In the bibliography I've included all the major sources from which I learned the topics of this lecture. The most accessible to me (and thus the ones I would most recommend for those seeking more information) are the surveys [1, 8] and the lecture notes [4, 5].

1 Background

The subspace theorem is a fundamental result proved by Schmidt in 1972. Although he originally developed it in order to prove bounds on the number of solutions of *norm form equations* (which we'll return to later), it has since found applications in a huge range of areas, including extremal combinatorics, which is the topic I'll focus on. However, at the end, I'll touch on some of the more "classical" topics of application, such as Diophantine approximation and number theory.

At an enormously high level, the subspace theorem is a result that says that certain systems of equations, which are in general underdetermined (and thus have infinitely many solutions), will only have finitely many solutions that are *restricted* in some way (e.g. solutions of bounded "size", or solutions in the integers or in some other substructure). Moreover, in many instances, this *qualitative* statement can actually be made *quantitative*, obtaining an effective bound on the number of solutions (usually in a way that depends only on the basic parameters of the problem, as opposed to other features like the specific coefficients involved).

To start with, we will use the following result as our main tool. It is a (highly non-trivial) consequence of the subspace theorem; later on, I will sketch how to derive such a result from the subspace theorem. Such a result was first proved independently by Evertse and by van der Poorten and Schlickewei, though the following theorem is due to Amoroso and Viada, and gives the best known quantitative estimate. Recall that every finitely generated abelian group is isomorphic to $\mathbb{Z}^r \times T$ for some finite group T and some non-negative integer r, which is called its *rank*. I'm stating this result for $\mathbb{C}$, but it actually holds for any algebraically closed field of characteristic 0.

Theorem 1. *Let Γ be a subgroup of $\mathbb{C}^*$ of rank r, and let $a_1, \dots, a_n \in \mathbb{C}$. Then the number of solutions to the equation*

$$a_1 x_1 + \cdots + a_n x_n = 1 \tag{1}$$

with $x_1, \dots, x_n \in \Gamma$ and no subsum on the left-hand side vanishing is at most

$$A(n, r) := (8n)^{4n^4(n+nr+1)}.$$

Solutions $(x_1, \dots, x_n)$ with no vanishing subsum $\sum_{i \in I} a_i x_i = 0$ for $\varnothing \subsetneq I \subsetneq [n]$ are called *non-degenerate solutions*. Note that some sort of non-degeneracy assumption is necessary in order to obtain a finite bound; for instance, if $a_1 = a_3 = 1$ and $a_2 = -1$, then for any $x \in \Gamma$, the vector $(x, x, 1)$ is a solution to (1), so this equation will have infinitely many solutions in general. However, such solutions are ignored in the bound of Theorem 1.

2 "Purely combinatorial" applications

2.1 The sum-product phenomenon

Recall that for a finite subset X of a ring, we denote by $X+X$ and XX, respectively, the sum-set $\{x+x' : x, x' \in X\}$ and the product-set $\{xx' : x, x' \in X\}$. We have the trivial bounds

$$|X| \leq |X+X|, |XX| \leq \frac{|X|^2}{2} + O(|X|).$$

Erdős and Szemerédi initiated the study of the *sum-product phenomenon*, which is a meta-statement of the form "no non-trivial set can have both its sum-set and its product-set be very small". The outstanding open problem here is the following conjecture of Erdős and Szemerédi.

Conjecture 2 (Sum-product conjecture)**.** *For any finite $X \subseteq \mathbb{R}$, we have*

$$\max\{|X+X|, |XX|\} \geq |X|^{2-o(1)}$$

as $|X| \to \infty$.

Note that the $o(1)$ in the exponent is necessary; for instance, if $X = [n]$, then it is easy to see that $|X+X| = 2n-1 = O(|X|)$, while Erdős showed that $|XX| \leq |X|^2/(\log|X|)^{\Omega(1)}$.

In general, Conjecture 2 remains wide open. Major progress was made by Elekes and later by Solymosi, and the current record, due to Rudnev and Stevens, is that

$$\max\{|X+X|, |XX|\} \geq |X|^{\frac{4}{3}+\frac{2}{1167}-o(1)}.$$

However, one can obtain much stronger lower bounds in certain special cases. For instance, Elekes and Ruzsa showed that if $X+X$ is very small (i.e. within a constant factor of its trivial minimum value $2|X|-1$), then XX is very large, namely at least $|X|^{2-o(1)}$. The following result, due to Chang, is the "complement" of this Elekes–Ruzsa theorem (and obtains an even stronger bound): it asserts that if XX is within a constant factor of its trivial lower bound, then $X+X$ is within an additive $O(|X|)$ of its trivial upper bound. Although the sum-product conjecture is usually stated for subsets of $\mathbb{R}$, Chang's theorem actually works even for subsets of $\mathbb{C}$.

Theorem 3 (Chang)**.** *For every $c > 0$, there exists some $C > 0$ such that the following holds. If $X \subseteq \mathbb{C}$ is finite and $|XX| \leq c|X|$, then*

$$|X+X| \geq \frac{|X|^2}{2} - C|X|.$$

Unlike many of the known results about the sum-product phenomenon, such as the Elekes–Ruzsa result mentioned above, Chang's proof does not use any connection to discrete geometry. Instead, she obtains this theorem as a consequence of Theorem 1.

Proof of Theorem 3. First, by deleting at most one element of X, we may assume without loss of generality that $X \subseteq \mathbb{C}^*$, which we will do from now on.

We begin by applying the following result, sometimes called Freiman's lemma (e.g. in Lemma 5.3 of Tao and Vu's *Additive Combinatorics*). It can be seen as a simple special case of the more general Freiman–Ruzsa theorem.

Lemma 4 (Freiman)**.** *If $X \subseteq \mathbb{C}^*$ satisfies $|XX| \leq c|X|$, then there exists a group $\Gamma \subseteq \mathbb{C}^*$ of rank $r \leq c$ such that $X \subseteq \Gamma$.*

We now fix a subgroup $\Gamma \subseteq \mathbb{C}^*$ of rank $r \leq c$ with $X \subseteq \Gamma$. In order to bound $|X+X|$, we will bound the *additive energy* of X. In fact, for technical reasons, it'll be easier to bound the *restricted additive energy*

$$E'(X) = |\{(x_1, x_2, x_3, x_4) \in X^4 : x_1 + x_2 \neq 0 \text{ and } x_1 + x_2 = x_3 + x_4\}|.$$

Note that we have the trivial bounds $2|X|^2 - 2|X| \leq E'(X) \leq |X|^3$. The following simple consequence of Cauchy–Schwarz shows that it suffices to prove that $E'(X)$ is close to its trivial lower bound in order to conclude that $|X + X|$ is close to its trivial upper bound of $|X|^2/2 + O(|X|)$.

Lemma 5. *For any $X \subseteq \mathbb{C}$, we have $|X + X| \geq (|X|^2 - |X|)^2/E'(X)$. In particular, if $E'(X) \leq 2|X|^2 + C|X|$ for some $C \geq 1$, then $|X + X| \geq \frac{1}{2}|X|^2 - C|X|$.*

Proof. We note that

$$|X|^2 = \sum_{z \in X+X} |\{(x_1, x_2) \in X^2 : x_1 + x_2 = z\}| \leq |X| + \sum_{z \in (X+X)\setminus\{0\}} |\{(x_1, x_2) \in X^2 : x_1 + x_2 = z\}|$$

since there are at most $|X|$ pairs $(x_1, x_2) \in X$ summing to 0. By Cauchy–Schwarz, we have that

$$\begin{aligned}
(|X|^2 - |X|)^2 &\leq \left(\sum_{z \in (X+X)\setminus\{0\}} |\{(x_1, x_2) \in X^2 : x_1 + x_2 = z\}| \right)^2 \\
&\leq |X + X| \sum_{z \in (X+X)\setminus\{0\}} |\{(x_1, x_2) \in X^2 : x_1 + x_2 = z\}|^2 \\
&= |X + X| \sum_{z \in (X+X)\setminus\{0\}} |\{(x_1, x_2, x_3, x_4) \in X^4 : x_1 + x_2 = z = x_3 + x_4\}| \\
&= |X + X| E'(X),
\end{aligned}$$

which proves the first claim. If we plug in $E'(X) \leq 2|X|^2 + C|X|$, then

$$\begin{aligned}
|X + X| &\geq \frac{(|X|^2 - |X|)^2}{E'(X)} \geq \frac{(|X| - 1)^2}{2} \cdot \frac{1}{1 + C/(2|X|)} \geq \frac{(|X| - 1)^2}{2} \left(1 - \frac{C}{2|X|}\right) \\
&= \frac{|X|^2}{2} - \frac{C + 4}{4}|X| > \frac{|X|^2}{2} - C|X|,
\end{aligned}$$

using the fact that $1/(1 + y) \geq 1 - y$ for any $y \in \mathbb{R}$, and our assumption that $C \geq 1$. □

Thus, to obtain our desired bound, it suffices to prove that $E'(X) \leq 2|X|^2 + O(|X|)$. Given an equation $x_1+x_2 = x_3+x_4$ with $x_i \in X$ and $x_1+x_2 \neq 0$, we may use our assumption that $0 \notin X$ to rearrange this as

$$\frac{x_1}{x_4} + \frac{x_2}{x_4} - \frac{x_3}{x_4} = 1.$$

If there is a vanishing subsum on the left, then we must have either $x_1 = x_3$ (which yields at most $|X|^2$ solutions), or $x_2 = x_3$ (which yields at most another $|X|^2$ solutions), or $x_1 = -x_2$ (which can't happen by our assumption). If there is no vanishing subsum on the left, then Theorem 1 shows that for any fixed $x_4 \in X$, the number of solutions $x_1, x_2, x_3 \in \Gamma$ is at most $A(3, r)$, so we get at most $A(3, r)|X|$ solutions in total. Since $A(3, r) \leq A(3, c)$, we may set $C = A(3, c) = 24^{1296+972c}$, and find that in total,

$$E'(X) \leq 2|X|^2 + C|X|,$$

which yields the desired bound by Lemma 5. □

2.2 Unit distances

Erdős's unit distance problem asks to understand the function $u(n)$, defined as the maximum number of unit distances among n points in $\mathbb{R}^2$. Equivalently, if we identify $\mathbb{R}^2$ with $\mathbb{C}$, we can define this as

$$u(n) = \max_{S \subseteq \mathbb{C}, |S|=n} \left| \left\{ \{s_1, s_2\} \in \binom{S}{2} : |s_1 - s_2| = 1 \right\} \right|.$$

Erdős gave a very elegant construction showing that $u(n) \geq n^{1+c/\log\log n}$ for some fixed $c > 0$. His construction is to take a $\sqrt{n} \times \sqrt{n}$ grid and then carefully scale it so that there are many unit distances. The appropriate scaling is given by finding an integer $m \in [n]$ that can be represented as the sum of two squares in many ways, and then dilating the grid by a factor of $1/\sqrt{m}$; the precise lower bound on $u(n)$ comes from classical number-theoretic estimates on the number of ways to represent an integer as the sum of two squares. Erdős conjectured that this lower bound is close to optimal, namely that $u(n) \leq n^{1+o(1)}$.

The best known upper bound is $u(n) = O(n^{4/3})$, originally proven by Spencer, Szemerédi, and Trotter. At this point, there are at least three rather different proofs of this bound, and there appears to be a real barrier at the exponent 4/3. It seems that a major obstruction is that all known techniques work if we replace the Euclidean metric by any strictly convex norm on $\mathbb{R}^2$, and there exist such norms for which $\Omega(n^{4/3})$ unit distances are actually attainable by a set of size n. It would be a major breakthrough to prove $u(n) = o(n^{4/3})$.

Because improving the upper bound seems so difficult, some people have focused on restricted problems. A fairly natural one is to try to restrict the directions of the unit distances we are interested in. Namely, for a given set of directions $D \subseteq \{z \in \mathbb{C} : |z| = 1\}$, we define the *restricted unit distances function* by

$$u_D(n) = \max_{S \subseteq \mathbb{C}, |S|=n} \left| \left\{ \{s_1, s_2\} \in \binom{S}{2} : s_1 - s_2 \in D \right\} \right|.$$

Here, let's assume for simplicity that $-D = D$, so that this is well-defined even when considering unordered pairs. Note that if we take D to equal the circle $\{z \in \mathbb{C} : |z| = 1\}$, then we simply recover $u(n)$.

Perhaps the most natural special case, first considered by Schwartz, Solymosi, and de Zeeuw, is when we let $D = \{e^{2\pi i\alpha} : \alpha \in \mathbb{Q}\}$, i.e. when we only consider unit distances which form rational angles with the x axis. This choice of D is a subgroup of $\mathbb{C}^*$ of rank 0 (since every element in it is torsion), which suggests that it may also be fruitful to let D be a multiplicative subgroup of $\mathbb{C}$ of bounded rank. This was done by Schwartz, who proved the following theorem.

Theorem 6 (Solymosi–Schwartz–de Zeeuw for $r = 0$, Schwartz for general r)**.** *For every $\varepsilon > 0$, there exist $c = c(\varepsilon) > 0$ and a positive integer $n_0 = n_0(\varepsilon)$ such that the following holds. For every $n > n_0$ and every subgroup $\Gamma \subseteq \mathbb{C}^*$ of rank $r < c \log n$, we have $u_\Gamma(n) \leq n^{1+\varepsilon}$.*

In other words, Erdős's conjecture $u(n) \leq n^{1+o(1)}$ is true if we restrict ourselves to configurations where all unit distances come from a multiplicative group whose rank is not too large. The reason this result is interesting is that in Erdős's grid-based construction of a lower bound for $u(n)$, it turns out that all the unit distances lie in a subgroup of rank $O(\log n/\log\log n)$. Since it is natural to expect that any extremal construction for $u(n)$ should be "structured" in some way (e.g. look somewhat like a grid), it seems plausible that Schwartz's result might yield the conjectured bound $u(n) \leq n^{1+o(1)}$.

I won't go into too many details, but I will sketch a proof of Theorem 6.

Proof sketch of Theorem 6. Let S be an extremal configuration for $u_\Gamma(n)$. We form a graph G_0 whose vertex set is S, and where two vertices are adjacent if their difference is an element of Γ. Then G_0 has n vertices and $u_\Gamma(n)$ edges.

By repeatedly deleting vertices of low degree, we may pass to a subgraph G with $m \geq \sqrt{n}$ vertices and minimum degree $\Omega(n^\varepsilon)$. We fix some parameter k depending only on ε, to be chosen later. We wish to count the number of *non-degenerate paths* of k edges in G, where we call a path $s_0, s_1, \ldots, s_k$ non-degenerate if $\sum_{i\in I}(s_i - s_{i-1}) \neq 0$ for every $\varnothing \neq I \subseteq [k]$. On the one hand, we can build such a path greedily as follows, using the minimum degree condition. Pick some vertex s_0 of G to be the first vertex. Having picked $s_0, \ldots, s_i$, we have at least $\deg(s_i) - i - 2^i$ options for s_{i+1}, since at most 2^i choices of a neighbor of s_i will cause the path to become degenerate, and at most i neighbors of s_i have already been used in the path. As long as $n^\varepsilon > 2^{k+1}$, all these numbers are $\Omega(n^\varepsilon)$, so in total we have at least $\Omega(n^{k\varepsilon})$ non-degenerate paths of length k in G.

So by picking $k = 3/\varepsilon$ and by averaging over the endpoints, we can find two vertices $u, v \in V(G)$ which have many (say $\Omega(n)$) non-degenerate paths between them. However, if we let $a = v - u$ (treating the vertices as complex numbers again), then we see that every non-degenerate path is a solution to the equation

$$\frac{1}{a}x_1 + \cdots + \frac{1}{a}x_k = 1$$

with no vanishing subsum on the left, and with $x_1, \ldots, x_k \in \Gamma$. If Γ has constant rank r, then Theorem 1 yields a contradiction, since there can be at most $A(k, r) = O(1)$ solutions

to this equation. Moreover, if one works through the numerical dependencies carefully, it is not hard to see that we may even take $r = O_\varepsilon(\log n)$ and obtain the same contradiction. □

2.3 Linear recurrence relations

A *linear recurrence relation of order k* is defined by k complex numbers $c_1, \dots, c_k \in \mathbb{C}$ and k initial values $R_1, \dots, R_k \in \mathbb{C}$, and later values in the sequence are determined inductively by the rule

$$R_n = c_1 R_{n-1} + c_2 R_{n-2} + \cdots + c_k R_{n-k}$$

for every $n > k$. The most well-known example is the Fibonacci sequence, where we take $R_1 = R_2 = c_1 = c_2 = 1$. We will always assume that $c_k \neq 0$, for otherwise the recurrence relation could be taken to have order $k-1$.

We may encode the structure of a linear recurrence relation in matrix form, as

$$\begin{pmatrix} R_{n+k} \\ R_{n+k-1} \\ \vdots \\ R_{n+2} \\ R_{n+1} \end{pmatrix} = \begin{pmatrix} c_1 & c_2 & \cdots & c_{k-1} & c_k \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}^n \begin{pmatrix} R_k \\ R_{k-1} \\ \vdots \\ R_2 \\ R_1 \end{pmatrix}.$$

Let T be the matrix above, called the *transition matrix*. Let its eigenvalues be $\lambda_1, \dots, \lambda_r$, with λ_i having multiplicity m_i, and $\sum m_i = k$. Note that all eigenvalues are non-zero since we assumed $c_k \neq 0$ and thus T is invertible. The recurrence relation is called *simple* if $r = k$, i.e. every eigenvalue has multiplicity 1 and thus $\lambda_1, \dots, \lambda_k$ are all distinct, and it is called *non-degenerate* if $|\lambda_i| \neq |\lambda_j|$ for all $i \neq j$. The importance of the eigenvalues comes from the fact that if the recurrence is simple, then we may diagonalize T and find that

$$R_n = a_1 \lambda_1^n + a_2 \lambda_2^n + \cdots + a_k \lambda_k^n \tag{2}$$

for some constants $a_1, \dots, a_k \in \mathbb{C}$ and all $n \geq 1$.

It is natural to study the *zero set* $Z(R_n)$ of the recurrence relation, that is the set of n such that $R_n = 0$. A famous theorem of Skolem, Mahler, and Lech states that $Z(R_n)$ is the union of a finite set and finitely many arithmetic progressions. Moreover, if the recurrence is non-degenerate, then it is not hard to show that in fact the $Z(R_n)$ is finite; indeed, if it's non-degenerate, then some λ_i has maximum absolute value, and thus $|R_n|$ is asymptotic to $|a_i||\lambda_i|^n$ as $n \to \infty$, and in particular it cannot be zero infinitely often[1]. It was conjectured that in fact, for non-degenerate recurrence relations, $|Z(R_n)|$ should be bounded by a constant depending only on the order k of the recurrence, and this conjecture was proved by Schmidt as a consequence of the subspace theorem. Below is an easier special case, where we restrict to simple recurrences.

[1]This proof sketch only works if the recurrence is simple, but a similar proof works without that assumption, by using a Jordan decomposition rather than a diagonalization as in equation (2).

Theorem 7 (Evertse–Schlickewei–Schmidt)**.** *Let $\{R_n\}$ be a simple, non-degenerate recurrence relation of order k. Then $|Z(R_n)| \leq C_k$, where we may take $C_k = 2^{k^{O(1)}}$*

Proof sketch. By (2), the set $Z(R_n)$ consists of solutions to the equation

$$a_1\lambda_1^n + a_2\lambda_2^n + \cdots + a_k\lambda_k^n = 0 \tag{3}$$

for some constants $a_1, \ldots, a_k$. First, consider those solutions where there is no vanishing subsum on the left-hand side. Letting $x_1 = (\lambda_1/\lambda_k)^n, \ldots, x_{k-1} = (\lambda_{k-1}/\lambda_k)^n$, these correspond to solutions to

$$\left(-\frac{a_1}{a_k}\right) x_1 + \cdots + \left(-\frac{a_{k-1}}{a_k}\right) x_{k-1} = 1,$$

where all x_i lie in the group Γ generated by $\lambda_1, \ldots, \lambda_k$, which has rank at most k. Therefore, by Theorem 1, there are at most $A(k-1,k) \leq 2^{k^{O(1)}}$ such solutions. Moreover, by our non-degeneracy assumption, we can recover n from such a solution $(x_1, \ldots, x_{k-1})$.

On the other hand, suppose that (3) has some vanishing subsum on the left, corresponding to some non-empty subset $I \subseteq [k]$. Then the subsum corresponding to $[k] \setminus I$ must also vanish. Moreover, each of these subsums corresponds to an element of $Z(R'_n)$, for some other recurrence R'_n of order strictly less than k. By induction on k, the total number of times this can happen is at most $C_{|I|}C_{k-|I|} = 2^{k^{O(1)}}$. Adding up over all 2^k choices for I still yields a bound of $2^{k^{O(1)}}$ on the number of such solutions with vanishing subsums. Combining this with the bound above, we get the desired result. □

3 The subspace theorem and its classical applications

So far, I haven't said what the subspace theorem actually is, and we've only used its (quantitative) consequence Theorem 1. In this section, I want to state the "real" subspace theorem, and sketch some of its more standard applications, in the fields of Diophantine approximation, transcendence theory, and counting integer points on varieties.

Here is the original statement of the subspace theorem.

Theorem 8 (Schmidt)**.** *Let $L_1, \ldots, L_n$ be linearly independent linear forms in n variables with algebraic coefficients. For any $\varepsilon > 0$, the set of $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{Z}^n$ satisfying the inequality*

$$|L_1(\boldsymbol{x}) \cdots L_n(\boldsymbol{x})| \leq \frac{1}{(\max|x_i|)^{\varepsilon}} \tag{4}$$

lies in a union of finitely many proper subspaces of $\mathbb{Q}^n$.

Loosely, there can only be finitely many "reasons" (linear relations) for why the polynomial $L_1(\boldsymbol{x}) \cdots L_n(\boldsymbol{x})$ can be "very small" (tending to 0 as a polynomial in the ℓ_∞ norm of $\boldsymbol{x}$) when we restrict $\boldsymbol{x}$ to the integer lattice $\mathbb{Z}^n$.

Note that this is very false if we allow $\boldsymbol{x}$ to be an arbitrary point of $\mathbb{R}^n$. For instance, if $L_1(\boldsymbol{x}) = x_1, L_2(\boldsymbol{x}) = x_2$, then any point on the curve $x_1x_2^2 = 1$ with $x_2 \geq 1$ will satisfy

$|L_1(\boldsymbol{x})L_2(\boldsymbol{x})| = |1/x_2| = (\max|x_i|)^{-1}$, even though this curve is not contained in a linear subspace. Crucially, however, this curve does not have many integer points on it (as it must not, by the subspace theorem).

We can easily find n subspaces where the inequality (4) holds. Namely, each linear form L_i vanishes on some codimension-one subspace V_i, so clearly any integer point on any V_i will yield a solution to (4). One might expect that these are actually the only subspaces where solutions can come from, but that turns out to be false. One example is given for $n = 3$ and

$$L_1(\boldsymbol{x}) = x_1 + \sqrt{2}x_2 + \sqrt{3}x_3 \qquad L_2(\boldsymbol{x}) = x_1 - \sqrt{2}x_2 + \sqrt{3}x_3 \qquad L_3(\boldsymbol{x}) = x_1 - \sqrt{2}x_2 - \sqrt{3}x_3.$$

By Dirichlet's theorem on Diophantine approximation, there eixst infinitely many pairs $(x_1, x_2) \in \mathbb{Z}^2$ such that

$$\left|\sqrt{2} - \frac{x_1}{x_2}\right| \leq |x_2|^{-2},$$

and for such solutions $\max\{|x_1|, |x_2|\} = \Theta(|x_2|)$. Then if we set $x_3 = 0$ and (x_1, x_2) to be such a solution, we can compute that

$$|L_1(\boldsymbol{x})L_2(\boldsymbol{x})L_3(\boldsymbol{x})| = |x_1 + \sqrt{2}x_2||x_1 - \sqrt{2}x_2|^2 \leq |x_1 + \sqrt{2}x_2||x_2|^{-2} = O(\max\{|x_1|, |x_2|\})^{-1},$$

even though $L_1(\boldsymbol{x})L_2(\boldsymbol{x})L_3(\boldsymbol{x}) \neq 0$. In other words, the subspace $x_3 = 0$ yields infinitely many solutions to (4), even though none of the three linear forms L_1, L_2, L_3 vanish on it. This example demonstrates that the finite list of subspaces given by the subspace theorem is highly non-trivial, and contains information about the arithmetic properties of the coefficients of the linear forms (rather than simply linear-algebraic information about where these forms vanish).

The original proof of the subspace theorem did not give any effective bound on how many subspaces one needs. However, several such quantitative results were developed later, none of which I will state. But roughly speaking, they all obtain a bound on the number of subspaces in terms of ε and in terms of the "complexity" of the coefficients of $L_1, \dots, L_n$. These sorts of quantitative estimates are of course needed to deduce other quantitative results like Theorem 1.

The proof of the subspace theorem is quite difficult, and I will not say anything about it.

3.1 Diophantine approximation and Roth's theorem

Recall that Dirichlet proved, as a simple application of the pigeonhole principle, that for any $\alpha \in \mathbb{R}$, there exist infinitely many pairs $(x, y) \in \mathbb{Z}^2$ such that

$$\left|\alpha - \frac{x}{y}\right| \leq \frac{1}{|y|^2}.$$

In other words, any irrational number can be approximated by rationals "quadratically well".

Roth's theorem, a seminal result in Diophantine approximation, says that for *algebraic* numbers, Dirichlet's theorem is essentially best possible.

Theorem 9 (Roth)**.** *Let α be a real algebraic number that is not rational. Then for any $\varepsilon > 0$, there exist finitely many pairs $(x, y) \in \mathbb{Z}^2$ with*

$$\left|\alpha - \frac{x}{y}\right| \leq \frac{1}{|y|^{2+\varepsilon}}. \tag{5}$$

It turns out that Roth's theorem is an easy corollary of the subspace theorem.

Proof of Theorem 9. Given a solution (x, y) to (5), we multiply both sides by $|y|^2$ to find that

$$|y(\alpha y - x)| \leq \frac{1}{|y|^{\varepsilon}}. \tag{6}$$

The linear forms $L_1(x, y) = y$ and $L_2(x, y) = \alpha y - x$ are linearly independent, and they both have algebraic coefficients by our assumption that α is algebraic. So we are essentially in the $n = 2$ setup of the subspace theorem, except that the right-hand side is $|y|^{-\varepsilon}$ as opposed to $\max(|x|, |y|)^{-\varepsilon}$. But this is easily solved since $|x/y| \leq |\alpha| + |y|^{-2} \leq |\alpha| + 1$, which implies that $|x| = O_\alpha(|y|)$. By absorbing the constant factor into the ε in the exponent, we find that the subspace theorem implies that the solutions (x, y) to (6) lie in a finite union of proper subspaces of $\mathbb{Q}^2$. Since the zero-dimensional subspace can't yield any solutions, all these subspaces must be one-dimensional, i.e. the span of some fixed (x_0, y_0). But all integer points (x, y) in such a subspace have $x/y = x_0/y_0$, and thus all have the same value of $|\alpha - x/y|$, which is strictly positive since α is irrational. Since only finitely many of these points will have this value less than $|y|^{-2-\varepsilon}$, we deduce that there are only finitely many solutions to (5). □

Since Roth's theorem is already a difficult and important result, it is perhaps not very surprising that the subspace theorem is hard to prove. In fact, one can view the subspace theorem as a higher-dimensional generalization of Roth's theorem, and Schmidt's proof of the subspace theorem is related to and modeled on Roth's proof of Theorem 9.

3.2 Integer points on varieties and norm form equations

Let $F \in \mathbb{Q}[x, y]$ be a *homogeneous* polynomial in two variables. We're interested in studying integer solutions to the equation $F(x, y) = m$ for some given $m \in \mathbb{Q}$.

If $\deg F = 1$, then it's easy to see that $F(x, y) = m$ has either zero or infinitely many integer solutions, depending on whether m is divisible by the gcd of the coefficients of F, after clearing denominators (this is sometimes called Bézout's theorem[2], and it's an easy consequence of the general Euclidean algorithm).

For $\deg F = 2$, the situation is somewhat more complicated. It is still true (and a well-known exercise in elementary algebraic geometry) that $F(x, y) = m$ has either zero or infinitely many *rational* solutions; in fact, the same holds for any degree-2 equation. However, it may have only finitely many *integer* solutions, such as the equation $x^2 + y^2 = 1$. It may also have infinitely many integer solutions, such as the Pell equation $x^2 - 2y^2 = 1$.

[2]Note that this is unrelated to Bézout's theorem about the number of intersection points of varieties.

It turns out that once $\deg F \geq 3$, the situation becomes much simpler. Indeed, Thue proved that if $F \in \mathbb{Q}[x, y]$ is an irreducible homogeneous polynomial of degree $d \geq 3$, then the equation $F(x, y) = m$ has only finitely many integer solutions, for any $m \in \mathbb{Q}$.

(It's worth remarking here about Faltings's famous theorem, proving the well-known Mordell conjecture. This theorem says that any smooth curve of genus $g > 1$ defined over $\mathbb{Q}$ has only finitely many rational points. Roughly speaking, a curve of high degree should have high genus, and vice versa, which suggests that Faltings's theorem is more general than Thue's. However, as far as I can tell, there is probably no formal reduction between them.)

Schmidt originally developed the subspace theorem to prove a generalization of Thue's theorem for homogeneous polynomials in more than 2 variables. To motivate this class of equations, known as *norm form equations*, let's return briefly to the setup of Thue's theorem. Let's assume without loss of generality that the coefficient of x^d in $F(x, y)$ is 1, and let $f(x) = F(x, 1)$. By the irreducibility of F, we can see that f has distinct roots in $\mathbb{C}$, say $\rho_1, \dots, \rho_d$. But then we may write $f(x) = (x - \rho_1) \cdots (x - \rho_d)$, which shows that

$$F(x, y) = \prod_{i=1}^{d} (x - \rho_i y). \tag{7}$$

Moreover, if we let K be the splitting field of f over $\mathbb{Q}$, then the numbers $\rho_1, \dots, \rho_d$ are all Galois conjugates in K. If $x, y \in \mathbb{Q}$, then they are fixed by all elements of the Galois group of K over $\mathbb{Q}$. So in short, we see that (7) is expressing $F(x, y)$ as a product of all Galois conjugates of $x - \rho_1 y$.

Recall that the field norm $N_{K/\mathbb{Q}} : K \to \mathbb{Q}$ is a multiplicative function defined by multiplying together all the Galois conjugates of an element of K. From this and the above, we see that we can equivalently express the polynomial $F(x, y)$ as $F(x, y) = N_{K/\mathbb{Q}}(x - \rho_1 y)$. Then the Thue equation $F(x, y) = m$ becomes the norm form equation $N_{K/\mathbb{Q}}(x - \rho_1 y) = m$, and we are searching for integer solutions x, y.

The natural generalization of this setup, which was considered by Schmidt, is the following. Let $K/\mathbb{Q}$ be a Galois extension of degree d, and let $\sigma_1, \dots, \sigma_d$ be the elements of the Galois group of $K/\mathbb{Q}$. For some $1 \leq n \leq d$, fix elements $\alpha_1, \dots, \alpha_n \in K$ which are linearly independent over $\mathbb{Q}$ (note that we need $n \leq d$ since the dimension of K as a $\mathbb{Q}$-vector space is d). Then we can define the associated *norm form* by

$$F(x_1, \dots, x_n) = N_{K/\mathbb{Q}}(\alpha_1 x_1 + \cdots + \alpha_n x_n) = \prod_{i=1}^{d} [\sigma_i(\alpha_1) x_1 + \cdots + \sigma_i(\alpha_n) x_n].$$

Then F is a homogeneous polynomial of degree d. Moreover, we see that every element of the Galois group of $K/\mathbb{Q}$ permutes the linear factors of F, which shows that the coefficients of F are in $\mathbb{Q}$. The norm form equation we will be interested in studying is the equation $F(x_1, \dots, x_n) = m$ for some fixed $m \in \mathbb{Q}$, and we are searching for solutions $(x_1, \dots, x_n) \in \mathbb{Z}^n$.

Schmidt proved a necessary and sufficient condition for the equation $F(x_1, \dots, x_n) = m$ to have finitely many integer solutions $(x_1, \dots, x_n) \in \mathbb{Z}^n$. Roughly speaking, it is "usually" the

case that there are only finitely many solutions—the only way for infinitely many solutions to arise is if there is some hidden algebraic structure that yields them. I'll give a precise statement of this shortly, but first I want to explain how the subspace theorem can be used to count solutions to such norm form equations. It will be convenient to have a slightly more general form of the subspace theorem, which can be proved from Theorem 8 by a simple induction argument.

Corollary 10 (Subspace theorem for more forms than variables)**.** *Let $1 \leq n \leq d$, and let $L_1, \dots, L_d$ be linear forms in n variables that are in general position (i.e. any n-subset is linearly independent) with algebraic coefficients. For any $\varepsilon > 0$ and $C > 0$, the set of $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{Z}^n$ satisfying the inequality*

$$|L_1(\boldsymbol{x}) \cdots L_d(\boldsymbol{x})| \leq C \left(\max|x_i|\right)^{d-n-\varepsilon}$$

lies in a union of finitely many proper subspaces of $\mathbb{Q}^n$.

Given this, we can try to prove that $N_{K/\mathbb{Q}}(\alpha_1 x_1 + \cdots + \alpha_n x_n) = m$ has finitely many solutions, as follows. We induct on n, with the $n = 1$ case being trivial since any degree-d polynomial in one variable has at most d roots, and thus there are only a finite number of solutions. For the inductive step, let us write $L_i(x_1, \dots, x_n) = \sigma_i(\alpha_i)x_i + \cdots + \sigma_i(\alpha_n)x_n$ for $i \in [d]$. The linear forms $L_1, \dots, L_d$ certainly have algebraic coefficients, and let's assume for a moment that they're in general position (which certainly seems plausible, since we picked $\alpha_1, \dots, \alpha_n$ to be linearly independent). Moreover, let's assume for simplicity that $n < d$.

Now, we claim that the solutions $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{Z}^n$ to the norm form equation $N_{K/\mathbb{Q}}(\alpha_1 x_1 + \cdots + \alpha_n x_n) = m$ lie in finitely many proper subspaces of $\mathbb{Q}^n$. Indeed, any such solution is either the zero vector or satisfies

$$|L_1(\boldsymbol{x}) \cdots L_d(\boldsymbol{x})| = |N_{K/\mathbb{Q}}(\alpha_1 x_1 + \cdots + \alpha_n x_n)| = |m| \leq |m|(\max|x_i|)^{d-n-\frac{1}{2}},$$

by our assumption that $n < d$. Applying Corollary 10 with $C = |m|$ and $\varepsilon = \frac{1}{2}$, we find that the solutions to the norm form equation lie in finitely many subspaces of $\mathbb{Q}^n$. If V is such a subspace, then on V, some linear combination of the x_i vanishes. Without loss of generality, this means that we may write $x_n = c_1 x_1 + \cdots + c_{n-1}x_{n-1}$ as a linear combination of $x_1, \dots, x_{n-1}$. So if we define $\beta_1 = \alpha_1 + c_1\alpha_n$, then we see that on V, we can reduce to the norm form equation $N_{K/\mathbb{Q}}(\beta_1 x_1 + \cdots + \beta_{n-1}x_{n-1}) = m$, which has $n-1$ variables. By the inductive assumption, there are finitely many solutions to such an equation (since the β_i are also linearly independent). Since there were only finitely many options for V, this proves that there are finitely many solutions in total.

We made two assumptions in this proof sketch. The first was that $n < d$, which turns out to not really matter and can be eliminated with a bit more work. The second was that the forms $L_1, \dots, L_d$ are in general position. However, this turns out to just be false in full generality, as must be the case since there *do* exist norm form equations with infinitely many solutions. In fact, we already saw such an example, the Pell equation $x^2 - 2y^2 = 1$.

What is the general phenomenon here? Recall that every algebraic extension K of $\mathbb{Q}$ has a *ring of integers* $\mathcal{O}_K$, consisting of all elements of K that are the roots of monic

polynomials with integer coefficients. It is a well-known fact that an element $\varepsilon \in \mathcal{O}_K$ is a unit of the ring $\mathcal{O}_K$ if and only if $N_{K/\mathbb{Q}}(\varepsilon) = \pm 1$. Moreover, an important result in algebraic number theory is Dirichlet's unit theorem, which implies that $\mathcal{O}_K$ will have infinitely many units unless K is equal to $\mathbb{Q}$ or to an imaginary quadratic field (i.e. equal to $\mathbb{Q}(\sqrt{d})$ for some integer $d < 0$). Therefore, if $\mathcal{O}_K$ is a subset of the lattice $\Lambda = \{\alpha_1 x_1 + \cdots + \alpha_n x_n : x_1, \dots, x_n \in \mathbb{Z}\}$, then there will be infinitely many solutions to the norm form equation $N_{K/\mathbb{Q}}(\alpha_1 x_1 + \cdots + \alpha_n x_n) = 1$. More generally, essentially the same argument shows that if there is a subfield $\mathbb{Q} \subsetneq L \subseteq K$ such that L is not imaginary quadratic, and if there exists some $\gamma \in K^*$ such that $\gamma \mathcal{O}_L \subseteq \Lambda$, then there will be infinitely many solutions to the norm form equation $N_{K/\mathbb{Q}}(\alpha_1 x_1 + \cdots + \alpha_n x_n) = m$, where $m = N_{K/\mathbb{Q}}(\gamma)$. Schmidt proved that this is in fact the only obstruction.

Theorem 11 (Schmidt)**.** *Let $K/\mathbb{Q}$ be a finite extension of degree d, and let $\alpha_1, \dots, \alpha_n$ be linearly independent elements of K. Then the following are equivalent.*

1. *There do not exist $\gamma \in K^*$ and a subfield $\mathbb{Q} \subsetneq L \subseteq K$ which is not imaginary quadratic such that $\gamma \mathcal{O}_L \subseteq \{\alpha_1 x_1 + \cdots + \alpha_n x_n : x_1, \dots, x_n \in \mathbb{Z}\}$.*
2. *For every $m \in \mathbb{Q}$, there are only finitely many integer solutions $x_1, \dots, x_n$ to the norm form equation $N_{K/\mathbb{Q}}(\alpha_1 x_1 + \cdots + \alpha_n x_n) = m$.*

3.3 Transcendence theory

One of the earliest results in Diophantine approximation is Liouville's theorem (a very weak form of Roth's theorem), which Liouville developed in order to prove that certain explicit real numbers are transcendental. Given this, it is not surprising that stronger results in Diophantine approximation like the subspace theorem can be used to prove that larger classes of number are transcendental.

Let's say that an infinite sequence $u_1, u_2, \dots$ has *long repetitions* if there exists some $\varepsilon > 0$ such that the word $u_1 u_2 \dots u_N$ has two disjoint equal subwords of length εN, for infinitely many choices of N. Similarly, for any $b \geq 2$, we say that a real number $\alpha \in (0, 1)$ has *b-ary long repetitions* if its expansion in base b has long repetitions. For instance, if α is rational, then its b-ary expansion is eventually periodic, and thus has long repetititions.

Theorem 12 (Adamczewski–Bugeaud–Luca)**.** *Let $b \geq 2$ and $\alpha \in (0, 1)$. If α has b-ary long repetitions, then α is either rational or transcendental.*

The proof of Theorem 12 uses a version of the subspace theorem, as I'll soon sketch. But for the moment, let's see some applications. One easy consequence is a new proof that Liouville numbers are transcendental, as in the following result.

Corollary 13. *The number $\alpha = \sum_{n \geq 1} 2^{-2^n}$ is transcendental.*

Proof. The binary representation of α is $0.0101000100000001\dots$. This is never periodic, so α is irrational. Moreover, the initial segment of length N has two disjoint blocks of 0 of length at least $N/8$, so α has binary long repetitions. So by Theorem 12, it must be transcendental. □

In fact, using Theorem 12, Adamczewski and Bugeaud proved a major generalization of this fact, which is that every *automatic number* is either rational or transcendental. An automatic number is one whose b-ary representation (for some b) is the output of some finite automaton, when we give it as input the sequence of binary representations of the natural numbers. This follows from Theorem 12, since automatic numbers are known to have b-ary long repetitions, which is hopefully intuitive: a fixed finite automaton can only generate a sequence of "bounded complexity", so we would expect to have long repeated blocks if we run it on the sequence of binary representations of the natural numbers.

Moreover, Theorem 12 can also be used to say interesting things about *algebraic* numbers. Recall that a real number α is called *b-normal* if every pattern appears in its b-ary expansion with equal asymptotic density, i.e. every word in $[b]^n$ appears in the b-ary expansion of α a b^{-n} fraction of the time, for every $n \geq 1$. Moreover, α is called *normal* if it is b-normal for every $b \geq 2$.

Although it is easy to prove that almost every real number is normal, there are very few explicit examples of normal numbers. However, it has long been conjectured that every irrational algebraic number is normal. This conjecture appears totally out of reach at the moment, but Adamczewski and Bugeaud were able to prove a (very) weak version of it using Theorem 12. Given $\alpha \in (0,1)$ and integers n and $b \geq 2$, we define the *complexity function* $\rho(\alpha, b, n)$ as the number of words in $[b]^n$ that appear in the b-ary representation of α. So $1 \leq \rho(\alpha, b, n) \leq b^n$. Moreover, having $\rho(\alpha, b, n) = b^n$ is a weakened version of normality (where every word in $[b]^n$ appears in the b-ary expansion of α, but we say nothing about its density).

Theorem 14 (Adamczewski–Bugeaud)**.** *Let $b \geq 2$ and $\alpha \in (0,1)$. If α is irrational and algebraic, then*

$$\lim_{n\to\infty} \frac{\rho(\alpha, b, n)}{n} = \infty.$$

Again, the truth is probably that $\rho(\alpha, b, n) = b^n$, so this superlinearity result is not so shocking, but it is one of the strongest results we can prove in the direction of showing that irrational algebraic numbers are normal.

Theorem 14 is actually a pretty simple consequence of Theorem 12. Indeed, it's fairly easy to show that if $\rho(\alpha, b, n) = O(n)$, then α has b-ary long repetitions: only $O(n)$ patterns of length n appear in any initial segment, so one can argue that two disjoint ones have to be equal. But if α has b-ary long repetitions then it must be rational or transcendental by Theorem 12, and the contrapositive yields Theorem 14.

To prove Theorem 12, we will need a p-adic extension of the subspace theorem, originally due to Schlickewei (and in fact, most applications of the subspace theorem need this generalization). Recall that for a prime p, the *p-adic norm* $|\cdot|_p$ on $\mathbb{Q}$ is defined by $|x|_p = p^{-v_p(x)}$, where $v_p(x)$ is the largest power of p which divides x (which is taken to be negative if there are powers of p in the denominator of x). An important (and easy) fact is the *product formula*, which says that for any $x \in \mathbb{Q}$,

$$|x| \prod_p |x|_p = 1,$$

where the product runs over all primes p. Note that $|x|_p \neq 1$ for only finitely many primes p, so that this infinite product is really finite and there are no convergence issues.

With these preliminaries, we can state Schlickewei's p-adic extension of the subspace theorem. It is usually stated for n linearly independent forms in n variables, but for convenience, we state the following more general result where the number of forms can be larger than the number of variables, just like Corollary 10.

Theorem 15 (Schlickewei)**.** *Let $n \geq 2$ and $C, \varepsilon > 0$, and let $p_1, \dots, p_s$ be distinct primes. Let $L_{1,\infty}, \dots, L_{d,\infty}$ (with $d \geq n$) be linear forms in n variables with algebraic coefficients that are in general position. Similarly, for each $j \in [s]$, let $L_{1,p_j}, \dots, L_{d_j,p_j}$ (with $d_j \geq n$) be a collection of linear forms in n variables that are in general position and have algebraic coefficients. Then the set of solutions $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{Z}^n$ to the inequality*

$$|L_{1,\infty}(\boldsymbol{x}) \cdots L_{d,\infty}(\boldsymbol{x})| \prod_{j=1}^{s} |L_{1,p_j}(\boldsymbol{x}) \cdots L_{d_j,p_j}(\boldsymbol{x})|_{p_j} \leq (\max|x_i|)^{d-n-\varepsilon}$$

lies in a union of finitely many proper subspaces of $\mathbb{Q}^n$.

(One subtlety I've swept under the rug is the fact that I haven't defined $|\cdot|_p$ for algebraic numbers, so it's not clear that quantities like $|L_{1,p_j}(\boldsymbol{x})|_{p_j}$ make sense. One way of getting around this is to insist that the coefficients of each L_{i,p_j} lie in $\overline{\mathbb{Q}} \cap \mathbb{Q}_{p_j}$, which is indeed what will always happen in our applications (in fact, these coefficients will just be rational). However, it turns out that one can also just pick an extension of the p-adic norm to $\overline{\mathbb{Q}}$, which is perhaps conceptually easier to think about.)

Once we have the p-adic subspace theorem, it is not too hard to prove Theorem 12.

Proof of Theorem 12. Let $\alpha \in (0,1)$, and let's assume that α is algebraic; our goal is to prove that in this case, it is rational. Since α has b-ary long repetitions, we can pick some fixed $\varepsilon > 0$ and some N such that the first N b-ary digits of α have disjoint identical blocks of length εN. This means that the b-ary expansion of α starts $0.ABCB$, where A and C may be empty strings, and the length of B, denoted $\ell(B)$, is εN.

The goal is now to pick a rational number that approximates α very well, and to derive a contradiction if they are actually distinct. Namely, we take ξ to be the number with b-ary expansion $0.ABCBCBCBC\dots$, and note that ξ is rational since it has an eventually periodic b-ary expansion. Let $r = \ell(A)$ and $s = \ell(BC)$. Then $b^r\xi = A.BCBC\dots$ and $b^r b^s \xi = ABC.BCBC\dots$, which shows that $b^r(b^s - 1)\xi = ABC.0 - A.0$ is an integer. Therefore, we may write

$$\xi = \frac{M}{b^r(b^s-1)} = \frac{M}{b^{r+s} - b^r}$$

for some $M \in \mathbb{Z}$. Moreover, since ξ agrees with α on the first $\ell(ABCB) = r + s + \ell(B) = r + s + \varepsilon N$ b-ary digits, we have that

$$|\alpha - \xi| \leq b^{-r-s-\varepsilon N}.$$

Combining the last two equations, we find that

$$|b^{r+s}\alpha - b^r\alpha + M| = (b^{r+s} - b^r)|\alpha - \xi| \leq b^{r+s}|\alpha - \xi| \leq b^{-\varepsilon N}.$$

Let $p_1, \dots, p_s$ be the set of primes dividing b, and let $L_{i,p_j}(x_1, x_2, x_3) = x_i$ for every $i \in [3]$ and $j \in [s]$. Additionally, let

$$L_{1,\infty}(x_1, x_2, x_3) = x_1, \qquad L_{2,\infty}(x_1, x_2, x_3) = x_2, \qquad L_{3,\infty}(x_1, x_2, x_3) = \alpha x_1 - \alpha x_2 - x_3.$$

Since α is algebraic and since each triple of linear forms is linearly independent, we are in a position to apply Theorem 15.

Now, let $\boldsymbol{x} = (b^{r+s}, b^r, M)$, and note that $|M| \leq b^{r+s}$ since $\xi \in (0, 1)$. Therefore, we have that $\max|x_i| = b^{r+s} \leq b^N$. Additionally, by our choices of linear forms, we have that

$$\begin{aligned}|L_{1,\infty}(\boldsymbol{x})L_{2,\infty}(\boldsymbol{x})L_{3,\infty}(\boldsymbol{x})| \prod_{j=1}^{s} |L_{1,p_j}(\boldsymbol{x})L_{2,p_j}(\boldsymbol{x})L_{3,p_j}(\boldsymbol{x})|_{p_j} = \\ = \left(|b^r| \prod_{j=1}^{s} |b^r|_{p_j}\right) \left(|b^{r+s}| \prod_{j=1}^{s} |b^{r+s}|_{p_j}\right) |\alpha b^{r+s} - \alpha b^s - M| \left(\prod_{j=1}^{s} |M|_{p_j}\right)\end{aligned}$$

Note that the first two factors equal 1 by the product formula and the choice of $p_1, \dots, p_s$ as the set of primes dividing b. Moreover, since $M \in \mathbb{Z}$, we have that $|M|_{p_j} \leq 1$ for every p_j. So we find that

$$\begin{aligned}|L_{1,\infty}(\boldsymbol{x})L_{2,\infty}(\boldsymbol{x})L_{3,\infty}(\boldsymbol{x})| \prod_{j=1}^{s} |L_{1,p_j}(\boldsymbol{x})L_{2,p_j}(\boldsymbol{x})L_{3,p_j}(\boldsymbol{x})|_{p_j} &\leq |\alpha b^{r+s} - \alpha b^s - M| \\ &\leq b^{-\varepsilon N} \\ &\leq (\max|x_i|)^{-\varepsilon} \qquad (8)\end{aligned}$$

by our computations above.

Note that this argument worked for any N such that the initial segment of length N of the b-ary expansion of α has long repeated blocks. By assumption, there are infinitely many such N. Therefore, we may construct an infinite sequence of vectors $\boldsymbol{x}^{(N)} = (b^{r(N)+s(N)}, b^{r(N)}, M(N))$, with r, s, M defined as above. Infinitely many of these vectors will actually be distinct, since $s(N) \geq \varepsilon N$ tends to infinity with N, and thus we'll see infinitely many distinct first coordinates of $\boldsymbol{x}^{(N)}$. For all of them, (8) will hold with the same value of ε (since it's the value given by the long repetititions assumption). Therefore, by Theorem 15, this set of $\boldsymbol{x}^{(N)}$ must lie in a finite union of proper subspaces of $\mathbb{Q}^3$, and in particular some infinite subsequence must lie in a single proper subspace. So that means that there exist rational numbers β, γ, δ, not all equal to 0, such that for infinitely many N,

$$\beta b^{r(N)} + \gamma b^{r(N)+s(N)} + \delta M(N) = 0$$

If $\delta = 0$, then we could factor out $b^{r(N)}$ from this equation and obtain a contradiction since $s(N) \to \infty$ and $b \geq 2$. If we divide by $b^{r(N)}(b^{s(N)} - 1)$, then we find that

$$\beta \frac{1}{b^{s(N)} - 1} + \gamma \frac{b^{s(N)}}{b^{s(N)} - 1} + \delta \xi(N) = 0$$

If we let $N \to \infty$ (and thus $s(N) \to \infty$), then the first term vanishes and the second term tends to γ, while $\xi(N) \to \alpha$ by our definition of ξ as a good approximation of α. So in the limit, we find that $\gamma + \delta\alpha = 0$, which means that $\alpha = -\gamma/\delta \in \mathbb{Q}$, as desired. □

3.4 Back to the beginning

In all our earlier applications, we didn't use the subspace theorem, but rather its consequence Theorem 1. In this section, I want to sketch how the subspace theorem (or rather, its p-adic version, Theorem 15) can be used to prove such a result. Recall that we fix $a_1, \dots, a_n \in \mathbb{C}$ and group $\Gamma \subset \mathbb{C}^*$ of rank r, and are trying to count solutions to

$$a_1 x_1 + \cdots + a_n x_n = 1 \tag{9}$$

with $x_1, \dots, x_n \in \Gamma$ and no subsum on the left-hand side vanishing.

For simplicity, let's assume that $a_1, \dots, a_n$ are algebraic and our group Γ is actually a subgroup of $\mathbb{Q}^*$, meaning that searching for solutions $x_i \in \Gamma$ is the same as searching for solutions $x_i \in \mathbb{Q}$ whose numerators and denominators are only divisible by a fixed set $p_1, \dots, p_r$ of primes, where r is the rank of Γ. (This is not a huge simplifying assumption: one can reduce from the case the general case to the case of $\Gamma \subset \overline{\mathbb{Q}}^*$ and algebraic coefficients using a straightforward argument from algebraic geometry. It then suffices to work in a fixed number field containing Γ and $a_1, \dots, a_n$, and it turns out that the entire machinery of the subspace theorem works over number fields as well as over $\mathbb{Q}$.)

We apply Theorem 15 with $d = n + 1$ and the linear forms

$$L_{1,\infty}(\boldsymbol{x}) = L_{1,p_j}(\boldsymbol{x}) = x_1, \qquad L_{2,\infty}(\boldsymbol{x}) = L_{2,p_j}(\boldsymbol{x}) = x_2, \quad \dots \quad L_{n,\infty}(\boldsymbol{x}) = L_{n,p_j}(\boldsymbol{x}) = x_n$$

and

$$L_{n+1,\infty}(\boldsymbol{x}) = L_{n+1,p_j}(\boldsymbol{x}) = a_1 x_1 + \cdots + a_n x_n,$$

and these forms are in general position because we may assume without loss of generality that all a_i are non-zero. Moreover, by the product formula, we see that any solution $\boldsymbol{x} = (x_1, \dots, x_n) \in \Gamma^n$ to (9) satisfies

$$|L_{1,\infty}(\boldsymbol{x}) \cdots L_{n+1,\infty}(\boldsymbol{x})| \prod_{j=1}^{s} |L_{1,p_j}(\boldsymbol{x}) \cdots L_{n+1,p_j}(\boldsymbol{x})|_{p_j} = 1.$$

We can't quite apply Theorem 15 yet, because the vector $\boldsymbol{x}$ does not lie in $\mathbb{Z}^n$. However, if we let D be the least common denominator of $x_1, \dots, x_n$ and observe that D is an integer divisible only by $p_1, \dots, p_s$, we see that we can set $\boldsymbol{y} = D\boldsymbol{x} \in \mathbb{Z}^n$ and have that

$$|L_{1,\infty}(\boldsymbol{y}) \cdots L_{n+1,\infty}(\boldsymbol{y})| \prod_{j=1}^{s} |L_{1,p_j}(\boldsymbol{y}) \cdots L_{n+1,p_j}(\boldsymbol{y})|_{p_j} = 1 \leq (\max |y_i|)^{\frac{1}{2}} = (\max |y_i|)^{d-n-\frac{1}{2}}.$$

So by Theorem 15, all such solutions $\boldsymbol{y}$ lie in a finite union of proper subspaces of $\mathbb{Q}^n$, and therefore the same is true of solutions $\boldsymbol{x}$ to (9). On any such subspace, we have a linear relation among $x_1, \dots, x_n$, which means that we can replace one of them by a linear combination of the others and obtain (upon some rearrangement) another equation of the form (9), except this time with $n-1$ variables. Moreover, one can check that vanishing subsums in the original equation correspond to vanishing subsums in the new equation, which means that by induction, there are only finitely many non-degenerate solutions coming from any such subspace. Since there are only finitely many such subspaces, we deduce the desired result.

4 Magic

In this section, I just want to remark on a few other applications of the subspace theorem that I find totally amazing. I won't really get into any of the proofs, because it's somewhat far afield from the main topic of the talk.

4.1 Sub-exponential GCDs

It is an easy fact that if b is a power of a, then $a^n - 1$ divides $b^n - 1$ for every $n \geq 1$. Indeed, if $b = a^k$, then $b^n - 1 = a^{nk-1} = (a^n - 1)(a^{n(k-1)} + a^{n(k-2)} + \cdots + a^n + 1)$. Moreover, it turns out[3] that this is an if and only if condition: if $a^n - 1$ divides $b^n - 1$ for all $n \geq 1$, then b is a power of a.

Similarly, if a and b are both powers of some c, then $\gcd(a^n - 1, b^n - 1) \geq c^n - 1$, for the same reason. However, it is reasonable to expect that apart from this "obvious" obstruction, we should have that $\gcd(a^n - 1, b^n - 1)$ is fairly small as $n \to \infty$; indeed, the set of primes dividing $a^n - 1$ and $b^n - 1$ should have "little to do with one another", and thus the gcd should be small.

This vague intuition is correct, as shown by the following remarkable theorem of Bugeaud, Corvaja, and Zannier, whose proof uses the subspace theorem.

Theorem 16 (Bugeaud–Corvaja–Zannier)**.** *Suppose that a and b are positive integers that are not both powers of some fixed c. Then $\gcd(a^n - 1, b^n - 1) = 2^{o(n)}$. In other words, for every $\varepsilon > 0$, there exists some $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that $\gcd(a^n - 1, b^n - 1) \leq 2^{\varepsilon n}$ for all $n \geq n_0$.*

Their proof is ineffective, meaning that they get no control on how fast the $o(n)$ term tends to 0 (or equivalently, on how large $n_0(\varepsilon)$ is). I believe that one should be able to make their theorem effective by using a quantitative version of the subspace theorem, but it is possible that there are some subtleties that I'm missing.

How good is this barely sub-exponential upper bound? Well, we always have that $\gcd(a-1, b-1)$ divides $\gcd(a^n - 1, b^n - 1)$. Ailon and Rudnick conjectured that this bound is

[3]Bugeaud, Corvaja, and Zannier say that this "is a known amusing elementary problem", but I actually have no idea how to prove it.

tight infinitely often, i.e. that there exist infinitely many n such that $\gcd(a^n-1, b^n-1) = \gcd(a-1, b-1)$. However, one can also show that $\gcd(a^n-1, b^n-1)$ is quite large for some n. Namely, there exist infinitely many n such that

$$\gcd(a^n-1, b^n-1) = 2^{2^{\Omega(\log n/\log\log n)}}.$$

Since $2^{\log n/\log\log n}$ is just barely sub-polynomial, this lower bound is barely sub-exponential, which shows that Theorem 16 is in some sense fairly close to best possible. To see how to get such a lower bound, a fairly simple argument in analytic number theory shows that we can pick infinitely many integers n such that n is divisible by $p-1$ for at least $2^{\Omega(\log n/\log\log n)}$ choices of a prime p. By Fermat's little theorem, for any such p, we have that a^n-1 and b^n-1 are both divisible by p. Therefore $\gcd(a^n-1, b^n-1)$ is at least the product over all such p, which is at least $2^{2^{\Omega(\log n/\log\log n)}}$ since each such p is at least 2.

I won't say much about the proof of Theorem 16, except that it's somewhat similar to the proof of Theorem 12 which I presented above. Let d_n be the denominator of $(b^n-1)/(a^n-1)$, and assume for contradiction that $d_n \le a^{(1-\varepsilon)n}$ for infinitely many n (which is equivalent to $\gcd(a^n-1, b^n-1) \ge a^{\varepsilon n}$). For $j \ge 1$, let

$$z_j(n) = \frac{b^{jn}-1}{a^n-1},$$

and note that the denominator of $z_j(n)$ is also d_n since $b^{jn}-1$ is divisible by b^n-1. We use the geometric series to approximate

$$\frac{1}{a^n-1} = \sum_{i=1}^{\infty} \frac{1}{a^{in}} \approx \sum_{i=1}^{M} \frac{1}{a^{in}}$$

for some large cutoff M. By multiplying both sides by $b^{jn}-1$ and rearranging, we find that

$$z_j(n) \approx \sum_{i=1}^{M} \left(\frac{b^j}{a^i}\right)^n - \sum_{i=1}^{M} \frac{1}{a^{in}} \qquad \Longrightarrow \qquad \left| z_j(n) - \sum_{i=1}^{M} \left(\frac{b^j}{a^i}\right)^n + \sum_{i=1}^{M} \frac{1}{a^{in}} \right| \approx 0$$

This says that a linear form in the variables $z_j(n), (b^j/a^1)^n, \dots, (b^j/a^M)^n, 1/a^n, \dots, 1/a^{Mn}$ is small, and this holds for every fixed j. By applying this for many j at once, we get many linear forms whose evaluation is small. Additionally, we define linear forms for every prime dividing ab, with the ith linear form just equalling the ith variable. Then all of these linear forms are p-adically bounded by d_n, since the only denominators that can appear in our variables are the denominators d_n in $z_j(n)$. Working all of this out carefully, one can check that we are in a position to apply Theorem 15, and to conclude that all the vectors we produce from this operation (when we run over all n for which $\gcd(a^n-1, b^n-1)$ is large) lie in a finite union of proper subspaces. Again by passing to a subsequence which all lies in a single subspace, we can then derive a contradiction to our assumption that a and b were not both powers of a fixed c.

4.2 Nearby irreducible polynomials

A consequence of Hilbert's irreducibility theorem, first observed by van der Waerden, is that if P is a polynomial whose coefficients are randomly chosen integers in $[-N, N]$, then P will be irreducible over $\mathbb{Q}$ with high probability as $N \to \infty$. In other words, almost all integer polynomials are irreducible.

It is natural to wonder if there is a "local" version of this fact, i.e. if every integer polynomial is "close" to an irreducible one. One version of this question was posed by Turán, who asked whether there exists a constant C such that for every polynomial $P \in \mathbb{Z}[x]$, there exists some *irreducible* $Q \in \mathbb{Z}[x]$ such that $\deg Q \leq \deg P$ and such that the sum of the absolute differences of the coefficients of P and Q is at most C. In other words, Turán asked if we can perturb the coefficients of P by at most C in order to make it irreducible, while also ensuring that we don't raise the degree. If we drop the degree condition, then Schinzel proved that we may take $C = 3$: for every $P \in \mathbb{Z}[x]$, one of the polynomials $x^m + x^n + P(x)$ and $x^m + x^n + P(x) + 1$ is irreducible for infinitely many choices of m and n. However, the full version of Turán's problem is still open.

In a different direction, Szegedy asked whether we can find a nearby irreducible polynomial by changing only the constant coefficient of P. Namely, he asked whether for every $d \geq 1$, there exists some $C_d > 0$ such that for every $P \in \mathbb{Z}[x]$ of degree d, one of the shifts $P(x) - C_d, P_x - C_d + 1, \dots, P(x) + C_d - 1, P(x) + C_d$ is irreducible. While this question remains open, Győry was able to prove it for *monic* polynomials. In fact, he proved the following stronger result.

Theorem 17 (Győry)**.** *For all integers $a, d \geq 0$, there exists a constant $C_{a,d}$ such that the following holds. If $P \in \mathbb{Z}[x]$ has leading coefficient a and degree d, then there is some integer $b \in [-C_{a,d}, C_{a,d}]$ such that $P(x) + b$ is irreducible.*

The result about monic polynomials follows from setting $a = 1$ for all d, while Szegedy's conjecture is that one may remove the dependence on a entirely. I won't say much about the proof of Győry's result; at a high level, he uses the theory of resultants to understand the set of b for which $P(x)+b$ is reducible, and then uses a consequence of the subspace theorem to conclude that the relevant resultant equations have a bounded number of solutions. This implies that if $C_{a,d}$ is sufficiently large, some $b \in [-C_{a,d}, C_{a,d}]$ cannot be a solution, and hence $P(x) + b$ must be irreducible.

References

[1] Y. F. Bilu, The many faces of the subspace theorem [after Adamczewski, Bugeaud, Corvaja, Zannier...], *Astérisque* **317** (2008), 1–38. Séminaire Bourbaki. Vol. 2006/2007.

[2] Y. Bugeaud, Quantitative versions of the subspace theorem and applications, *J. Théor. Nombres Bordeaux* **23** (2011), 35–57.

[3] Y. Bugeaud, P. Corvaja, and U. Zannier, An upper bound for the G.C.D. of $a^n - 1$ and $b^n - 1$, *Math. Z.* **243** (2003), 79–84.

[4] J.-H. Evertse, The subspace theorem, 2011. Lecture notes available at `http://www.math.leidenuniv.nl/~evertse/dio2011-subspace.pdf`.

[5] J.-H. Evertse, The p-adic subspace theorem, 2017. Lecture notes available at `http://www.math.leidenuniv.nl/~evertse/dio17-9.pdf`.

[6] J.-H. Evertse, H. P. Schlickewei, and W. M. Schmidt, Linear equations in variables which lie in a multiplicative group, *Ann. of Math. (2)* **155** (2002), 807–836.

[7] K. Győry, On the irreducibility of neighbouring polynomials, *Acta Arith.* **67** (1994), 283–294.

[8] R. Schwartz and J. Solymosi, Combinatorial applications of the subspace theorem, in *Geometry, structure and randomness in combinatorics*, *CRM Series*, vol. 18, Ed. Norm., Pisa, 2015, pp. 123–140.