very large class of natural problems. Fix an algebraic variety in $U \subset \mathbb{F}^m$

33 Valiant's completeness result [172], mentioned in Section 10.1, analogously makes SDIT the word problem for the commutative field $\mathbb{F}(x)$.

(e.g., for SDIT, $m = dn^2$). Given a point $u \in \mathbb{F}^m$ (e.g., for SDIT u is a matrix tuple), determine if $u \in U$. Of course, it is natural to work here with algebraically closed fields, e.g., $\mathbb{F} = \mathbb{C}$.

Such membership problems in algebraic varieties obviously arise naturally in numerous settings. One way to view the developments of the previous section is that if the variety U is actually the nullcone of a (nice) group action, then continuous, convex optimization methods (extended to the geodesic setting), such a gradient descent, may be far more efficient than symbolic, algebraic algorithms, and indeed in some cases may have polynomial-time convergence. Thus, a first question to ask is whether SDIT itself is the nullcone of some group action. Unfortunately, it is not (unless $d \leq 2$ or $n \leq 2$), again possibly helping to understand its difficulty.

Theorem 10.7 ([125]). *For $d, n \geq 3$, SDIT for a d-tuple of $n \times n$ matrices is not the nullcone of any linear group action.*

A central part of the proof of this theorem is the characterization the symmetries of the SDIT variety, extending to d-tuples for any d the aforementioned theorem of Frobenuis [73] for the case $d = 1$. Among some of the natural directions suggested by this work we name three basic ones: (1) Find methods of determining the symmetries of naturally given algebraic varieties; (2) Find methods to determine if a given algebraic variety is the nullcone of a linear group action; (3) Extend the convex optimization methods of Section 10.4 to prove membership in other algebraic varieties, beside nullcones.

FUNDING

This work was partially supported by NSF grants CCF-1412958 and CCF-1900460.

REFERENCES

[1] D. Achlioptas, A. Coja-Oghlan, and F. Ricci-Tersenghi, On the solution-space geometry of random constraint satisfaction problems. *Random Structures Algorithms* **38** (2011), no. 3, 251–268.

[2] J. F. Adams, Vector fields on spheres. *Ann. of Math. (2)* **75** (1962), 603–632.

[3] J. F. Adams, P. D. Lax, and R. S. Phillips, On matrices whose real linear combinations are nonsingular. *Proc. Amer. Math. Soc.* **16** (1965), no. 2, 318–322.

[4] L. Adleman and M. Huang, *Primality testing and abelian varieties over finite fields*. Lecture Notes in Math. 1512, Springer, 1992.

[5] L. Adleman, C. Pomerance, and R. Rumely, On distinguishing prime numbers from composite numbers. *Ann. of Math.* (1983), 173–206.

[6] B. Adsul, S. Nayak, and K. Subrahmanyam, *A geometric approach to the Kronecker problem II: rectangular shapes, invariants of $n \times n$ matrices, and a generalization of the Artin-Procesi theorem*. Citeseer, 2010.

[7] M. Agrawal and S. Biswas, Primality and identity testing via Chinese remaindering. *J. ACM* **50** (2003), no. 4, 429–443.

[8] M. Agrawal, N. Kayal, and N. Saxena, Primes is in P. *Ann. of Math.* **160** (2004), no. 2, 781–793.

[9] M. Ajtai, Generating hard instances of lattice problems. In *Proceedings of the 28th annual ACM symposium on theory of computing*, pp. 99–108, ACM, 1996.

[10] M. Ajtai and C. Dwork, A public-key cryptosystem with worst-case/average-case equivalence. In *Proceedings of the 29th annual ACM symposium on theory of computing*, pp. 284–293, ACM, 1997.

[11] D. J. Aldous, Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de probabilités xvii 1981/82*, pp. 243–297, Springer, 1983.

[12] D. J. Aldous, The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.* **3** (1990), no. 4, 450–465.

[13] Z. Allen-Zhu, A. Garg, Y. Li, R. Oliveira, and A. Wigderson, Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pp. 172–181, ACM, 2018.

[14] N. Alon and B. Klartag, Economical toric spines via Cheeger's inequality. *J. Topol. Anal.* **1** (2009), no. 02, 101–111.

[15] C. Améndola, K. Kohn, P. Reichenbach, and A. Seigal, Invariant theory and scaling algorithms for maximum likelihood estimation. *SIAM J. Appl. Algebra Geom.* **5** (2021), no. 2, 304–337.

[16] N. Anari, V. Jain, F. Koehler, H. T. Pham, and T.-D. Vuong, Entropic independence in high-dimensional expanders: Modified log-Sobolev inequalities for fractionally log-concave polynomials and the Ising model. 2021, arXiv:2106.04105.

[17] N. Anari, K. Liu, and S. O. Gharan, Spectral independence in high-dimensional expanders and applications to the hardcore model. *SIAM J. Comput.* (2021), FOCS20-1–FOCS20-37. DOI 10.1137/20M1367696.

[18] N. Anari, K. Liu, S. O. Gharan, and C. Vinzant, Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid. In *Proceedings of the 51st annual ACM SIGACT symposium on theory of computing*, pp. 1–12, ACM, 2019.

[19] S. Arora and B. Barak, *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[20] S. Arora, J. Lee, and A. Naor, Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.* **21** (2008), no. 1, 1–21.

[21] S. Arora, S. Rao, and U. Vazirani, Expander flows, geometric embeddings and graph partitioning. *J. ACM* **56** (2009), no. 2, 1–37.

[22] H. Bäärnhielm and C. Leedham-Green, The product replacement prospector. *J. Symbolic Comput.* **47** (2012), no. 1, 64–75.

[23] L. Babai, Local expansion of vertex-transitive graphs and random generation in finite groups. In *Proceedings of the 23rd annual ACM symposium on theory of computing, STOC 1991*, pp. 164–174, Citeseer, 1991.

[24] L. Babai, Graph isomorphism in quasipolynomial time. 2015, arXiv:1512.03547.

[25] L. Babai and E. Szemerédi, On the complexity of matrix group problems I. In *Proceedings of 25th annual IEEE symposium on foundations of computer science*, pp. 229–240, IEEE, 1984.

[26] E. Bach and J. Shallit, *Algorithmic number theory: Efficient algorithms. 1*. MIT Press, Cambridge, 1997.

[27] B. Barak, Z. Dvir, A. Wigderson, and A. Yehudayoff, Fractional Sylvester–Gallai theorems. *Proc. Natl. Acad. Sci.* **110** (2013), no. 48, 19213–19219.

[28] B. Barak, R. Impagliazzo, and A. Wigderson, Extracting randomness using few independent sources. *SIAM J. Comput.* **36** (2006), no. 4, 1095–1118.

[29] J. Batson, D. A. Spielman, and N. Srivastava, Twice-Ramanujan sparsifiers. *SIAM Rev.* **56** (2014), no. 2, 315–334.

[30] M. Ben-Or and N. Linial, Collective coin flipping, robust voting schemes and minima of Banzhaf values. In *Proceedings of 26th annual IEEE symposium on foundations of computer science*, pp. 408–416, IEEE, 1985.

[31] I. Benjamini, G. Kalai, and O. Schramm, Noise sensitivity of Boolean functions and applications to percolation. *Publ. Math. IHÉS* **90** (1999), no. 1, 5–43.

[32] E. Berlekamp, Factoring polynomials over finite fields. *Bell Syst. Tech. J.* **46** (1967), no. 8, 1853–1859.

[33] A. Besicovitch, Sur deux questions d'integrabilité des fonctions. *J. Soc. Phys. Math.* **2** (1919), 105–123.

[34] Y. Bilu and N. Linial, Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica* **26** (2006), no. 5, 495–519.

[35] C. Borell, Geometric bounds on the Ornstein–Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete* **70** (1985), no. 1, 1–13.

[36] J. Bourgain, On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.* **52** (1985), no. 1–2, 46–52.

[37] J. Bourgain, N. Katz, and T. Tao, A sum-product estimate in finite fields, and applications. *Geom. Funct. Anal.* **14** (2004), no. 1, 27–57.

[38] J. Bourgain and L. Tzafriri, On a problem of Kadison and Singer. *J. Reine Angew. Math.* **420** (1991), 1–43.

[39] A. Broder, Generating random spanning trees. In *Proceedings of 30th annual IEEE symposium on foundations of computer science*, pp. 442–447, IEEE, 1989.

[40] P. A. Brooksbank and E. M. Luks, Testing isomorphism of modules. *J. Algebra* **320** (2008), no. 11, 4020–4029.

[41] P. Bürgisser, C. Franks, A. Garg, R. Oliveira, M. Walter, and A. Wigderson, Towards a theory of non-commutative optimization: geodesic first and second order methods for moment maps and polytopes. 2019, arXiv:1910.12375.

[42] P. Bürgisser, C. Ikenmeyer, and G. Panova, No occurrence obstructions in geometric complexity theory. 2016, arXiv:1604.06431.

[43] A. Cayley, *On the theory of linear transformations*. E. Johnson, 1845.

[44] P. M. Cohn, *Skew field constructions*. 27, CUP Archive, 1977.

[45] G. Cooperman, Towards a practical, theoretically sound algorithm for random generation in finite groups. 2002, arXiv:math/0205203.

[46] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Undergrad. Texts Math., Springer, New York, 1992.

[47] D. Dadush and O. Regev, Towards strong reverse Minkowski-type inequalities for lattices. In *Proceedings of 57th annual IEEE symposium on foundations of computer science*, pp. 447–456, IEEE, 2016.

[48] R. Davies, Some remarks on the Kakeya problem. *Math. Proc. Cambridge Philos. Soc.* **69** (1971), 417–421.

[49] H. Derksen and G. Kemper, *Computational invariant theory*. Springer, 2015.

[50] H. Derksen and V. Makam, Polynomial degree bounds for matrix semi-invariants. 2015, arXiv:1512.03393.

[51] H. Derksen and V. Makam, Algorithms for orbit closure separation for invariants and semi-invariants of matrices. *Algebra Number Theory* **14** (2020), no. 10, 2791–2813.

[52] H. Derksen and V. Makam, Maximum likelihood estimation for matrix normal models via quiver representations. *SIAM J. Appl. Algebra Geom.* **5** (2021), no. 2, 338–365.

[53] H. Derksen, V. Makam, and M. Walter, Maximum likelihood estimation for tensor normal models via castling transforms. 2020, arXiv:2011.03849.

[54] H. Derksen and J. Weyman, Semi-invariants of quivers and saturation for Littlewood–Richardson coefficients. *J. Amer. Math. Soc.* **13** (2000), no. 3, 467–479.

[55] H. Derksen and J. Weyman, The combinatorics of quiver representations. 2006, arXiv:math/0608288.

[56] P. Diaconis, Group representations in probability and statistics. *Lecture Notes Monogr. Ser.* **11** (1988), i–192.

[57] J. Dieudonné, Sur une généralisation du groupe orthogonal à quatre variables. *Arch. Math.* **1** (1948), no. 4, 282–287.

[58] J. Dixon, Generating random elements in finite groups. *Electron. J. Combin.* **13** (2008), no. R94, 1.

[59] M. Domokos and A. Zubkov, Semi-invariants of quivers as determinants. *Transform. Groups* **6** (2001), no. 1, 9–24.

[60] S. Donkin, Invariants of several matrices. *Invent. Math.* **110** (1992), no. 1, 389–401.

[61] Z. Dvir, On the size of Kakeya sets in finite fields. *J. Amer. Math. Soc.* **22** (2009), no. 4, 1093–1097.

[62] Z. Dvir, Guest column: From randomness extraction to rotating needles. *ACM SIGACT News* **40** (2010), no. 4, 46–61.

[63] Z. Dvir, S. Kopparty, S. Saraf, and M. Sudan, Extensions to the method of multiplicities, with applications to Kakeya sets and mergers. *SIAM J. Comput.* **42** (2013), no. 6, 2305–2328.

[64] Z. Dvir and A. Shpilka, Locally decodable codes with two queries and polynomial identity testing for depth 3 circuits. *SIAM J. Comput.* **36** (2007), no. 5, 1404–1434.

[65] M. Dyer, A. Frieze, and R. Kannan, A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM* **38** (1991), no. 1, 1–17.

[66] M. Dyer, A. Sinclair, E. Vigoda, and D. Weitz, Mixing in time and space for lattice spin systems: a combinatorial view. *Random Structures Algorithms* **24** (2004), no. 4, 461–479.

[67] J. Edmonds, Systems of distinct representatives and linear algebra. *J. Res. Natl. Bur. Stand., B* **71** (1967), 241–245.

[68] D. Eisenbud and J. Harris, Vector spaces of matrices of low rank. *Adv. Math.* **70** (1988), no. 2, 135–155.

[69] U. Feige, G. Kindler, and R. O'Donnell, Understanding parallel repetition requires understanding foams. In *Proceedings of the 22nd annual IEEE conference on computational complexity*, pp. 179–192, IEEE, 2007.

[70] W. Feller, *An introduction to probability theory and its applications*. Wiley Ser. Prob. Math. Stat., Wiley, 1971.

[71] M. Forbes and A. Shpilka, Explicit Noether normalization for simultaneous conjugation via polynomial identity testing. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, pp. 527–542, Springer, 2013.

[72] E. Formanek, Invariants and the ring of generic matrices. *J. Algebra* **89** (1984), no. 1, 178–223.

[73] F. G. Frobenius, Über die Darstellung der endlichen Gruppen durch lineare Substitutionen. *Sitz. Kön. Preuss. Akad. Wiss. Berlin* (1897), 944–1015.

[74] P. Gabriel, Unzerlegbare darstellungen I. *Manuscripta Math.* **6** (1972), no. 1, 71–103.

[75] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson, A deterministic polynomial time algorithm for non-commutative rational identity testing. 2015, arXiv:1511.03730.

[76] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson, Algorithmic aspects of Brascamp–Lieb inequalities. 2016, arXiv:1607.06711.

[77] C. Gentry, Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st annual ACM symposium on theory of computing, STOC '09*, pp. 169–178, ACM, 2009.

[78] M. Goemans, Semidefinite programming in combinatorial optimization. *Math. Program.* **79** (1997), no. 1–3, 143–161.

[79] O. Goldreich, *Computational complexity: a conceptual perspective*. Cambridge University Press, Cambridge, 2008.

[80] S. Goldwasser and J. Kilian, Almost all primes can be quickly certified. In *Proceedings of the 18th annual ACM symposium on theory of computing*, pp. 316–329, ACM, 1986.

[81] S. Goldwasser, S. Micali, and C. Rackoff, The knowledge complexity of interactive proof systems. *SIAM J. Comput.* **18** (1989), no. 1, 186–208.

[82] A. Granville, It is easy to determine whether a given integer is prime. *Bull. Amer. Math. Soc.* **42** (2005), 3–38.

[83] M. Gromov, Essays in group theory. *Math. Sci. Res. Inst. Publ.* **8** (1987), 75–263.

[84] H. Guo, M. Jerrum, and J. Liu, Uniform sampling through the Lovász local lemma. 2016, arXiv:1611.01647.

[85] V. Guruswami, J. Lee, and A. Razborov, Almost Euclidean subspaces of ℓ_1^n via expander codes. *Combinatorica* **30** (2010), no. 1, 47–68.

[86] L. Gurvits, Van der Waerden/Schrijver–Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: One theorem for all. *Electron. J. Combin.* **15** (2008), no. 1, R66.

[87] L. Guth and N. Katz, On the Erdős distinct distance problem in the plane. 2010, arXiv:1011.4105.

[88] M. Hamada and H. Hirai, Computing the NC-rank via discrete convex optimization on CAT(0) spaces. *SIAM J. Appl. Algebra Geom.* **5** (2021), no. 3, 455–478.

[89] L. Hamilton and A. Moitra, The Paulsen problem made simple. *Israel J. Math.* **246** (2021), no. 1, 299–313.

[90] O. Heilmann and E. Lieb, Theory of monomer-dimer systems. *Comm. Math. Phys.* **25** (1972), no. 3, 190–232.

[91] D. Hilbert, Über die vollen Invariantensysteme. *Math. Ann.* **42** (1893), no. 3, 313–373.

[92] D. Holt, B. Eick, and E. O'Brien, *Handbook of computational group theory*. CRC Press, 2005.

[93] S. Hoory, N. Linial, and A. Wigderson, Expander graphs and their applications. *Bull. Amer. Math. Soc.* **43** (2006), no. 4, 439–561.

[94] R. Impagliazzo and A. Wigderson, $P = \mathrm{BPP}$ unless E has subexponential circuits: Derandomizing the XOR lemma. In *Proceedings of the 29th annual ACM symposium on theory of computing*, pp. 220–229, ACM, 1997.

[95] G. Ivanyos, Y. Qiao, and K. Subrahmanyam, Non-commutative Edmonds' problem and matrix semi-invariants. 2015, arXiv:1508.00690.

[96] M. Jerrum, The Markov chain Monte Carlo method: An approach to approximate counting and integration. In *Approximation algorithm for NP-hard problems*, pp. 482–520, PWS Publishing, 1996.

[97] M. Jerrum, A. Sinclair, and E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM* **51** (2004), no. 4, 671–697.

[98] M. Jerrum and G. Sorkin, Simulated annealing for graph bisection. In *Proceedings of 34th annual IEEE symposium on foundations of computer science*, pp. 94–103, IEEE, 1993.

[99] M. Jerrum, L. Valiant, and V. Vazirani, Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.* **43** (1986), 169–188.

[100] Z. Ji, A. Natarajan, T. Vidick, J. Wright, and H. Yuen, MIP* = RE. 2020, arXiv:2001.04383.

[101] V. Kabanets and R. Impagliazzo, Identity tests means proving circuit lower bounds. *Comput. Complexity* **13** (2004), no. 1–2, 1–46.

[102] R. Kadison and I. Singer, Extensions of pure states. *Amer. J. Math.* **81** (1959), no. 2, 383–400.

[103] J. Kahn, G. Kalai, and N. Linial, The influence of variables on Boolean functions. In *Proceedings of 29th annual IEEE symposium on foundations of computer science*, pp. 68–80, IEEE, 1988.

[104] G. Kasparov and G. Yu, The coarse geometric Novikov conjecture and uniform convexity. *Adv. Math.* **206** (2006), no. 1, 1–56.

[105] N. Kayal and S. Saraf, Blackbox polynomial identity testing for depth 3 circuits. In *Proceedings of 50th annual IEEE symposium on foundations of computer science*, pp. 198–207, IEEE, 2009.

[106] G. Kempf and L. Ness, The length of vectors in representation spaces. In *Algebraic geometry*, pp. 233–243, Springer, 1979.

[107] S. Khot, On the power of unique 2-prover 1-round games. In *Proceedings of the 34th annual ACM symposium on theory of computing*, pp. 767–775, ACM, 2002.

[108] S. Khot, Inapproximability of NP-complete problems, discrete Fourier analysis, and geometry. In *Proceedings of the international congress of mathematicians, vol. 5*, pp. 2676–2697, Hindustan Book Agency, New Delhi, 2010.

[109] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell, Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.* **37** (2007), no. 1, 319–357.

[110] S. Khot and N. Vishnoi, The unique games conjecture, integrality gap for cut problems and embeddability of negative-type metrics into ℓ_1. In *Proceedings of 46th annual IEEE symposium on foundations of computer science*, pp. 53–62, IEEE, 2005.

[111] G. Kindler, R. O'Donnell, A. Rao, and A. Wigderson, Spherical cubes and rounding in high dimensions. In *Proceedings of 49th annual IEEE symposium on foundations of computer science*, pp. 189–198, IEEE, 2008.

[112] R. Krauthgamer, J. Lee, M. Mendel, and A. Naor, Measured descent: A new embedding method for finite metrics. *Geom. Funct. Anal.* **15** (2005), no. 4, 839–858.

[113] T. C. Kwok, L. C. Lau, Y. T. Lee, and A. Ramachandran, The Paulsen problem, continuous operator scaling, and smoothed analysis. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pp. 182–189, ACM, 2018.

[114] V. Lafforgue, Un renforcement de la propriété (T). *Duke Math. J.* **143** (2008), no. 3, 559–602.

[115] J. Lagarias, Knapsack public key cryptosystems and Diophantine approximation. In *Advances in cryptology*, pp. 3–23, Springer, 1984.

[116] J. Landsberg, *Geometry and complexity theory*. Cambridge University Press, 2017.

[117] A. Lenstra, H. Lenstra, and L. Lovász, Factoring polynomials with rational coefficients. *Math. Ann.* **261** (1982), no. 4, 515–534.

[118] H. Lenstra, Integer programming with a fixed number of variables. *Math. Oper. Res.* **8** (1983), no. 4, 538–548.

[119] N. Linial, Finite metric spaces–combinatorics, geometry and algorithms. In *Proceedings of the international congress of mathematicians III*, pp. 573–586, 2002.

[120] N. Linial, E. London, and Y. Rabinovich, The geometry of graphs and some of its algorithmic applications. *Combinatorica* **15** (1995), no. 2, 215–245.

[121] K. Liu, From coupling to spectral independence and blackbox comparison with the down-up walk. 2021, arXiv:2103.11609.

[122] C. Lu, O. Reingold, S. Vadhan, and A. Wigderson, Extractors: Optimal up to constant factors. In *Proceedings of the 35th annual ACM symposium on theory of computing*, pp. 602–611, ACM, 2003.

[123] A. Lubotzky, High dimensional expanders. In *Proceedings of the international congress of mathematicians: Rio de Janeiro 2018*, pp. 705–730, World Scientific, 2018.

[124] A. Lubotzky, R. Phillips, and P. Sarnak, Ramanujan graphs. *Combinatorica* **8** (1988), no. 3, 261–277.

[125] V. Makam and A. Wigderson, Singular tuples of matrices is not a null cone (and the symmetries of algebraic varieties). *J. Reine Angew. Math.* **2021** (2021), no. 780, 79–131.

[126] A. Marcus, D. Spielman, and N. Srivastava, Interlacing families I: Bipartite Ramanujan graphs of all degrees. In *Proceedings of 54th annual IEEE symposium on foundations of computer science*, pp. 529–537, IEEE, 2013.

[127] A. Marcus, D. Spielman, and N. Srivastava, Interlacing families II: Mixed characteristic polynomials and the Kadison–Singer problem. 2013, arXiv:1306.3969.

[128] G. Margulis, Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Probl. Inf. Transm.* **24** (1988), 39–46.

[129] E. Mayr and S. Ritscher, Space-efficient Gröbner basis computation without degree bounds. In *Proceedings of the 36th international symposium on symbolic and algebraic computation*, pp. 257–264, ACM, 2011.

[130] M. Mendel and A. Naor, Nonlinear spectral calculus and super-expanders. *Publ. Math. IHÉS* **119** (2014), no. 1, 1–95.

[131] M. Mézard, G. Parisi, and R. Zecchina, Analytic and algorithmic solution of random satisfiability problems. *Science* **297** (2002), no. 5582, 812–815.

[132] T. Mignon and N. Ressayre, A quadratic bound for the determinant and permanent problem. *Int. Math. Res. Not.* **2004** (2004), no. 79, 4241–4253.

[133] G. Miller, Riemann's hypothesis and tests for primality. *J. Comput. System Sci.* **13** (1976), no. 3, 300–317.

[134] H. Minkowski, *Geometrie der Zahlen*. Teubner, 1910.

[135] A. Moitra, Approximate counting, the Lovász local lemma and inference in graphical models. 2016, arXiv:1610.04317.

[136] R. Moser, A constructive proof of the Lovász local lemma. In *Proceedings of the 41st annual ACM symposium on theory of computing*, pp. 343–350, ACM, 2009.

[137] R. Moser and G. Tardos, A constructive proof of the general Lovász local lemma. *J. ACM* **57** (2010), no. 2, 11.

[138] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, Noise stability of functions with low influences: Invariance and optimality. *Ann. of Math.* **171** (2010), no. 1, 295–341.

[139] K. Mulmuley, On P vs. NP and geometric complexity theory. *J. ACM* **58** (2011), no. 2, 5.

[140] K. Mulmuley, The GCT program toward the P vs. NP problem. *Commun. ACM* **55** (2012), no. 6, 98–107.

[141] K. Mulmuley, Geometric complexity theory V: Equivalence between blackbox derandomization of polynomial identity testing and derandomization of Noether's normalization lemma. In *Proceedings of 53rd annual IEEE symposium on foundations of computer science*, pp. 629–638, IEEE, 2012.

[142] D. Mumford, *Algebraic geometry: Complex projective varieties. 1*. Springer, 1995.

[143] D. Mumford, J. Fogarty, and F. Kirwan, *Geometric invariant theory*, 2nd edn. Ergeb. Math. Grenzgeb. 34, Springer, 1982.

[144] A. Naor and R. Young, Vertical perimeter versus horizontal perimeter. 2017, arXiv:1701.00620.

[145] R. O'Donnell, *Analysis of Boolean functions*. Cambridge University Press, 2014.

[146] A. M. Odlyzko and H. J. J. te Riele, Disproof of the Mertens conjecture. *J. Reine Angew. Math.* **357** (1985), 138–160.

[147] C. Peikert, A decade of lattice cryptography. *Found. Trends Theor. Comput. Sci.* **10** (2016), no. 4, 283–424.

[148] M. Pohst and H. Zassenhaus, *Algorithmic algebraic number theory*. Cambridge University Press, 1989.

[149] C. Procesi, The invariant theory of $n \times n$ matrices. *Adv. Math.* **19** (1976), no. 3, 306–381.

[150] M. Rabin, Probabilistic algorithm for testing primality. *J. Number Theory* **12** (1980), no. 1, 128–138.

[151] R. Raz, A parallel repetition theorem. *SIAM J. Comput.* **27** (1998), no. 3, 763–803.

[152] R. Raz, A counterexample to strong parallel repetition. *SIAM J. Comput.* **40** (2011), no. 3, 771–777.

[153] J. Razmyslov, Trace identities of full matrix algebras over a field of characteristic zero. *Math. USSR, Izv.* **8** (1974), no. 4, 727.

[154] O. Regev and N. Stephens-Davidowitz, A reverse Minkowski theorem. 2016, arXiv:1611.05979.

[155] O. Reingold, S. Vadhan, and A. Wigderson, Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Ann. of Math.* (2002), 157–187.

[156] A. Schofield and M. Van den Bergh, Semi-invariants of quivers for arbitrary dimension vectors. *Indag. Math.* **12** (2001), no. 1, 125–138.

[157] Á. Seress, *Permutation group algorithms*. Cambridge Tracts in Math. 152, Cambridge University Press, 2003.

[158] P. Shor, Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings of 35th annual IEEE symposium on foundations of computer science*, pp. 124–134, IEEE, 1994.

[159] A. Shpilka and A. Yehudayoff, Arithmetic circuits: a survey of recent results and open questions. *Found. Trends Theor. Comput. Sci.* **5** (2010), no. 3–4, 207–388.

[160] D. Simon, Selected applications of LLL in number theory. In *The LLL algorithm*, pp. 265–282, Springer, 2010.

[161] A. Sinclair and M. Jerrum, Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.* **82** (1989), no. 1, 93–133.

[162] A. Sly, Computational transition at the uniqueness threshold. In *Proceedings of 51st annual IEEE symposium on foundations of computer science*, pp. 287–296, IEEE, 2010.

[163] R. Solovay and V. Strassen, A fast Monte-Carlo test for primality. *SIAM J. Comput.* **6** (1977), no. 1, 84–85.

[164] D. Spielman and N. Srivastava, An elementary proof of the restricted invertibility theorem. *Israel J. Math.* **190** (2012), no. 1, 83–91.

[165] D. Spielman and S. Teng, Spectral sparsification of graphs. *SIAM J. Comput.* **40** (2011), no. 4, 981–1025.

[166] V. Strassen, Relative bilinear complexity and matrix multiplication. *J. Reine Angew. Math.* **375** (1987), 406–443.

[167] B. Sturmfels, *Algorithms in invariant theory*. Springer, 2008.

[168] E. Szemerédi and W. Jr. Trotter, Extremal problems in discrete geometry. *Combinatorica* **3** (1983), no. 3–4, 381–392.

[169] T. Tao, Recent progress on the Kakeya conjecture. 2009, http://terrytao.wordpress.com/2009/05/11/.

[170] S. W. Thomson, On the division of space with minimum partitional area. *Acta Math.* **11** (1887), no. 1–4, 121–134.

[171] S. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.* **7** (2011), no. 1–3, 1–336.

[172] L. Valiant, Completeness classes in algebra. In *Proceedings of the 11th annual ACM symposium on theory of computing*, pp. 249–261, ACM, 1979.

[173] L. Valiant, The complexity of computing the permanent. *Theoret. Comput. Sci.* **8** (1979), no. 2, 189–201.

[174] S. Vempala, Geometric random walks: a survey. *Comb. Comput. Geom.* **52** (2005), no. 2, 573–612.

[175] D. Weitz, Counting independent sets up to the tree threshold. In *Proceedings of the 38th annual ACM symposium on theory of computing*, pp. 140–149, ACM, 2006.

[176] A. Wigderson, *Mathematics and computation*. Princeton University Press, 2019.

[177] T. Wolff, Recent work connected with the Kakeya problem. *Prospects Math.* **2** (1999), 129–162.

AVI WIGDERSON

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA, avi@ias.edu