Algebraic Methods in Combinatorics
Lecture Notes (2001)

Oleg Pikhurko
University of Warwick

Contents

1 Preface

This course deals with results where a combinatorial problem is solved by applying non-combinatorial (mostly algebraic) arguments. Of course, it is always pleasant when methods and ideas developed for tacking one sort of problems can can be applied to another. Also, this often leads to a better understanding of the subject and to further results. Last, but not least, the proofs obtained via this method are often very elegant and beautiful.

The drawbacks of the method is that it does not work in many (seemingly suitably) situations and that each problem usually requires an individual approach. It seems that the whole theory is still in the stage of development.

A few words should be said how the material is presented. An old Chinese proverb says: "If you give a fish to a man, you'll feed him for a day, but if you teach him to catch fish, you'll feed him for life". One can draw interesting parallels by comparing the mathematical universe with an ocean where various notions, ideas, and proofs swim like fish and it is a mathematicians task to 'catch fish'. The moral of the proverb which we try to follow is that instead presenting just a proof to digest, we try to show the process how the proof was possibly found ("how the fish was caught"). Unfortunately, the actual reasoning, attempts, associations that preceeded the proof can hardly be recovered as they are usually not included in the papers written for publication. But at least we can try to reconstruct (or make up) how each individual proof was discovered. It is often the case that the proof preceeds the theorem: first, we play around with notions and then we see what we have actually proved.

This might create some dilemmas for those students who will take the examination in this course. The rule is: the answer should contains a clear statement of the result plus a correct proof (in any order). The proof–and–then–theorem order would usually require more writing at the exam whereas the other order would require student's own rethinking and rewriting the material before the examination.

2 Lindstrom's Theorem

The typical argument of this course is as follows. Given some combinatorial object we construct a certain algebraic object (e.g. a set of vectors or a polynomial). Then we apply some known theorem to the latter object, to get some extra information about it, and finally translate this information back, into combinatorial language.

It is often the case that the algebraic theorem we apply is an easy basic property, sometimes bordering with triviality, but its combinatorial consequence may be very deep and hard to prove by purely combinatorial means (even if we know a non-combinatorial proof).

For example, let us try to exploit the almost trivial claim that any maximal independent set in a vector space V over a field has the same cardinality $\dim(\mathsf{V})$ which is called the *dimension* of V.

Let $A \subset [n]$, where we write $[n] := \{1, \ldots, n\}$. What is the most obvious way to correspond a vector to A? Just take the *characteristic vector* $\boldsymbol{\chi}_A$ which is defined by $\chi_{A,i} = \texttt{IF[}i \in A\texttt{]}$, where $\texttt{IF[}P\texttt{]}$ is 1 if P is true, and 0 otherwise. Where $\boldsymbol{\chi}_A$ 'lives' will depend on the context. For our application let us take the real vector space, that is, $\boldsymbol{\chi}_A \in \mathbb{R}^n$.

Now, if we take $m \geq n+1$ non-empty sets $A_1, \dots, A_m$, then there is some linear dependence between their characterictic vectors $\boldsymbol{v}_1 := \boldsymbol{\chi}_{A_1}, \dots, \boldsymbol{v}_m := \boldsymbol{\chi}_{A_m}$, that is, there are reals $(\alpha_1, \dots, \alpha_m)$, not all zero, such that

$$\sum_{i=1}^{m} \alpha_i \boldsymbol{v}_i = \mathbf{0}. \tag{1}$$

Define $I_1 = \{i \in [m] : \alpha_i > 0\}$ and $I_2 = \{i \in [m] : \alpha_i < 0\}$, and rewrite (1) as

$$\sum_{i \in I_1} \alpha_i \boldsymbol{v}_i = \sum_{i \in I_2} (-\alpha_i) \boldsymbol{v}_i. \tag{2}$$

Now we have to "combinatorially interpret" (2), maybe with some coarsening (loss of information). In our case, the trick is to note that if the vectors in (2) are equal, then the sets of indexes with non-zero coordinate are also equal, that is,

$$\bigcup_{i \in I_1} A_i = \bigcup_{i \in I_2} A_i. \tag{3}$$

That's it, so easy. Let us state it as a theorem.

Theorem 1 *For any family $A_1, \dots, A_m$ of $m \geq n+1$ subsets of $[n]$ there are disjoint $I_1, I_2 \subset I$ such that $I_1 \cup I_2 \neq \emptyset$ and (3) holds.* ▮

The obtained result (which can also be proved by purely combinatorial means) is not trivial at all.

Let us go one step futher in our exploitations. Now we correspond to A_i a vector $\boldsymbol{v}_i \in \mathbb{R}^{2n}$ as follows: $\boldsymbol{v}_i = (x_1, \dots, x_n, y_1, \dots, y_n)^T$ with $x_j = \texttt{IF[}j \in A_i\texttt{]}$ and $y_j = 1 - x_j$, $j \in [n]$; in other words, $\boldsymbol{v}_i = \begin{bmatrix} \boldsymbol{\chi}_{A_i} \\ \boldsymbol{\chi}_{\overline{A}_i} \end{bmatrix}$, where $\overline{A} := [n] \setminus A$. (Note that we represent vectors as columns.)

Clearly, if we have $m \geq 2n+1$ sets, then we can guarantee a linear dependence. However, we can do far better by observing that all $\boldsymbol{v}_i$'s belong to the subspace $\mathsf{V} \subset \mathbb{R}^{2n}$ of all vectors for which $x_1 + y_1 = \dots = x_n + y_n$. The dimension of V is $n+1$. (*Prove!*) Thus, if $m \geq n+2$, we can find α's satisfying (1). Define I_1, I_2 as above and deduce (2) and (3).

But we can also use the y_i-coordinates! Looking at the sets, where the y-th coordinate of the vector in the right-hand side (and the left-hand side) of (2) is non-zero we obtain $\cup_{i \in I_1} \overline{A}_i = \cup_{i \in I_2} \overline{A}_i$, which is equivalent to

$$\bigcap_{i \in I_1} A_i = \bigcap_{i \in I_2} A_i. \tag{4}$$

Theorem 2 (Lindstrom [Lin93]) *For any family $A_1, \dots, A_m$ of $m \geq n+2$ subsets of $[n]$, there are disjoint $I_1, I_2 \subset I$ such that (3) and (4) hold and $I_1 \cup i_2 \neq \emptyset$.* ▮

Now, the latter theorem seems much harder than Theorem 1. Lindstrom [Lin93] asks if there is a combinatorial proof of Theorem 2 and this problem seems to be still open. Perhaps, Theorem 2 itself is not very important but a combinatorial proof of it should be of interest as it may introduce new interesting ideas.

Notes

I do not know to whom attribute Theorem 1: Lindstrom [Lin93] does not indicate its author, so it is probably a 'folklore' result.

Execises and Further Reading

[Lin93]: Find a purely combinatorial proof of Theorem 2. (Research problem!)

3 The Addressing Problem for Graphs

3.1 $\{0,1,*\}$-Addressing

While the previous theorems seem somewhat artificial, here is a natural problem with a gem of proof.

Suppose that we have a computer network represented by a graph G where a node $x \in V(G)$ has to send a message to a node $y \in V(G)$. Each node $z \in V(G)$ has some label $l(z)$. What x "knows" is the label of the destination y and the labels of the G-neighbours of x. Given this information, x should choose a neighbour to whom pass the message for further transmission. Ideally, the path that the message travels should be shortest possible.

For example, the vertices of the *n-cube* Q_n can be identified with $(0,1)$-sequences of length n so that two sequence are adjacent if and only if they differ in precisely one position. It is easy to see that the Q_n-distance between $\boldsymbol{x}, \boldsymbol{y} \in Q_n$ is precisely the *hamming distance*, that is, the number of indexes where $x_i \neq y_i$. Now, the routing problem has a simple and optimal algorithm: choose i such that $x_i \neq y_i$ and send the message to $\boldsymbol{x}'$ which is obtained from $\boldsymbol{x}$ by replacing x_i by y_i.

Not every graph admits a distance-preserving $(0,1)$-addressing. (*For example, prove that the triangle does not.*) To get around this obstacle we simply extend our addressing alphabet to contain a special *joker symbol* $*$, which matches both 0 and 1 (and itself). Formally, let $S := \{0,1,*\}$ and

$$d_S(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} d_S(x_i, y_i), \quad \boldsymbol{x}, \boldsymbol{y} \in S^n,$$

where $d_S(a,b) = \texttt{IF[}\{a,b\} = \{0,1\}\texttt{]}$, $a,b \in S$. Note that d_S does not satisfy the triangle inequality so it is not, strictly speaking, a distance function. However, any graph G admits a *$\{0,1,*\}^m$-addressing* which is a function $l : V(G) \to S^m$ such that $d_G(x,y) = d_S(l(x), l(y))$, if m is sufficiently large. Let $m(G)$ be the smallest such m. (*Prove that G admits an S^k-addressing for any $k \geq m(G)$.*) Given such an addressing, any message can be transmitted along the shortest path possible. (*How?*)

An important function is

$$m(n) := \max\{m(G) : v(G) = n, \kappa(G) > 0\},$$

the smallest m such that every connected graph on n vertices admits an S^m-addressing. Winkler [Win83] showed that $m(n) \leq n-1$.

One the other hand, let us show that $m(n) \geq n-1$, namely that K_n, the complete graph on n vertices, does not admit an S^m-addressing for $m < n-1$.

3.2 Reduction to Decompositions

Let $l : K_n \to S^m$ be an addressing for K_n. For any $u, v \in V(K_n)$ the S-distance between $l(u)$ and $l(v)$ is 1. This means that there exists exactly one index $i \in [m]$ with $\{l(u)_i, l(v)_i\} = \{0, 1\}$. Let H_i be the subgraph consisting of the edges corresponding to an index $i \in [m]$. Clearly, each H_i is a complete bipartite graph with parts

$$X_i := \{v \in V(K_n) : l(v)_i = 0\} \text{ and } Y_i := \{v \in V(K_n) : l(v)_i = 1\},$$

and these graphs partition the edge set of K_n.

3.3 Graham–Pollack Theorem

A general and very useful tool is the *adjacency matrix* $\mathsf{A}(G)$ of a graph G defined by $\mathsf{A}_{i,j}(G) = \texttt{IF[}\{x_i, x_j\} \in E(G)\texttt{]}$, where $v(G) = \{x_1, \ldots, x_n\}$. In particular, as we consider loopless graphs, the diagonal entries of $\mathsf{A}(G)$ are zero.

An observaton that suggests the above line of attack is that the rank of the adjacency matrix of a complete biparite graph is only 2 whereas that of K_n is large. Namely, $\mathsf{A}(K_n) = \mathsf{J}_n - \mathsf{I}_n$, where J_n is the $n \times n$-matrix composed entirely of 1's and I_n is the $n \times n$ identity matrix. It is easy to see that

$$\det(\mathsf{J}_n - \mathsf{I}_n) = \det \begin{pmatrix} & & & 1 \\ & -\mathsf{I}_{n-1} & & \vdots \\ & & & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} = (-1)^{n-1}(n-1)$$

and thus it has the full rank n (over the reals). The partition property translates into

$$\sum_{i=1}^{m} \mathsf{A}(H_k) = \mathsf{A}(K_n). \tag{5}$$

We know that

$$\text{rank}(\mathsf{A} + \mathsf{B}) \leq \text{rank}(\mathsf{A}) + \text{rank}(\mathsf{B}). \tag{6}$$

Proof: choose $a = \text{rank}(\mathsf{A})$ independent rows in A and $b = \text{rank}(\mathsf{B})$ such rows in B; any row of $\mathsf{A} + \mathsf{B}$ is a linear combination of these $a + b$ selected rows, which implies the claim.

As $\text{rank}(\mathsf{A}(H_i)) = 2$, the identity (5) implies $m \geq \text{rank}(\mathsf{A}(K_n))/2 = n/2$. Unfortunately, this bound is weaker than that we hoped to prove.

This difficulty is resolved by the suprisingly simple observation that the rank of the adjacency matrix becomes 1 if we make H_k into a directed graph DH_i by orienting its edges, say from X_i to Y_i. (The *adjacency matrix of a digraph* G is $\mathsf{A}_{i,j}(G) = \texttt{IF[}(x_i, x_j) \in E(G)\texttt{]}$.)

Now, the union of the DH_i's is a *tournament*, that is, a directed graph with the property that for every pair u, v of vertices either (u, v) or (v, u) is an edge (but not both). In the matrix language this reads

$$\mathsf{D} + \mathsf{D}^T = \mathsf{J}_n - \mathsf{I}_n, \tag{7}$$

where $\mathsf{D} := \sum_{i=1}^{m} \mathsf{A}(DH_i)$.

By (6) we have $m \geq \operatorname{rank}(\mathsf{D})$ and we aim at showing that $\operatorname{rank}(\mathsf{D}) \geq n-1$. We know that $\operatorname{rank}(\mathsf{J}_n - \mathsf{I}_n) = n$. If we apply the rank subadditivity to (7) we obtain $m \geq \operatorname{rank}(\mathsf{D}) \geq n/2$ again. Here one might be inclined to give up declaring that $n/2$ is perhaps the best bound obtainable with linear algebra.

But let us persist — after all D and D^T are not two arbitrary matrices. The key observation is that if $\mathsf{D}\boldsymbol{x} = \mathbf{0}$, then $\boldsymbol{x}^T\mathsf{D}^T = 0$ and, by (7),

$$0 = \boldsymbol{x}^T(\mathsf{D} + \mathsf{D}^T)\boldsymbol{x} = \boldsymbol{x}^T\mathsf{J}_n\boldsymbol{x} - \boldsymbol{x}^T\mathsf{I}_n\boldsymbol{x} = \boldsymbol{x}^T\mathsf{J}_n\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{x}. \tag{8}$$

From this we conclude that $\mathsf{J}_n\boldsymbol{x} = \mathbf{0}$ implies $\boldsymbol{x} = \mathbf{0}$ (otherwise $\boldsymbol{x}^T\boldsymbol{x} > 0$ contradicting (8)). The system $\mathsf{J}_n\boldsymbol{x} = \mathbf{0}$ is nothing else as a single equation $\mathbf{1} \cdot \boldsymbol{x} = 0$, where $\mathbf{1} \in \mathbb{R}^n$ is the all-1 vector. In other words, there is no non-zero solution to $\begin{bmatrix} \mathsf{D} \\ \mathbf{1}^T \end{bmatrix} \boldsymbol{x} = \mathbf{0}$, which implies that $\operatorname{rank} \begin{bmatrix} \mathsf{D} \\ \mathbf{1}^T \end{bmatrix} = n$ and $\operatorname{rank}(\mathsf{D}) \geq n-1$, as required.

Putting all together, we have proved the following results.

Lemma 3 *The identity* (7) *implies that* $\operatorname{rank}(\mathsf{D}) \geq n-1$. *In particular, the adjacency matrix of an order-n tournament T has rank at least $n-1$.* ▮

Theorem 4 (Graham–Pollak [GP71]) *There is no partition of the edge-set of K_n into fewer than $n-1$ complete bipartite graphs, $n > 1$.* ▮

Theorem 5 (Graham–Pollak [GP71]) $m(K_n) \geq n-1$ *and, consecutively,* $m(n) \geq n-1$. ▮

It is usually Theorem 4 that is referred to as the *Graham–Pollak Theorem.*

Notes

The original proof of Theorem 4 by Graham and Pollack [GP71] was rather complicated. Tverberg [Tve82] was first to find a simple proof (which is similar to the one presented here). Pritikin [Pri86] extended Tberberg's method to digraphs and multigraphs, and Alon [Alo86a] to complete r-partite graphs.

Execises and Further Reading

[BF92]: Try exercises to Chapter 1.4.

[LW92]: Read Chapter 9 for a nice exposition of Winkler's [Win83] proof that $m(n) \leq n-1$.

4 Rules and Clubs

4.1 A Few Problems

To control the number of university clubs, the council suggested the following *Even Rules*:

1. Every club must consist of an even number of members.
2. Every two clubs must share an even number of members.
3. No two clubs can have exactly the same set of members.

The Council hoped to restrict the maximum possible number of clubs, given that there are in total n students who are eligible for a club membership. However, within a short time $2^{\lfloor n/2 \rfloor}$ clubs have been formed (**How? Can more clubs be formed?**), each club pestering the university for money for squashes and other little social events (such as, for example, the annual trip to Hawaii). The next (emergency) council meeting hired a mathematician and introduced a little amendment into the rules, the new ones being called *Odd Rules*:

1. Every club must consist of an odd number of members.
2. Every two clubs must share an even number of members.

In no way could the students form more than n clubs. (**Prove!**) The outrage was greater even than that ever produced by loosing the rowing race to Oxford! The situation was stabilised only after the next (emergency) council meeting promised to change the club rules every year. The clever mathematician came up with following *Rule* λ for $1 \leq \lambda \leq n$:

1. Any two clubs have precisely λ members in common.
2. No two clubs can have exactly the same set of members.

Prove that the mathematician was clever indeed (provided he survived the next students' outbreak) by proving that **no more than n clubs can be formed at any time**.

Notes

The Odd/Even Rules problems come from Babai and Frankl [BF92, Chapter 1.1]; the stated bound for "Rule i" is due to Fisher [Fis40].

I will explain solutions next time. The impatient ones can find them in [BF92, Chapter 1.1] and in [LW92, Chapter 19]. But please do try to think about the problems yourself before looking up solutions. (A 'helpful' hint: use linear algebra for proving upper bounds!)

4.2 Solutions

Let us give the solutions to the above problems which we restate in the combinatorial language. A *set system* (or a *hypergraph*) on a set X is $\mathcal{H} \subset 2^X$, a collection of subsets of X which are called *edges*. The *size* $|\mathcal{H}|$ is the number of edges.

Theorem 6 (Even Rules) *Let $\mathcal{H} \subset 2^{[n]}$ be a collection of distinct subsets of $[n]$ such that $|A \cap B|$ is even for any $A, B \in \mathcal{H}$. (In particular, each $|A|$ is even.) Then the maximum size of $\mathcal{H}$ is $2^{\lfloor n/2 \rfloor}$.*

Proof. It is easy to find a suitable hypergraph with $2^{\lfloor n/2 \rfloor}$ edges: form $\lfloor n/2 \rfloor$ pairs and let $\mathcal{H}$ consist of all possible unions of pairs. (All (but at most one) students suddenly marry each other with each newly-wed couple staying together all the time.)

On the other hand, note that $|A \cap B| = \boldsymbol{\chi}_A \cdot \boldsymbol{\chi}_B$. As we are interested not in actual intersection sizes but in their residues modulo 2, let the characteristic vectors live in $(\mathbb{F}_2)^n$, the n-dimensional vector space over $\mathbb{F}_2$, the field on 2 elements. Our assumption reads now that $\boldsymbol{\chi}_A \cdot \boldsymbol{\chi}_B = 0$ for any $A, B \in \mathcal{H}$. In other words, the spanned subspace

$$\mathsf{U} = \mathrm{Span}\{\boldsymbol{\chi}_A : A \in \mathcal{H}\} \subset (\mathbb{F}_2)^n.$$

is a subspace of its own *orthocomplement*

$$\mathsf{U}^\perp := \{\boldsymbol{x} \in (\mathbb{F}_2)^n : \boldsymbol{u} \in \mathsf{U} \Rightarrow \boldsymbol{u} \cdot \boldsymbol{x} = 0\}.$$

For arbitrary $\mathsf{U} \subset (\mathbb{F}_2)^n$ we have

$$\dim(\mathsf{U}^\perp) = n - \dim(\mathsf{U}), \tag{9}$$

which follows from the fact that $\mathsf{U}^\perp$ can be defined by $d = \dim(\mathsf{U})$ linearly independent equations $\boldsymbol{x} \cdot \boldsymbol{u}_i = 0$, where $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ is a basis of U.

As $\mathsf{U} \subset \mathsf{U}^\perp$, we have $\dim(\mathsf{U}) \le \dim(\mathsf{U}^\perp)$. Now $m \le 2^{\dim(\mathsf{U})} \le 2^{\lfloor n/2 \rfloor}$, as required. ▮

Remark. Beware that $\mathsf{U} \cup \mathsf{U}^\perp$ need not span $(\mathbb{F}_q)^n$ if q is a power of 2 (although the analogeous claim is true for fields of odd or zero characteristic).

Theorem 7 (Odd Rules) *Let $\mathcal{H} \subset 2^{[n]}$ be such that $|A|$ is odd for each $A \in \mathcal{H}$ but $|A \cap B|$ is even for any distinct $A, B \in \mathcal{H}$. Then the maximum size of $\mathcal{H}$ is n.*

Proof. Lower bound: take $\mathcal{H} = \{\{i\} : i \in [n]\}$. (Each student forms his/her own club.)

For the upper bound we consider again the characteristic vectors $\boldsymbol{\chi}_A \in (\mathbb{F}_2)^n$, $A \in \mathcal{H}$. The assumptions say that every two are orthogonal while each has non-zero norm (that is, $\boldsymbol{\chi}_A \cdot \boldsymbol{\chi}_A \ne 0$). Our intuition tells us that these vectors should be independent. Indeed, they are: if

$$\sum_{A \in \mathcal{H}} \lambda_A \boldsymbol{\chi}_A = \boldsymbol{0},$$

then the scalar product of this identity with $\boldsymbol{\chi}_A$ shows that each $\lambda_A = 0$. Hence, $m \le \dim((\mathbb{F}_2)^n) = n$. ▮

Execises and Further Reading

[BF92]: Chapter 1.1 contains many excellent exercises.

5 Modular Frankl–Wilson Inequality

One way to show that some vectors are independent is as follows.

Lemma 8 (Diagonal Principle) *Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in \mathbb{R}^n$. Suppose that we can find linear functions $u_1, \ldots, u_m \in (\mathbb{R}^n)^*$ such that $u_j(\boldsymbol{v}_i) = 0$ for all $1 \le i < j \le m$ and $u_i(\boldsymbol{v}_i) \neq 0$ for any $i \in [m]$. Then $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ are linearly independent.*

Proof. Suppose on the contrary that we have $\sum_{i=1}^m c_i \boldsymbol{v}_i = \boldsymbol{0}$ for some scalars c_i, not all zero. Let j be the largest index such that $c_j \neq 0$. Then $0 = u_j(\sum_{i=1}^m c_i \boldsymbol{v}_i) = c_j u_j(\boldsymbol{v}_j) \neq 0$, a contradiction. ▮

The following result is a nice illustration of the above principle. This theorem appears in Grolmusz and Sudakov [GS01] but it was built upon the results of other people. Please read the notes for the historical remarks which explain the name of the result, *Modular k-Wise Frankl–Wilson Inequality.*

Theorem 9 (Modular k-Wise Frankl–Wilson Inequality) *Let p be a prime and $L = \{l_1, \ldots, l_s\} \subset [0, p-1]$. Let $k \ge 2$ be an integer and let $\mathcal{H}$ be a family of subsets of $[n]$ such that $|C| \notin L \pmod p$ for every $C \in \mathcal{H}$ but $|C_1 \cap \ldots \cap C_k| \in L \pmod p$ for any collection of k distinct sets from $\mathcal{H}$. Then*

$$|\mathcal{H}| \le (k-1) \sum_{i=0}^{s} \binom{n}{i}. \tag{10}$$

Proof. We repeat the following procedure until $\mathcal{H}$ is empty. At Round i, if $\mathcal{H} \neq \emptyset$ we choose a collection $C_1, \ldots, C_d$ from $\mathcal{H}$ such that $|\cap_{j=1}^d C_j| \notin L \pmod p$ but for any additional set $C' \in \mathcal{H}$ we have $|(\cap_{j=1}^d C_j) \cap C'| \in L \pmod p$. Clearly, such family always exists (build it by adding one new edge at a time) and $d \in [k-1]$. Denote $A_i = C_1$, $B_i = \cap_{j=1}^d C_j$ and remove all sets $C_1, \ldots, C_d$ from $\mathcal{H}$. Suppose that as the result of this process we obtain m pairs of sets A_i, B_i, $i \in [m]$. Note that $m \ge |\mathcal{H}|/(k-1)$. By definition, $|A_i \cap B_i| = |B_i| \notin L \pmod p$ but $|A_j \cap B_i| \in L \pmod p$ for any $j > i$.

For $i = 1, \ldots, m$ let us define the multilinear polynomial f_i in n variables $\boldsymbol{x} = (x_1, \ldots, x_n) \in (\mathbb{F}_p)^n$ by

$$f_i(\boldsymbol{x}) = \prod_{j=1}^{s} (\boldsymbol{x} \cdot \boldsymbol{\chi}_{B_i} - l_j) \in \mathbb{F}_p, \tag{11}$$

The point of this definition is that

$$f_i(\boldsymbol{\chi}_{A_i}) = \prod_{j=1}^{s} (|A_i \cap B_i| - l_j) = \prod_{j=1}^{s} (|B_i| - l_j) \neq 0, \quad i \in [m], \tag{12}$$

but

$$f_i(\boldsymbol{\chi}_{A_h}) = \prod_{j=1}^{s} (|A_h \cap B_i| - l_j) = 0, \quad 1 \le i < h \le m. \tag{13}$$

Now, the Diagonal Principle (Lemma 8) implies that the polynomials $f_1, \ldots, f_m$ are linearly independent: correspond $\boldsymbol{\chi}_{A_i}$ to the linear functional on the space of polynomials which maps f to $f(\boldsymbol{\chi}_{A_i})$. On the other hand, each f_i belongs to the linear subspace F of all polynomials of degree at most s; so $m \le \dim(\mathsf{F})$. Clearly, the dimension of F is at most (in fact, is equal to) the number of *monomials* $x_1^{d_1} \ldots x_n^{d_n}$ with $d_1 + \ldots + d_n \le s$. It is easy to see that the number of non-negative integer solutions $\boldsymbol{d}$ to $d_1 + \ldots + d_n = i$ is $\binom{n+i-1}{n-1} = \binom{n+i-1}{i}$. (*Prove this.*) Hence, we obtain

$$|\mathcal{H}| \le (k-1)m \le (k-1) \sum_{i=0}^{s} \binom{n+i-1}{i}. \tag{14}$$

This is a good upper bound, just 'slightly' weaker than the one we have to prove.

The following cute trick improves on (14). Observe that in (12) and (13) we evaluate the f_i's on $(0,1)$-vectors only. But clearly $0^d = 0$ and $1^d = 1$ for $d \ne 0$. Thus, if $\bar{f}_i$ is obtained from f_i by reducing for each monomial the exponent of each occurring variable to 1, then we still have

$$\begin{array}{rcll} \bar{f}_i(\boldsymbol{\chi}_{A_i}) & \ne & 0, & i \in [m], \\ \bar{f}_i(\boldsymbol{\chi}_{A_h}) & = & 0, & 1 \le i < h \le m. \end{array}$$

Again by the Diagonal Principle (Lemma 8), the polynomials $\bar{f}_1, \ldots, \bar{f}_m$ are linearly independent. However, each $\bar{f}_j$ is a *multilinear polynomial* (that is, linear in each separate variable); these are generated by monomials $\prod_{i \in I} x_i$, $I \in \binom{[n]}{\le s}$. Hence, $m \le \sum_{i=0}^{s} \binom{n}{i}$, implying the desired bound on $|\mathcal{H}|$. ∎

Now let us discuss how far the obtained upper bound is from being sharp. The following construction shows that if $s < p$ are fixed while $k = 2^{o(n)}$, then the bound of Theorem 9 is asymptocally best possible (if it be a function of n, k, s and p only).

Let $L = [0, s-1]$. Choose t with $2^{t-1} < k-1 \le 2^t$. We have $t = o(n)$. Let $Y_1, \ldots, Y_t$ be disjoint p-subsets of $[n]$ and let $Y := [n] \setminus \cup_{i=1}^{t} Y_i$. By definition $|Y| = n - o(n)$. Choose any distinct $I_1, \ldots, I_{k-1} \in 2^{[t]}$. (This is possible as $k - 1 \le 2^t$.) Finally, the hypergraph $\mathcal{H}$ consists of all subsets of $[n]$ of the form $A \cup \left(\cup_{i \in I_j} Y_i\right)$ for all $A \in \binom{Y}{s}$ and $j \in [k-1]$. Clearly,

$$|\mathcal{H}| = (k-1)\binom{|Y|}{s} = (1 - o(1))(k-1)\binom{n}{s},$$

and every $C \in \mathcal{H}$ has size equal to s modulo p. But for any distinct $C_1, \ldots, C_k \in \mathcal{H}$ not all intersections $C_i \cap Y$ can be the same; hence $C := C_1 \cap \cdots \cap C_k$ intersects Y in at most $s-1$ vertices and $|C| \in [0, s-1]$ (mod p) — the hypergraph $\mathcal{H}$ satisfies all assumptions of Theorem 9.

However, for some concrete L better bounds can be obtained. Unfortunately, no theory which gives (asymptotically) sharp upper bounds on $|\mathcal{H}|$ for any L has been developed yet.

Some 'non-modular' results can be deduced from the Modular Frankl–Wilson Inequality (Theorem 9). Here is one example.

Corollary 10 (Weak Fisher's Inequality) *Let $0 \le \lambda \le n$ and let $\mathcal{H} \subset 2^{[n]}$ satisfy $|A \cap B| = \lambda$ for any distinct $A, B \in \mathcal{H}$. Then $|\mathcal{H}| \le n+1$.*

Proof. If $|A| = \lambda$ for some $A \in \mathcal{H}$, then any two other sets intersect precisely in A and it follows that $|\mathcal{H}| \le n - \lambda + 1 \le n+1$, as required. So, let us assume that $|A| > \lambda$ for each $A \in \mathcal{H}$.

Choose any prime $p > n - \lambda$. For any $A \in \mathcal{H}$, we have $\lambda < |A| \le n < \lambda + p$; hence, $|A| \not\equiv \lambda \pmod{p}$. Thus all assumptions of Theorem 9 are satisfied (for $s = 1$, $L = \{\lambda\}$, and $k = 2$), which implies that $|\mathcal{H}| \le n+1$, as required. ▮

Remark. The bound in Corollary 10 is sharp for $\lambda = 0$ (take $\mathcal{H} = \{\{i\} : i \in [n]\} \cup \{\emptyset\}$). However, if $\lambda \ge 1$, then it is possible to show that $|\mathcal{H}| \le n$, the latter being known as the *Fisher's Inequality*, a version of which was first proved by Fisher [Fis40].

Notes

Let L be a set consisting of s non-negative integers and let $\mathcal{H}$ be a hypergraph on $[n]$. We say that $\mathcal{H}$ is *L-intersecting* if $|E \cap D| \in L$ for any pair of distinct edges $E, D \in \mathcal{H}$.

Ray-Chaudhuri and Wilson [RCW75] showed that if $\mathcal{H}$ is *k-uniform* (i.e. $|A| = k$ for each $A \in \mathcal{H}$) and L-intersecting, then $|\mathcal{H}| \le \binom{n}{s}$. Frankl and Wilson [FW81] proved that if H is L-intersecting (but not necessarily uniform), then $|\mathcal{H}| \le \sum_{i=0}^{s} \binom{n}{i}$, which can be viewed as the *non-uniform version of Ray-Chaudhuri–Wilson inequality* (and is sometime referred to by this name).

Deza, Frankl and Singhi [DFS83] proved the modular version of the Frankl–Wilson Inequality (our Theorem 9 for $k = 2$). In the modular version we allow not only L-intersections but also their translations by a multiple of a prime p while on the other side we restrict the possible sizes of edges. Although, the assumptions and conclusions sound very similar, one result does not seem to imply the other.

Also, in Theorem 9 we go for a more general result by restricting k-wise intersections (not just pairwise). This explains the name of Theorem 9.

What we have touched upon is but a tiny fraction of the beautiful (albeit far from being complete) theory of intersecting hypergraphs. Some further hallmarks of the theory are indicated in the abstract of the Part III Essay "Intersecting Set Systems" which places emphasis on 'non-modular' intersection theorems. To demonstrate the usefulness of the theory we will present a few applications of our 'modular' Theorem 9, hopefully staying clear off the main line of the essay.

Those writing the essay should seek the advice of Dr. Thomason regarding its scope.

Execises and Further Reading

[BF92]: See Chapter 5.4 for further discussion and exercises.

6 Chromatic Number of $\mathbb{R}^n$

The *chromatic number* of $\mathbb{R}^n$, denoted by $\chi(\mathbb{R}^n)$, is the smallest number of colours needed to colour the points of $\mathbb{R}^n$ so that no two points at distance 1 have the same colour. Alternatively, $\chi(\mathbb{R}^n)$ is the smallest number m of parts in a partition $\mathbb{R}^n = X_1 \cup \cdots \cup X_m$ such that no pair of points within the same X_i is at the unit distance.

6.1 Small Dimensions

To get a feeling of this function, let us consider some easy examples.

Clearly, two colours are enough to colour $\mathbb{R}^1$: we just have to ensure that $x, x+1 \in \mathbb{R}$ always receive different colours. This can be achieved, for example, by the 2-colouring $c : \mathbb{R} \to \{0,1\}$ with $c(x) = \lfloor x \rfloor \pmod 2$.

Theorem 11 $\chi(\mathbb{R}^1) = 2$. ▮

Things get complicated already for $\mathbb{R}^2$. The unit-length equilateral triangle T shows that $\chi(\mathbb{R}^2) \geq 3$. Does there exist a 3-colouring? If yes, then any copy of T receives all 3 colours; thus any two vertices at distance $\sqrt{3}$ have the same colour (see Figure 1). But then the isosceles triangle with sides $(\sqrt{3}, \sqrt{3}, 1)$ (or, in other words, the configuration of Figure 2) establishes a contradiction.

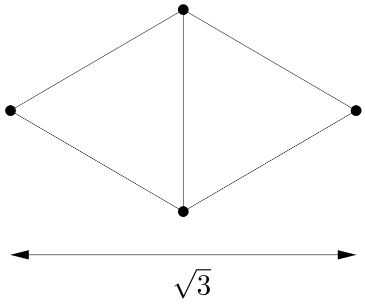

Fig. 1: Two length-1-equilateral triangles.

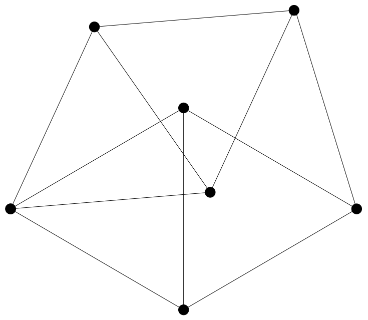

Fig. 2: All edges have length 1.

Theorem 12 $\chi(\mathbb{R}^2) \geq 4$. ▮

6.2 General Upper Bounds

Here we are interested in the 'large-scale' behaviour of $\chi(\mathbb{R}^n)$. It is not hard to see that $\chi(\mathbb{R}^n)$ is finite. The idea is to tile $\mathbb{R}^n$ into parts of small diameter and assign a colour to each part in a periodical fashion so that two parts of the same colour are always far apart. It is straightforward to realise this idea. For example, partition $\mathbb{R}^n = \cup_{\boldsymbol{i}\in\mathbb{Z}^n} Q_{\boldsymbol{i}}$ into 'cubes'

$$Q_{\boldsymbol{i}} := \{\boldsymbol{x} \in \mathbb{R}^n : \lfloor x_j/\sqrt{n} \rfloor = i_j \text{ for all } j \in [n]\}, \quad \boldsymbol{i} \in \mathbb{Z}^n,$$

and colour the whole of $Q_{\boldsymbol{i}}$ by the colour $(i_1 \pmod k), \ldots, i_n \pmod k))$, where $k = \lceil \sqrt{n}\,\rceil + 1$. This colouring shows that

$$\chi(\mathbb{R}^n) \le (\lceil \sqrt{n}\,\rceil + 1)^n = \mathrm{e}^{(\frac{1}{2}+o(1))\,n \ln n}. \tag{15}$$

However, it seems that our colouring of the cubes is uneconomical. Let us apply the *greedy algorithm*, wherein one colours the cubes straightforwardly, one by one. When we try to choose a colour for $Q_{\boldsymbol{i}}$, then restrictions on the colour can come only from those cubes $Q_{\boldsymbol{j}}$ for which there are $\boldsymbol{x} \in Q_{\boldsymbol{i}}$ and $\boldsymbol{y} \in Q_{\boldsymbol{j}}$ with $\|\boldsymbol{x} - \boldsymbol{y}\| = 1$. But every such $Q_{\boldsymbol{j}}$ lies entirely within a ball B of radius $5/2$ whose centre coincides with the centre $\boldsymbol{q}$ of $Q_{\boldsymbol{i}}$ because for any $\boldsymbol{z} \in Q_{\boldsymbol{j}}$

$$\|\boldsymbol{z} - \boldsymbol{q}\| \le \|\boldsymbol{z} - \boldsymbol{y}\| + \|\boldsymbol{y} - \boldsymbol{x}\| + \|\boldsymbol{x} - \boldsymbol{q}\| < 1 + 1 + 1/2 = 5/2.$$

The number of such $Q_{\boldsymbol{j}}$'s is at most

$$\frac{\mathrm{vol}(B)}{\mathrm{vol}(Q)} = \frac{\pi^{n/2}(5/2)^n}{\Gamma(n/2+1)} \times n^{n/2} = (10.3318...)^n$$

and a colour for $Q_{\boldsymbol{i}}$ can always be chosen if the number of available colours is at least $\lfloor \mathrm{vol}(B)/\mathrm{vol}(Q) \rfloor + 1$. Therefore we have for all large n

$$\chi(\mathbb{R}^n) \le 10.4^n$$

which improves on (15).

Still, the bound can be imporved, which comes from the fact that the volume ratio of a ball and a cube of the same diameter is large. (A cube has corners that protrude far away.) If we would have had a partition of $\mathbb{R}^n$ into diameter-1 balls, then we would have probably obtained a better upper bound. There is, of course, no hope of a ball tiling (except for $\mathbb{R}^1$), but what we actually need for our argument is a *covering*: we can assign a colour to each part and then for every $\boldsymbol{x} \in \mathbb{R}^n$ choose one part P containing $\boldsymbol{x}$ and colour $\boldsymbol{x}$ by the colour assigned to P.

A good explicit covering by balls is not easy to come up with, but the following (non-constructive) definition produces magic results. Let $C \subset \mathbb{R}^n$ be a maximal set with respect to the property that its any two points are at distance at least $1/2$. Clearly, the union of open balls of radius $1/2$ (that is, of diameter 1) centred at points of C covers the whole space: $\mathbb{R}^n = \cup_{\boldsymbol{x} \in C} B_{\boldsymbol{x}}(1/2)$. On the other hand, the open balls of radius $1/4$ about the points of C are disjoint, so C cannot get too dense.

Take some ball $B = B_{\boldsymbol{x}}(1/2)$ centred at $\boldsymbol{x} \in C$. At most $|Y|$ other balls can interfere when we try to assign a colour to B, where

$$Y = \{\boldsymbol{y} \in C : \|\boldsymbol{x}' - \boldsymbol{y}'\| = 1 \text{ for some } \boldsymbol{x}' \in B_{\boldsymbol{x}}(1/2) \text{ and } \boldsymbol{y}' \in B_{\boldsymbol{y}}(1/2)\}.$$

Each $\boldsymbol{y} \in Y$ lies within distance less than 2 from $\boldsymbol{x}$ so the open balls of radius $1/4$ about points of Y lie entirely within $B_{\boldsymbol{x}}(9/4)$ (and are disjoint). Hence,

$$|Y| < \frac{\mathrm{vol}(B(9/4))}{\mathrm{vol}(B(1/4))} = 9^n.$$

Thus $|Y| + 1 \le 9^n$ colours always suffice.

Theorem 13 $\chi(\mathbb{R}^n) \le 9^n$ *for any* n. ▮

6.3 Lower Bounds

Let us turn to lower bounds. The obvious idea is to find a finite $Y \subset \mathbb{R}^n$ such that the graph $G(Y)$ has large chromatic number, where $G(Y)$ has Y for the vertex set with $\boldsymbol{x}, \boldsymbol{y} \in Y$ being connected if and only if $\|\boldsymbol{x} - \boldsymbol{y}\| = 1$.

A hint that we take for granted is to let

$$Y = \left\{\boldsymbol{\chi}_A : A \in \tbinom{n}{k}\right\}.$$

for some suitable k. Of course, there is nothing special about the distance 1 in the definition of $\chi(\mathbb{R}^n)$, so instead of scaling Y we will prefer to specify later some distance d, which need not equal 1, as 'forbidden'. In other words, we will be proving a lower bound on $\chi(G(Y,d))$, where in $G(Y,d)$ we connect points of Y at distance d.

The canonical way of bounding the chromatic number of a graph G is to use the trivial inequality

$$\chi(G) \geq \frac{v(G)}{\alpha(G)}, \tag{16}$$

where $v(G)$ is the *order* (the number of vertices) and $\alpha(G)$ is the *independence number*, the largest size of an *independent set* (a set that spans no edge).

Of course, we know $v(G) = \binom{n}{k}$. Let $X \subset Y$ be an independent set in $G := G(Y,d)$. By the definition of Y the set X corresponds to the hypergraph

$$\mathcal{H} = \left\{A \in \tbinom{[n]}{k} : \boldsymbol{\chi}_A \in X\right\}.$$

What does the G-independence of X mean in terms of $\mathcal{H}$? The distance between $\boldsymbol{\chi}_A, \boldsymbol{\chi}_B \in X$ is

$$\|\boldsymbol{\chi}_A - \boldsymbol{\chi}_B\| = \sqrt{|A \bigtriangleup B|} = \sqrt{2(k - |A \cap B|)},$$

that is, it depends on the size of $|A \cap B|$ only (for fixed k). Thus, if we define $d = \sqrt{2(k-l)}$, then the 'forbidden' distance d will correspond to intersection size l. Now, X is independent in G if and only if no two distinct edges of $\mathcal{H}$ intersect in precisely l elements. Our aim is to bound $|X| = |\mathcal{H}|$ from above. And we have a tool for this: the Modular 2-Wise Frankl–Wilson Inequality (Theorem 9)!

To apply Theorem 9 we have to choose some p and $L \subset [0, p-1]$ so that

1. $k \notin L \pmod p$;
2. $([0, k-1] \setminus \{l\}) \subset L \pmod p$.

If Conditions 1 and 2 are satisfied, then Theorem 9 implies that

$$|\mathcal{H}| \leq \sum_{i=0}^{s} \binom{n}{i}, \tag{17}$$

where $s = |L|$. From (16) we obtain

$$\chi(\mathbb{R}^n) \geq \chi(G) \geq \frac{v(G)}{\alpha(G)} \geq \frac{\binom{n}{k}}{\sum_{i=0}^{s} \binom{n}{i}}. \tag{18}$$

6.4 Figuring the Optimal Parameters

It remains to choose k, l, p, L so that the lower bound (18) is as large as possible. Observe that (17) increases very rapidly with s (when $s < (\frac{1}{2} - \varepsilon)n$) so we should probably keep $s = |L|$ as small as possible. By Condition 1 $L \neq [0, p-1]$; by Condition 2 L must include (modulo p) two intervals containing together $k-1$ integers. Thus $|L| \geq (k-1)/2$. To achieve equality we have to ensure that these two intervals have the same size $(k-1)/2$ and superimpose each with the other when taken modulo p. This gives us no choice but to define

$$k = 2p-1,\ l = p-1 \text{ and } L = [0, p-2]. \tag{19}$$

Assuming that $\alpha := p/n$ is smaller than $\frac{1}{2} - \varepsilon$, we infer that $\sum_{i=0}^{p-1} \binom{n}{i} = \Theta(\binom{n}{p-1})$ because the ratio of two consecutive summands

$$\binom{n}{i} \Big/ \binom{n}{i-1} = \frac{n-i+1}{i} \leq \frac{n-p+2}{p-1}$$

is strictly below 1. Applying the rough version $n! = \Theta(n^{1/2}(n/\mathrm{e})^n)$ of *Stirling's formula*, the lower bound (18) reads

$$\frac{\binom{n}{2p-1}}{\Theta(\binom{n}{p-1})} = \left(\frac{\alpha^\alpha (1-\alpha)^{1-\alpha}}{(2\alpha)^{2\alpha}(1-2\alpha)^{1-2\alpha}} + o(1) \right)^n. \tag{20}$$

The expression in α is maximised for $\alpha_0 = \frac{2-\sqrt{2}}{4} = 0.1464...$, which gives $\chi(\mathbb{R}^n) \geq (1.207... + o(1))^n$. (The industrious reader can take the derivative of (20) with respect to α, where the factor $\ln(\frac{(1-2\alpha)^2}{4\alpha(1-\alpha)})$ having α_0 for a root pops up.)

It is well-known that for any $\varepsilon > 0$ there is a prime between m and $(1+\varepsilon)m$ for all large m, see e.g. Corollary 18. In particular, we can find a prime $p = (\alpha_0 + o(1))n$.

We were a bit slopy in the above calculations, but this is not crucial: we have come up with certain values ($p = (\alpha_0 + o(1))n$, $k = 2p-1$, etc.) and they (be they optimal or not) produce the following bound on $\chi(\mathbb{R}^n)$.

Theorem 14 (Frankl & Wilson [FW81]) $\chi(\mathbb{R}^n) > 1.2^n$ *if n is sufficiently large.* ▮

(*Write a correct, beautiful proof of Theorem 14!*)

6.5 Some Improvements

Can we improve our lower bound on $\chi(\mathbb{R}^n)$ by considering more general sets of vectors, not just from $\{0,1\}^n$ but, for example, from $\{-1,0,1\}^n$?

So, fix some positive integers a and b with $a + b < n$. To a pair

$$(A, B) \in \binom{[n]}{a, b} := \left\{ (A', B') : A' \in \binom{[n]}{a},\ B' \in \binom{[n] \setminus A'}{b} \right\}.$$

we correspond a vector $\boldsymbol{\chi}_{A,B} \in \mathbb{R}^n$ which is $+1$ on A, -1 on B and zero otherwise. Let $Y = \{\boldsymbol{\chi}_{A,B} : (A,B) \in \binom{[n]}{a,b}\}$. Clearly, every element of Y has norm $\sqrt{a+b}$, so the distance d between any two points of Y is determined by their scalar product l: $d^2 = 2(a+b) - 2l$.

As before we want to show that if no pair of elements of $X \subset Y$ has scalar product equal to the forbidden value l, then $|X|$ is 'small'. However, the formula

$$\boldsymbol{\chi}_{A,B} \cdot \boldsymbol{\chi}_{A',B'} = |A \cap A'| + |B \cap B'| - |A \cap B'| - |A' \cap B|,$$

hardly suggest any way to attack the claim.

But... let us digress and try to expand the proof of (17) from Section 6.3. We obtain the following outline: for any $\boldsymbol{\chi}_A \in Y$ define the polynomial $f_A(\boldsymbol{x}) = \prod_{i \in L}(\boldsymbol{\chi}_A \cdot \boldsymbol{x} - i)$ and show that f_A, $A \in X$, are linearly independent. (We had one more step of reducing each f_A to the multilinear polynomial $\bar{f}_A$, which improves the bound.)

Let us generalise this scheme of proof to the present settings. In order to have the linear independence, our polynomials $f_{A,B}$ should satisfy

$$f_{A,B}(\boldsymbol{\chi}_{A,B}) \neq 0, \quad \forall\ (A,B) \in \binom{[n]}{a,b}, \tag{21}$$

$$f_{A,B}(\boldsymbol{\chi}_{A',B'}) = 0, \quad \forall\ (A,B) \neq (A',B') \text{ and } \boldsymbol{\chi}_{A,B} \cdot \boldsymbol{\chi}_{A',B'} \neq l \tag{22}$$

The second condition prompts us to take a finite field $\mathbb{F}_p$, $L \subset \mathbb{F}_p$, and define

$$f_{A,B}(\boldsymbol{x}) = \prod_{i \in L}(\boldsymbol{\chi}_{A,B} \cdot \boldsymbol{x} - i) \in \mathbb{F}_p, \quad x \in \mathbb{R}^n. \tag{23}$$

To satisfy (21) we must have $a + b \notin L$; to satify (22) we have to require that any distinct $\boldsymbol{\chi}_{A,B}, \boldsymbol{\chi}_{A',B'} \in Y$ have scalar product either l (exactly) or in L (modulo p).

Suppose that (21) and (22) are satisfied. Then we conclude that the polynomials $f_{A,B}$, $(A,B) \in \binom{[n]}{a,b}$, are linearly independent and so $|X|$ is at most the corresponding dimension.

But before we compute the dimension observe that the vectors we feed to our polynomial in (21) and (22) have coordinates in $\{-1, 0, +1\}$. As $(x+1)x(x-1) = x^3 - x$ is then zero, we can replace each occurence of x_i^3 by x_i without changing the value of the polynomial. Let $\bar{f}_{A,B}$ be the obtained polynomials; they too satisfy (21) and (22).

Each $\bar{f}_{A,B}$ has degree at most $s := |L|$ while its degree in each variable is at most 2. To generate a monomial with these properties we have to choose numbers f, g with $f + 2g \leq s$, and then specify f variables which have exponent 1 and g variables with exponent 2. This shows that there are

$$d = \sum_{f=0}^{s} \sum_{g=0}^{\lfloor (s-f)/2 \rfloor} \binom{n}{f,g} \tag{24}$$

such monomials and the dimension argument implies that $|X| \leq d$.

We expect $s = \Theta(n)$, in which case d grows exponentially in n and its rate of growth is determined by the largest term in the sum (24). Up to a negligible error, to maximise $\binom{n}{f,g}$ it is enough to consider the following plausible pairs (f,g): either $f \approx g \approx n/3$ or on the 'border' which is defined by $f = 0$, or $g = 0$, or $f + 2g = s$. For example, if $f+2g=s$, then the pertubation $f'=f\pm 2$ and $g'=g\mp 1$ shows that we must have $f^2\approx g(n-f-g)$.

The bound that we would obtain is $\chi(\mathbb{R}^n)\geq \binom{n}{a,b}/d$. As before the best choice seems to make $|L|$ as small as possible, by having $a+b=2p-1$, $l=p-1$ and $L=[0,p-2]$, cf. (19). Rather than doing all optimisation symbolically, the reader is encouraged to make experiments (with *Mathematica* for example). For example, let $p=\gamma n$ with γ ranging from 0 to 1 with step 0.01 say; then run $a=\alpha n$ for $0\leq\alpha\leq\gamma$, computing numerically ψ in the obtained bound $\chi(\mathbb{R}^n)\geq(\psi+o(1))^n$. Making the range and increment of γ smaller, we can compute the best bound given by this method with arbitrary precision.

Theorem 15 (Raigorodski [Rai01]) $\chi(\mathbb{R}^n)\geq(1.239...+o(1))^n$.

Proof. Let $\gamma=0.4884...$ and $\alpha=0.03606...$ above. ▮

Do we get any further improvements by considering vectors in $\{\mp 2,\mp 2,0\}^n$? The author does not know, although it seems rather not (unless there are extra ideas). Allowing entries like ± 2 may make scalar products four times bigger, so the corresponding p, $|L|=p-1$ and d get rather large which seems to compensate our gain of 5^n rather than 3^n possible vectors. However, it is quite possible that an improvement can be achieved by some extra ideas. For example, Raigorodski [Rai01] believes that the upper bound (24) on $|X|$ is far from being sharp.

Notes

The problem of computing $\chi(\mathbb{R}^n)$ (in a different but equivalent form) can be traced back to Hadwiger [Had44]. Let us mention the current records in this area.

Surprisingly, it is not known whether the simple bound of Theorem 12 is sharp or not. We know only that $4\leq\chi(\mathbb{R}^2)\leq 7$, the bounds pointed by Nelson yet in 1950 (see [Gra94]). Raigorodski [Rai01] surveys known results on $\chi(\mathbb{R}^n)$ for concrete small n.

The best known general bounds are

$$(1.239...+o(1))^n\leq\chi(\mathbb{R}^n)\leq(3+o(1))^n,$$

where the upper bound is due to Larman and Rogers [LR72] and the lower to Raigorodski [Rai01].

Regarding the distribution of primes the following classical result has an accessible elementary proof (due to Erdős [Erd32]) which is presented in, for example, Aigner and Ziegler [AZ98, Chapter 2].

Theorem 16 (Bertrand's Postulate) *For every n there is some prime number p with $n<p\leq 2n$.* ▮

A related problem is to find the least value of λ so that there exists at least one prime between n and $n+O(n^\lambda)$ for all sufficiently large n. The current record seems to be the following.

Theorem 17 (Lou & Yao [LY92]) *For any $\varepsilon > 0$ there is $n_0 = n_0(\varepsilon)$ such that for all $n \geq n_0$ the there is at least one prime between n and $n + n^{6/11+\varepsilon}$.* ▮

Corollary 18 (Prime Distribution Theorem) *For any $\varepsilon > 0$ there is n_0 such that for any $n \geq n_0$ there is at least one prime between n and $(1+\varepsilon)n$.* ▮

7 Borsuk's Conjecture

Let $f(n)$ be the smallest integer f such that every bounded set in $\mathbb{R}^n$ can be partitioned into f sets of smaller diameter.

Borsuk [Bor33] introduced this problem and conjectured that $f(n) = n + 1$. The regular n-dimensional simplex shows that a partition into n sets of smaller diameter need not exist, that is, that $f(n) \geq n+1$. The conjecture was spectacularly disproved by Kahn and Kalai [KK93], whose proof (slightly modified) we now present.

Theorem 19 (Kahn & Kalai 1993) *$f(m) > 1.2^{\sqrt{m}}$ for all sufficiently large m. In particular, Borsuk's conjecture is false for all sufficiently large dimensions.*

Proof. Choose the largest n such that $k := \binom{n}{2} + n = n(n+1)/2$ is at most m. Choose a prime $p = (\frac{1}{4} + o(1))n$ with $p \geq n/4$, which is possible by known results on the distribution of primes. Fix an orthonormal basis $((\boldsymbol{e}_i)_{i\in[n]}, (\boldsymbol{f}_{ij})_{1\leq i\leq j\leq n})$ in $\mathbb{R}^k$, and define

$$\Phi(x_1,\ldots,x_n) = \sum_{1\leq i<j\leq n} x_i x_j \boldsymbol{f}_{ij} + \alpha \sum_{i=1}^{n} x_i \boldsymbol{e}_i,$$

where $\alpha = \sqrt{4p-n}$. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we have

$$\begin{aligned}\Phi(\boldsymbol{x}) \cdot \Phi(\boldsymbol{y}) &= \sum_{1\leq i<j\leq n} x_i x_j y_i y_j + \alpha^2 \sum_{i=1}^{n} x_i y_i \\ &= \frac{1}{2}(\boldsymbol{x}\cdot\boldsymbol{y})^2 + \alpha^2\, \boldsymbol{x}\cdot\boldsymbol{y} - \frac{1}{2}\sum_{i=1}^{n} x_i^2 y_i^2. \end{aligned} \tag{25}$$

For $A \subset [n]$ define $\boldsymbol{v}_A = 2\boldsymbol{\chi}_A - \boldsymbol{1} \in \mathbb{R}^n$, that is, $v_{A,i} = 1$ if $i \in A$ and $v_{A,i} = -1$ if $i \notin A$. Let $Y = \{\boldsymbol{v}_A : A \in \binom{[n]}{p-1}\}$. For $\boldsymbol{x}, \boldsymbol{y} \in Y$ we have by (25) that

$$\Phi(\boldsymbol{x}) \cdot \Phi(\boldsymbol{y}) = (\boldsymbol{x}\cdot\boldsymbol{y})^2/2 + \alpha^2\, \boldsymbol{x}\cdot\boldsymbol{y} - n/2.$$

In particular, $\|\Phi(\boldsymbol{x})\| = n^2/2 + \alpha^2 n - n/2$, $\boldsymbol{x} \in Y$, so $\Phi(Y)$ lies on a sphere centred at the origin. Hence, two points of $\Phi(Y)$ are at distance $\mathrm{diam}(\Phi(Y))$ if and only if their scalar product is the smallest possible.

The minimum of $x^2/2 + \alpha^2 x - n/2$ is attained for $x = -\alpha^2 = n - 4p$. Note that if $|A \cap B| = p - 1$ for some $A, B \in \binom{[n]}{2p-1}$, then $\boldsymbol{v}_A \cdot \boldsymbol{v}_B = n - 4p$. Thus if $Z \subset \Phi(Y)$ has smaller diameter, then no two edges of

$$\mathcal{H} := \left\{ A \in \binom{[n]}{2p-1} : \Phi(\boldsymbol{v}_A) \in Z \right\}$$

intersect in precisely $p-1$ vertices. The Modular 2-Wise Frankl–Wilson Inequality (with $L = [0, p-2]$) implies that $|\mathcal{H}| \le \sum_{i=0}^{p-1} \binom{n}{i}$, which is in turn at most $2\binom{n}{p-1}$ as, for $i \in [p-1]$,

$$\binom{n}{i}\binom{n}{i-1}^{-1} = \frac{n-i+1}{i} \ge \frac{n-p+2}{p-1} \ge 2.$$

(We assume that $p \le (n+4)/3$.)

Thus we need at least $|\Phi(Y)|/2\binom{n}{p-1}$ parts of smaller diameter to partition $\Phi(Y)$. This implies that

$$f(m) \ge f(k) \ge \frac{\binom{n}{2p-1}}{2\binom{n}{p-1}} = \left(\frac{(3/4)^{3/4}(1/4)^{1/4}}{(1/2)^{1/2}(1/2)^{1/2}} + o(1)\right)^n > (1.1397 + o(1))^n.$$

(We used Stirling's formula.) As $n \ge \sqrt{2m} - 1$, we have

$$f(m) \ge (1.1397 + o(1))^{\lceil \sqrt{2m}-1 \rceil} > (1.203 + o(1))^{\sqrt{m}},$$

which implies the theorem. ▮

Notes

Borsuk's conjecture is true under some extra restrictions (centrally symmetric bodies, bodies with smooth surface, and for all bodies in dimension 2 and 3): see Boltyanski and Gohberg [BG85].

The current record bounds on $f(n)$ are

$$(1.225... + o(1))^{\sqrt{n}} < f(n) < (1.224... + o(1))^n,$$

where the lower bound is due to Raigorodski (see [Rai01]) and the upper to Schramm [Sch88].

It is of interest to find the smallest n for which Borsuk's conjecture fails. The above argument of Kahn and Kalai (with carefully chosen n, k, l) provides a counterexample for $n = 1325$. Raigorodski's [Rai01] modification of their method provides counterexamples for all $n \ge 651$.

The smallest known n with $f(n) > n+1$ is 298, due to Hinrichs and Richter [HR02].

8 Borsuk's Conjecture and Leech Lattice

Here we present the results from Hinrichs [Hin02]. His idea was to use the following properties of $M \subset \Lambda_{24}$, the set of vectors of minimal length 1 in the Leech lattice Λ_{24}.

1. $M \subset 2^{-5/2}\mathbb{Z}^{24}$.

2. $|M| = 196560$.

3. M is a *spherical P-code* with $P := \{-1, \mp\frac{1}{2}, \mp\frac{1}{4}, 0\}$. (That is, M lies on the *unit sphere*

$$S^n := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| = 1\}$$

and the scalar products of pairs of distinct elements of M belong to P.)

In other words, $|M|$ is very large, while there are only 6 possible distances between pairs of elements from M. So it is plausible that some distance is 'hard to miss'.

Itself M is not good for disproving Borsuk's conjecture. But let us play with the idea of Kahn and Kalai: take $n = 24 + 24 + \binom{24}{2} = 324$, fix an orthonormal basis

$$\big((\boldsymbol{e}_i)_{i\in[24]}, (\boldsymbol{f}_i)_{i\in[24]}, (\boldsymbol{g}_{i,j})_{1\le i<j\le 24}\big)$$

in $\mathbb{R}^{324}$, and define

$$\Phi(\boldsymbol{x}) = c_1 \sum_{i=1}^{24} x_i^2 \boldsymbol{e}_i + c_2 \sum_{1\le i<j\le 24} x_i x_j \boldsymbol{g}_{ij} + c_3 \sum_{i=1}^{24} x_i \boldsymbol{f}_i, \quad \boldsymbol{x} \in \mathbb{R}^{24},$$

for some constants c_1, c_2, c_3. (We include x_i^2's as now x_i is not confined to ± 1.) As before, we want $\Phi(M)$ to lie on a sphere. For $\boldsymbol{x} \in S^{24}$ we have

$$\begin{aligned}\Phi(\boldsymbol{x}) \cdot \Phi(\boldsymbol{x}) &= c_1^2 \sum_{i=1}^{24} x_i^4 + c_2^2 \sum_{1\le i<j\le 24} x_i^2 x_j^2 + c_3^2 \sum_{i=1}^{24} x_i^2 \\ &= c_1^2 + (c_2^2 - 2c_1^2)\left(\sum_{1\le i<j\le 24} x_i^2 x_j^2\right) + c_3^2,\end{aligned}$$

where we used the fact that $\sum_{i=1}^{24} x_i^2 = 1$.

Let us require that $\Phi(S^{24}) \subset S^n$, which is achieved by having $c_2 = c_1\sqrt{2}$ and $c_1^2 + c_3^2 = 1$. Now, we have

$$\Phi(\boldsymbol{x}) \cdot \Phi(\boldsymbol{y}) = \psi(\boldsymbol{x} \cdot \boldsymbol{y}),$$

where $\psi(a) = c_1^2 a^2 + c_3^2 a$, $a \in \mathbb{R}$. Recall that the smaller is the scalar product of two unit vectors, then the larger is the distance between them. Tweaking c_1 and c_3 we have much freedom in deciding which a minimises $\psi(a)$. But $\boldsymbol{x} \cdot \boldsymbol{y} \in P$ for $\boldsymbol{x}, \boldsymbol{y} \in M$ so, with a foresight, let $c_1 = 2/\sqrt{5}$ and $c_3 = 1/\sqrt{5}$. Then we have

$$\Phi(\boldsymbol{x}) \cdot \Phi(\boldsymbol{y}) = \frac{4}{5}\,(\boldsymbol{x} \cdot \boldsymbol{y})(\boldsymbol{x} \cdot \boldsymbol{y} + 1/4),$$

that is, for $\boldsymbol{x}, \boldsymbol{y} \in M$,

$$\|\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{y})\| = \operatorname{diam}(\Phi(M)) \iff \boldsymbol{x} \cdot \boldsymbol{y} \in \{-1/4, 0\}.$$

Let C be a subset of M such that $\operatorname{diam}(\Phi(C)) < \operatorname{diam}(\Phi(M))$. Observe that the scalar products in C lie in $Q := \{-1, \mp\frac{1}{2}, \frac{1}{4}\}$. Let us try to deduce from this that $|C|$ is small.

Of course, the smaller Q is, the better bounds we can obtain. A helpful observation is that the scalar product -1 occurs very rarely. (For any vector there is at most one 'antipodal' vector.) If C' is obtained from C by deleting one element from each pair of antipodal vectors, then C' is a spherical $\{\mp\frac{1}{2}, \frac{1}{4}\}$-code and $|C| \le 2|C'|$.

However, we can improve on the last inequality by noting that $C'' := C \setminus C'$ is a sperical $\{\mp\frac{1}{2}\}$-code: $\boldsymbol{x} \cdot \boldsymbol{y} \ne \frac{1}{4}$ for any $\boldsymbol{x}, \boldsymbol{y} \in C''$, since otherwise $\boldsymbol{x}, (-\boldsymbol{y}) \in C$ would have inner product $-\frac{1}{4}$, which contradicts our assumption on C. So, we should be able to deduce better bounds on $|C''|$ and consequently on

$$|C| = |C'| + |C''|. \tag{26}$$

Let us first estimate $|C''|$. The familiar idea (from Theorem 9) is to define some polynomials $L_{\boldsymbol{c}}$ such that $L_{\boldsymbol{c}}(\boldsymbol{x})$ is zero if and only if $\boldsymbol{c} = \boldsymbol{x}$, $\boldsymbol{c}, \boldsymbol{x} \in C''$. For example, $L_{\boldsymbol{c}}(\boldsymbol{x}) = (2\,\boldsymbol{c}\cdot\boldsymbol{x} + 1)(2\,\boldsymbol{c}\cdot\boldsymbol{x} - 1)$ does the job. However, the following $\pmod 2$-trick produces far better results!

Consider the linear polynomials $L_{\boldsymbol{c}}$, $\boldsymbol{c} \in C''$, defined by

$$L_{\boldsymbol{c}}(\boldsymbol{x}) = 2\,\boldsymbol{c}\cdot\boldsymbol{x} + 1.$$

These polynomials have coefficients in the field $\mathbb{Q}(\sqrt{2})$. (Recall that $M \subset 2^{-5/2}\,\mathbb{Z}^{24}$.) Moreover, $L_{\boldsymbol{c}}(\boldsymbol{c}) = 3$, $L_{\boldsymbol{c}}(\boldsymbol{x}) = 0$ if $\boldsymbol{c}\cdot\boldsymbol{x} = -1/2$, and $L_{\boldsymbol{c}}(\boldsymbol{x}) = 2$ if $\boldsymbol{c}\cdot\boldsymbol{x} = 1/2$. In other words, $L_{\boldsymbol{c}}(\boldsymbol{x})$, $\boldsymbol{x} \in C''$, is odd if and only if $\boldsymbol{x} = \boldsymbol{c}$.

This implies that the polynomials $\{L_{\boldsymbol{c}} : \boldsymbol{c} \in C''\}$ are linearly independent over $\mathbb{Q}(\sqrt{2})$. Indeed, assuming that

$$\sum_{\boldsymbol{c}\in C''} (\alpha_{\boldsymbol{c}} + \beta_{\boldsymbol{c}}\sqrt{2}) L_{\boldsymbol{c}} = 0$$

for some not-all-zero $\alpha_{\boldsymbol{c}}, \beta_{\boldsymbol{c}} \in \mathbb{Q}$, we may as well assume that the $\alpha_{\boldsymbol{c}}, \beta_{\boldsymbol{c}}$ are integers which are not all even. But then the evaluation at the point $\boldsymbol{c} \in C''$ shows that $\alpha_{\boldsymbol{c}}$ and $\beta_{\boldsymbol{c}}$ have to be even for each $\boldsymbol{c} \in C''$, a contradiction.

Thus the cardinality of C'' cannot exceed 25, the dimension of the space of all linear polynomials in 24 indeterminates.

Let us move to C', which is a sperical $\{\mp\frac{1}{2}, \frac{1}{4}\}$-code. Here too we use the $\pmod 2$-trick! Consider the quadratic polynomials $P_{\boldsymbol{c}}$, $\boldsymbol{c} \in C'$, given by

$$P_{\boldsymbol{c}}(\boldsymbol{x}) = (2\,\boldsymbol{c}\cdot\boldsymbol{x} - 1)(4\,\boldsymbol{c}\cdot\boldsymbol{x} - 1), \quad \boldsymbol{x} \in \mathbb{R}^{24}.$$

Then $P_{\boldsymbol{c}}(\boldsymbol{c}) = 3$ is odd, but $P_{\boldsymbol{c}}(\boldsymbol{x})$ is even for any distinct points $\boldsymbol{c}, \boldsymbol{x}$ of a spherical $\{\mp\frac{1}{2}, \frac{1}{4}\}$-code. Thus these polynomials are linearly independent and

$$|C'| \le \binom{24}{2} + 24 + 24 + 1 = 325,$$

is at most the dimension of the linear space of polynomials of total degree at most 2 in 24 indeterminates

Putting everything together we conclude that any part $\Phi(C)$ of smaller diameter has at most $325 + 25 = 350$ points; so we need at least $\frac{196560}{350} > 651$ such parts to partition $\Phi(M)$.

Also, observe that $\Phi(M)$ lies within a 323-dimensional affine subspace of $\mathbb{R}^{324}$ consisting of all vectors for which the coordinates of the $\boldsymbol{e}_i$-basis vectors sum up to 1. This implies the following theorem.

Theorem 20 (Hinrichs [Hin02]) *Borsuk's conjecture is false in all dimensions* $n \in [323, 560]$. ∎

Notes

As with all 'Notes' sections, the material below is for your information only and therefore is not examinable.

A *lattice* is a discrete additive subgroup of $\mathbb{R}^n$. Here we define the Leech lattice $\Lambda_{24} \subset \mathbb{R}^{24}$ and describe its elements of minimal length. For further information please consult Conway and Sloane [CS99] where a definite treatment of the Leech lattice is presented.

First, we have to define the (extended) *Golay code* $\mathcal{C}_{24}$. Table 1 shows one of its many possible generator matrices G. How is it constructed? The left-hand half is obvious. The last column is the parity bit. The right-hand part of the first row is defined by the quadratic residues modulo 11, namely, $\mathsf{G}_{1,13+j} = 0$ if and only if j is a quadratic residue modulo 11. This pattern is cyclically rotated in subsequent rows.

1												1	0	1	0	0	0	1	1	1	0	1	1
	1											1	1	0	1	0	0	0	1	1	1	0	1
		1										0	1	1	0	1	0	0	0	1	1	1	1
			1									1	0	1	1	0	1	0	0	0	1	1	1
				1								1	1	0	1	1	0	1	0	0	0	1	1
					1							1	1	1	0	1	1	0	1	0	0	0	1
						1						0	1	1	1	0	1	1	0	1	0	0	1
							1					0	0	1	1	1	0	1	1	0	1	0	1
								1				0	0	0	1	1	1	0	1	1	0	1	1
									1			1	0	0	0	1	1	1	0	1	1	0	1
										1		0	1	0	0	0	1	1	1	0	1	1	1
											1	1	1	1	1	1	1	1	1	1	1	1	0

Table 1: A generator matrix G for the binary Golay code $\mathcal{C}_{24}$. Blank entries are zero.

The Golay code $\mathcal{C}_{24}$ is a 12-dimensional linear subspace of $(\mathbb{F}_2)^{24}$. In particular, it has 2^{12} codewords: to obtain a codeword $\boldsymbol{v}_I$ "decoding" $I \subset [12]$, sum up the corresponding rows: $\boldsymbol{v}_I = (\sum_{i \in I} \mathsf{G}_{i,*})^T$. This mapping is bijective as $\mathsf{G}_{[12],[12]}$ is the identity matrix.

We claim that for any $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{C}_{24}$ the *hamming distance*

$$d_H(\boldsymbol{u}, \boldsymbol{v}) := \sum_{i=1}^{24} \texttt{IF}[u_i \neq v_i]$$

is at least 8.

Here is a sketch of proof. First, check that all pairs of rows in Table 1 have inner product zero (over $\mathbb{F}_2$). Next, observe that $d_H(\boldsymbol{u} + \boldsymbol{v}) \equiv 0 \pmod 4$ for any $\boldsymbol{u}, \boldsymbol{v} \in (\mathbb{F}_2)^{24}$ with $\boldsymbol{v} \cdot \boldsymbol{u} = 0$ and $d_H(\boldsymbol{u}, \boldsymbol{0}) \equiv d_H(\boldsymbol{v}, \boldsymbol{0}) \equiv 0 \pmod 4$. Conclude that $d_H(\boldsymbol{u}, \boldsymbol{v}) \equiv 0 \pmod 4$ for any $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{C}_{24}$. It remains to prove that, for example, there is no $\boldsymbol{u} \in \mathcal{C}_{24}$ with $d_H(\boldsymbol{u}, \boldsymbol{0}) = 4$. Since $\mathsf{G}_{[12],[12]}$ is the identity matrix, it is enough to verify the last claim for sums of at most 4 rows of G. The symmetries of G reduce the last claim to a few easy cases.

If we remove one column from $\mathcal{C}_{24}$ we obtain the *Golay code* $\mathcal{C}_{23} \subset (\mathbb{F}_2)^{23}$ which also has 2^{12} codewords. Its minimal distance is 7. Since

$$1+\binom{23}{1}+\binom{23}{2}+\binom{23}{3}=2^{11}=\frac{2^{23}}{|\mathcal{C}_{23}|},$$

$\mathcal{C}_{23}$ gives a perfect packing of balls of d_H-radius 3. Let

$$W_i=\{\boldsymbol{u}\in\mathcal{C}_{23}: d_H(\boldsymbol{u},\boldsymbol{0})=i\}$$

consists of $\mathcal{C}_{23}$-codewords of *weight* i. As for each codeword $\boldsymbol{u}\in W_7$ there are $\binom{7}{4}$ vectors $\boldsymbol{v}\in(\mathbb{F}_2)^{23}$ of weight 4 and at distance 3 from $\boldsymbol{u}$, we conclude that $|W_7| = \binom{23}{4}/\binom{7}{4}=253$. Counting all vectors of $(\mathbb{F}_2)^{24}$ of weight 5 we obtain

$$|W_7|\binom{7}{5}+|W_8|\binom{8}{5}=\binom{23}{5}$$

and deduce that $|W_8|=506$. Hence, $\mathcal{C}_{24}$ has $|W_7|+|W_8|=759$ codewords of weight 8 (and the same number codewords of weight 16).

Theorem 21 *For each codeword of $\mathcal{C}_{24}$ there are 759 codewords at the hamming distance 8 or 16, one 'antipodal' codeword at distance 24, and $2^{12}-2\cdot 759-2=2576$ codewords at distance 12.* ∎

Now we are ready to define the *Leech lattice* Λ_{24} which we scale here so that its minimal-length elements lie on the unit sphere. It is generated by integer combinations of the following vectors

$$\frac{1}{4\sqrt{2}}(\mp 3,\pm 1^{(\times 23)}),\tag{27}$$

where the ∓ 3 may be in any position and the upper signs are taken on a set of coordinates where a codeword of $\mathcal{C}_{24}$ is 1. (Also, $a^{(\times k)}$ denotes k copies of a.)

The min-length elements $M\subset\Lambda_{24}$ can be explicitly described:

- $2^7\cdot 759$ of the form $\frac{1}{4\sqrt{2}}(\pm 2^{(\times 8)},0^{(\times 16)})$, where the positions of the ± 2's form one of the 759 $\mathcal{C}_{24}$-codewords of weight 8 and there are an even number of minus signs;
- $24\cdot 2^{12}$ of the form (27);
- $4\cdot\binom{24}{2}$ of the form $\frac{1}{4\sqrt{2}}(\pm 4^{(\times 2)},0^{(\times 22)})$.

Thus in total we have

$$|M|=2^7\cdot 759+24\cdot 2^{12}+4\cdot\binom{24}{2}=97152+98304+1104=196560.$$

min-length elements in Λ_{24}. Just imagine: if we put balls of radius 1/2 about the elements of Λ_{24}, then each ball touches 196560 other balls!

Using the fact that the distance between any two $\mathcal{C}_{24}$-codewords is 8, 12, 16 or 24, one should be able to deduce that the scalar product of any two distinct elements of M lies in $\{-1,\mp\frac{1}{2},\mp\frac{1}{4},0\}$.

Hinrich's argument uses the properties of M only so one could have explicitly described M as above without appealing to the Leech lattice at all.

9 Constructive Lower Bounds on Ramsey Numbers

The *Ramsey number* $r(s,t)$ is the smallest n such that any blue-red colouring of the edges of K_n yields a blue K_s or a red K_t (or both). Estimating $r(s,t)$, or even $r(t,t)$, is very difficult and the best known bounds are quite far apart.

9.1 Probabilistic Lower Bound

The following simple argument of Erdős [Erd47] gives an exponential lower bound on $r(t,t)$. Colour the edges of K_n independently, each being blue with probability 1/2. The expected number of monochromatic copies of K_t is $2^{1-\binom{k}{2}}\binom{n}{k}$. If this quantity is smaller than 1, then there is a K_t-free colouring and thus $r(t,t) > n$. It remains to estimate n. From *Stirling's formula*

$$\sqrt{2\pi k}\,(k/\mathrm{e})^k \le k! \le \mathrm{e}^{1/12k}\sqrt{2\pi k}\,(k/\mathrm{e})^k, \tag{28}$$

it follows that

$$\binom{n}{k} \le \frac{n^k}{k!} \le \frac{n^k}{\sqrt{2\pi k}\,(k/\mathrm{e})^k} \le \left(\frac{\mathrm{e}n}{k}\right)^k. \tag{29}$$

Thus, in Erdős' argument it is enough to take n with $(\mathrm{e}n/k)^k < 2^{\binom{k}{2}-1}$, which shows that

$$r(t,t) \ge \left(\tfrac{1}{\mathrm{e}} - o(1)\right) k2^{\frac{k-1}{2}} = \left(\tfrac{1}{\mathrm{e}\sqrt{2}} - o(1)\right) k2^{\frac{k}{2}}.$$

Unfortunately, the above argument (without any extra ideas) does not give any better algorithm for finding a K_t-free colouring than essentially checking all possible colourings of K_n. A challenging open problem is to find lower bounds on Ramsey numbers via explicit colourings.

9.2 Nagy's Construction

First we present the following simple construction due to Nagy [Nag72].

Theorem 22 (Nagy [Nag72]) $r(t+1,t) \ge \binom{t-1}{3}$.

Proof. Let $n = \binom{t-1}{3}$. Identify the vertices of K_n with 3-element subsets of $[t-1]$. Colour K_n by the following rule:

$$c(\{A,B\}) = \begin{cases} \text{blue}, & \text{if } |A\cap B| = 1,\\ \text{red}, & \text{otherwise}, \end{cases} \qquad A,B \in \binom{[t-1]}{3}.$$

Weak Fisher's inequality (Corollary 10) tells us that this colouring has no blue K_{t+1} and the Odd Rules Theorem (Theorem 7) tells us that it has no red K_t. ∎

9.3 Supexponential Bounds

Let t be large. Do we get better bounds on $r(t,t)$ if we generalise Nagy's idea by taking $\binom{[n]}{k}$ for the vertex set? Like in Nagy's construction we colour an edge $\{A,B\}$, $A,B \in \binom{[n]}{k}$, depending on the intersection size only; say blue (the first colour) if $|A\cap B| \in C_1$ and red (the second colour) if $|A\cap B| \in C_2$, where $C_1 \cup C_2 = [k-1]$ are some sets which we are about to define.

Let $i = 1,2$. The maximum order of a colour-i clique equals the maximum size of a hypergraph $\mathcal{H}_i \subset \binom{[n]}{k}$ with $|X\cap Y| \in C_i$ for any distinct $A,B \in \mathcal{H}_i$ and this should be less than t. The Modular 2-Wise Frankl–Wilson Inequality allows us to control the size of $\mathcal{H}_i$. In order to apply it for some p_i and L_i, we must have $k \notin L_i \pmod{p_i}$ but $C_i \subset L_i \pmod{p_i}$. Then we will obtain $|\mathcal{H}_i| \le \sum_{j=1}^{|L_i|}\binom{n}{j}$. So it is reasonable that we should try to make L_1 and L_2 have the same size s, where s is as small as possible.

If $p_i > k-1$, then it follows that $|L_i| \ge |C_i|$ because no two elements of C_i can have the same residue modulo p_i. If this happens for both $i=1,2$, then $s \ge k/2$ and the best bound we get is $r(t,t) = \Omega(t^2)$ only (*convince yourself of this*) — even worse than Nagy's bound.

So, suppose that $p_1 \le k-1$. Then $M \cap C_1 = \emptyset$, so $M \subset C_2$, where

$$M := \{k - jp_1 : 1 \le j \le k/p_1\}.$$

In particular, $p_2 \ne p_1$. But then no two elements $a,b \in M$ can have equal residues modulo p_2 for otherwise $k \equiv k - |a-b| \pmod{p_2}$, but the latter element belongs to $M \subset C_2$ which we do not allow. Thus $|L_2| \ge |M|$.

So, given all this, the most promising option is to let $s = p_1 - 1$, $C_2 = L_2 = M$, and let L_1 consist of all residues modulo p_1 except that of k. Now, that $|L_1| = |L_2| = s$ is fixed, it is clearly advantageous to have k as large as possible. The maximum k is easily seen to be $p_1^2 - 1$.

In summary, we made our choice on the following colouring. The vertex set is $\binom{[n]}{p_1^2-1}$ and we colour

$$c(\{A,B\}) = \begin{cases} \text{red}, & \text{if } |A\cap B| \equiv p_1 - 1 \pmod{p_1}, \\ \text{blue}, & \text{otherwise}, \end{cases} \qquad A,B \in \binom{[n]}{p_1^2-1}.$$

Now applying the Modular 2-Wise Frankl–Wilson Inequality (Theorem 9) twice, once for p_1 and $L_1 = [0, p_1 - 2]$ and another time for an arbitrary prime $p_2 \ge k$ and

$$L_2 = M = \{p_1^2 - 1 - jp_1 : j \in [p_1 - 1]\},$$

we obtain that the order of any monochromatic clique (in either colour) does not exceed $\sum_{j=0}^{p_1-1}\binom{n}{j}$.

Now, our objective is to choose a prime p_1 and an integer n such that $\sum_{j=0}^{p_1-1}\binom{n}{j} < t$; then we can conclude that $r(t,t) \ge \binom{n}{p_1^2-1}$. Of course, we wish to make $\binom{n}{p_1^2-1}$ as large as possible.

The remaining calculation are straightforward. Let us denote $p = p_1$ to avoid writing the subscript all the time.

As $n \geq p^2 - 1 \geq 3p - 4$, we have $\binom{n}{i}/\binom{n}{i-1} \geq 2$ for $i \leq p-1$ and $\sum_{j=0}^{p-1}\binom{n}{j} < 2\binom{n}{p-1}$. By (29) it is enough to have $t \geq 2(\mathrm{e}n/(p-1))^{p-1}$, so we let

$$n = \left\lfloor \frac{(p-1)(t/2)^{1/(p-1)}}{\mathrm{e}} \right\rfloor \tag{30}$$

On the other hand

$$\binom{n}{k} = \frac{n}{k} \times \frac{n-1}{k-1} \times \cdots \times \frac{n-k+1}{1} \geq (n/k)^k, \tag{31}$$

so let us choose p such that the last expression (given (30)) is large. Taking the logarithm

$$\begin{aligned} \ln((n/(p^2-1))^{p^2-1}) &= (p^2-1)\ln n - 2p^2 \ln p + O(p^2) \\ &= (p^2-1)\frac{\ln t}{p-1} - p^2 \ln p + O(p^2) \\ &= p \ln t - p^2 \ln p + O(p^2 + \ln t). \end{aligned}$$

Taking the derivative $\frac{\mathrm{d}}{\mathrm{d}p}(p \ln t - p^2 \ln p) = \ln t - 2p \ln p - p$, we see that the maximum is achieved at the (unique) root of $2p \ln p + p = \ln t$. We cannot compute it exactly but the approximation

$$p = (1 + o(1))\frac{\ln t}{2 \ln \ln t}, \tag{32}$$

gives us $\ln((n/k)^k) = (\frac{1}{4} + o(1))(\ln t)^2/\ln \ln t$. From the Prime Distribution Theorem (Corollary 18) we know that we can always choose a prime p of the form (32), which gives us the following result.

Theorem 23 (Frankl [Fra77]) *We can find an explicit K_t-free 2-colouring of K_n with*

$$n = t^{(1+o(1))\frac{\ln t}{4 \ln \ln t}}. \blacksquare$$

Notes

Theorem 23 was proved by Frankl [Fra77] but his proof of the non-existence of large monochromatic cliques was complicated (and his construction was slightly different too). Frankl and Wilson [FW81] gave a simpler proof, similar to the one presented here.

10 The Shannon Capacity of Graphs

10.1 Motivation

The following question was posed by Shannon [Sha56], the founder of information theory. Suppose we transmit messages (sequences of letters from an *alphabet* V) across a channel where some symbols may be distorted. What is the maximum rate of transmission such that the receiver can always detect if any errors have happened?

We define the *confusion graph* G which has V for the vertex set with $a, b \in V$ being connected if and only if these two letters can be confused during transmission. Clearly, two distinct messages of length n can be confused if and only if for any $i \in [n]$ their i-th letters either are equal or can be confused. In the graph-theoretic terms this prompts us to define the *graph product* $G_1 \times \cdots \times G_n$ which has the vertex set

$$V(G_1 \times \cdots \times G_n) := V(G_1) \times \cdots \times V(G_n),$$

with distinct $\boldsymbol{a}$ and $\boldsymbol{b}$ being connected by an edge if and only if $a_i = b_i$ or $\{a_i, b_i\} \in E(G_i)$ for all $i \in [n]$. The confusion graph for strings of length n is thus G^n, the product of n copies of the graph itself.

The sender and the receiver agree on some set $U \subset V^n$ of messages that are used for transmission. Clearly, errors can always be detected if and only if U is an independent set in G^n. Thus, we can have at most $\alpha(G^n)$ different messages, where $\alpha(G^n)$ is the *independence number* of G^n.

As m binary bits can generate 2^m messages, it is natural to measure the amount of information by taking the logarithm base 2 of the total number of possible messages. In our settings, the *information rate* (the number of bits per letter) is at most

$$\frac{\log_2 \alpha(G^n)}{n} = \log_2 \sqrt[n]{\alpha(G^n)}.$$

Disregarding the logarithm we thus arrive at the following definition: the *zero-error capacity* (or the *Shannon capacity*) of a graph G is

$$\Theta(G) := \sup_{n \geq 1} \sqrt[n]{\alpha(G^n)}.$$

Example 24 *For the* 5*-cycle* C_5 *we have* $\alpha(C_5) = 2$*. However,* $\alpha(C_5^2) \geq 5$*: let* $C_5 := ([5], \{\{i, i+1\} : i \in \mathbb{Z}_5\})$*, then the set*

$$\{(1,1), (2,3), (3,5), (4,2), (5,4)\}$$

is independent. Thus $\Theta(C_5) \geq \sqrt{5}$*.*

Another question to ask is “Given the confusion graph G, what is the maximum transmission rate with *error correction* (when we require that the receiver can always recover the original message)?”

The zero-error model is applicable, for example, when two computers communicate each with the other and the receiver can repeat the request if an error has occurred. But when it is, for example, a remote satellite transmitting data, then it is at least impractical (if altogether possible) to make it send the data again so we would rather try to correct errors.

If two messages $\boldsymbol{a}, \boldsymbol{b} \in V^n$ can be distorted during transmission into the same message, then for each $i \in [n]$ the vertices a_i, b_i of G are equal or adjacent, or have a common neighbour. Thus the maximum information rate with error correction is $\log_2(\Theta(H))$, where H is the graph on V with two vertices being connected if and only if the distance between them in G is at most 2. We see that this problem reduces to the computation of the Shannon capacity Θ (although for a different graph).

Execises and Further Reading

[AZ98]: Chapter 28 contains a very good exposition on the Shannon capacity.

10.2 The Shannon Capacity of a Union

Let *disjoint union of graphs* G and H, denoted by $G \sqcup H$, is the graph whose vertex set is the disjoint union of $V(G)$ and $V(H)$ and whose edge set is the (disjoint) union of $E(G)$ and $E(H)$.

Shannon [Sha56] proved that $\Theta(G \sqcup H) \geq \Theta(G) + \Theta(H)$ and conjectured that the equality always holds. This was disproved by Alon [Alo98] who found a graph G (or even a series of graphs) such that $\Theta(G) + \Theta(\overline{G}) < \Theta(G \sqcup \overline{G})$, where $\overline{G}$ denotes the graph-theoretic *complement* of G.

Let us present this result.

Why do we take a graph and its complement? Because then it is easy to bound $\Theta(G \sqcup \overline{G})$ from below as follows. Let

$$V(G \sqcup \overline{G}) = \{a_1, \ldots, a_m\} \cup \{b_1, \ldots, b_m\}$$

so that $\{a_i, a_j\}$ is an edge iff $\{b_i, b_j\}$ is not. Then the set

$$\{(a_i, b_i) : i \in [m]\} \cup \{(b_i, a_i) : i \in [m]\}$$

is an independent set of vertices in $(G \sqcup \overline{G})^2$, which shows the following lemma.

Lemma 25 *For any graph G of order m we have $\Theta(G \sqcup \overline{G}) \geq \sqrt{2m}$.* ▮

Thus our objective is to find a graph G of given order m such that both G and its complement have small capacity. We know that the capacity is at least the independence number. Thus, if we construct a counterexample this way, then both G and $\overline{G}$ have small independence numbers. Let us take some G that is known to have the latter property — maybe the capacities are also small by some magic!

From Section 9 we know two good candidates for G: the random graph with edge probability 1/2 and the intersection graph. Unfortunately, nobody has yet been able to make the probabilistic approach work: Alon [Alo98, Conjecture 5.1] conjectures that a random graph of order m has almost surely capacity $O(\log m)$ and this question is still wide open. We have to exploit the other direction.

Thus let G be the familiar graph on $\binom{[n]}{p^2-1}$ where $\{A, B\}$ is an edge if and only if $|A \cap B| \equiv p-1 \pmod{p}$. How can we bound the Shannon capacity of G from above? Whatever proof we find, it will also give a bound on the independence number of G. Maybe our proof of the upper bound on $\alpha(G)$ from Section 9 generalises to $\Theta(G)$? If we expand the whole proof (including the proof of Theorem 9), then we obtain the following outline.

To each vertex $A \in V(G)$ we associate a certain polynomial f_A (over some field) and a vector $\boldsymbol{c}_A$ (in our case $\boldsymbol{v}_A = \boldsymbol{\chi}_A$) such that

1. For each $A \in V(G)$ we have $f_A(\boldsymbol{v}_A) \neq 0$.

2. If A and B are distinct non-adjacent vertices of G, then $f_A(\boldsymbol{v}_B) = 0$.

Then we conclude that for any independent set $\mathcal{H} \subset V(G)$, the polynomials f_A, $A \in \mathcal{H}$, are linearly independent: indeed, if $\sum_{A\in\mathcal{H}} \beta_A f_A = 0$ for some scalars β's, then the evaluation of this identity on $\boldsymbol{v}_A$ shows that each $\beta_A = 0$. Thus, if all our polynomials belong to some linear subspace F, then we conlude that $\alpha(G) \le \dim(\mathsf{F})$.

Of course, the above method can be applied to any graph G. Namely, let F be a linear subspace of $\mathbb{F}[x_1, \ldots, x_r]$, the space of all polynomials in r variables $x_1, \ldots, x_r$ over the field $\mathbb{F}$. We say that a graph G admits a *respresentation* over F, if we can find $f_A \in \mathsf{F}$ and $\boldsymbol{v}_A \in (\mathbb{F})^r$, for each $A \in V(G)$, such that the above Conditions 1 and 2 hold. Then the following lemma is true.

Lemma 26 *If G has a representation over F, then $\alpha(G) \le \dim(\mathsf{F})$.* ▮

Can we hope to prove that $\Theta(G) \le \dim(\mathsf{F})$ under the same assumptions? To do this, we have to show that $\alpha(G^n) \le (\dim(\mathsf{F}))^n$. For the latter it is enough to find a representation of G^n over a subspace of dimension $(\dim(\mathsf{F}))^n$. One idea that comes to the mind is to take

$$(\mathsf{F})^{\otimes n} := \mathsf{F} \otimes \cdots \otimes \mathsf{F},$$

the *tensor product* of n copies of F, because this is an operation that gives the right dimension. Namely,

$$(\mathsf{F})^{\otimes n} := \mathrm{Span}\{f_1 \otimes \cdots \otimes f_n : f_i \in \mathsf{F},\ i \in [n]\},$$

where $f_1 \otimes \cdots \otimes f_n \in \mathbb{F}[(x_{i,j})_{i\in[n],j\in[r]}]$ is a polynomial in nr variables defined by

$$(f_1 \otimes \cdots \otimes f_n)(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \prod_{i=1}^{n} f_i(\boldsymbol{x}_i), \quad \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in (\mathbb{F})^r.$$

Please do not be scared by this notation. For example, if $f \in \mathbb{F}[x]$ and $g \in \mathbb{F}[y]$, then the polynomial $h = f \otimes g \in \mathbb{F}[x, y]$ is defined by $h(x, y) = f(x) \cdot g(y)$ — as simple as that!

It is easy to see that $\dim((\mathsf{F})^{\otimes n}) \le (\dim(\mathsf{F}))^n$: if $\{e_1, \ldots, e_d\}$ spans F, then $\{e_{i_1} \otimes \cdots \otimes e_{i_n} : \boldsymbol{i} \in [d]^n\}$ spans $(\mathsf{F})^{\otimes n}$. (*Prove that in fact* $\dim((\mathsf{F})^{\otimes n}) = (\dim(\mathsf{F}))^n$.)

The most obvious way to represent G^n over $(\mathsf{F})^{\otimes n}$ is to correspond to $\boldsymbol{a} = (a_1, \ldots, a_n) \in (V(G))^n$ the polynomial

$$f_{\boldsymbol{a}} := f_{a_1} \otimes \cdots \otimes f_{a_n} \in (\mathsf{F})^{\otimes n}$$

and the vector

$$\boldsymbol{v}_{\boldsymbol{a}} := (\boldsymbol{v}_{a_1}, \ldots, \boldsymbol{v}_{a_n}) \in (\mathbb{F})^{nr}.$$

(Here, for convenience, we view vectors as rows, not as columns.)

Let us check whether this gives a representation. For $\boldsymbol{a}, \boldsymbol{b} \in (V(G))^n$ we have

$$f_{\boldsymbol{a}}(\boldsymbol{v}_{\boldsymbol{b}}) = (f_{a_1} \otimes \cdots \otimes f_{a_n})(\boldsymbol{v}_{b_1}, \ldots, \boldsymbol{v}_{b_n}) = \prod_{i=1}^{n} f_{a_i}(\boldsymbol{v}_{b_i}).$$

Now, $f_{\boldsymbol{a}}(\boldsymbol{v}_{\boldsymbol{a}}) = \prod_{i=1}^n f_{a_i}(\boldsymbol{v}_{a_i}) \neq 0$ by Condition 1. On the other hand, if $\boldsymbol{a}, \boldsymbol{b}$ are not adjacent in G^n, then a_i, b_i are not adjacent in G for some $i \in [n]$; by Condition 2 we have $f_{a_i}(\boldsymbol{v}_{b_i})$ (and hence $f_{\boldsymbol{a}}(\boldsymbol{v}_{\boldsymbol{b}})$) is zero. By Lemma 26 we conclude that $\alpha(G^n) \leq (\dim(\mathsf{F}))^n$, which implies the following result.

Theorem 27 *If a graph G has a representation over F, then $\Theta(G) \leq \dim(\mathsf{F})$.* ▮

Now, the remainder is routine. Recall that G is a graph on $\binom{[n]}{p^2-1}$ in which $\{A, B\}$ is an edge if and only if $|A \cap B| \equiv p - 1 \pmod{p}$.

To prove that $\alpha(G)$ is small we applied in Section 9 the Modular Frankl–Wilson Inequality (Theorem 9) with $L = [0, p-2]$. Its proof tells us to assign to $A \in V(G)$ the vector $\boldsymbol{\chi}_A \in (\mathbb{F}_p)^n$ and the polynomial $\bar{f}_A \in \mathbb{F}_p[x_1, \ldots, x_n]$, which is obtained from

$$f_A(x_1, \ldots, x_n) = \prod_{j=0}^{p-2} (\boldsymbol{x} \cdot \boldsymbol{\chi}_A - j),$$

by repeatedly replacing each x_i^2 by x_i. Clearly, $\bar{f}_A \in \mathsf{F}$, where F is the linear subspace of all multilinear polynomials in n variables of degree at most $p-2$ over $\mathbb{F}_p$.

We know that this should give a representation over F, but let us double-check. We have

$$\bar{f}_A(\boldsymbol{\chi}_A) = f_A(\boldsymbol{\chi}_A) = \prod_{j=0}^{p-2} (|A| - j) \neq 0, \quad A \in V(G),$$

as $|A| = p^2 - 1$. If $A, B \in V(G)$ are not adjacent, then $|A \cap B| \in [0, p-2] \pmod{p}$ and $\bar{f}_A(\boldsymbol{\chi}_B) = f_A(\boldsymbol{\chi}_B) = 0$, as required. Hence, by Theorem 27 we conclude that

$$\Theta(G) \leq \dim(\mathsf{F}) = \sum_{j=0}^{p-1} \binom{n}{j}. \tag{33}$$

(The latter equality was established in the proof of Theorem 9.)

To prove that $\alpha(\overline{G})$ is small, in Section 9 we applied Theorem 9 for

$$L = \{p^2 - 1 - ip : i \in [1, p-1]\}$$

and any prime $p_2 > p^2 - 1$. (In fact, we could take $\mathbb{R}$ instead of $\mathbb{F}_{p_2}$, without any changes in the argument below.) The proof of Theorem 9 suggests to correspond $A \in V(\overline{G})$ to $\boldsymbol{\chi}_A \in (\mathbb{F}_{p_2})^n$ and to $\bar{f}_A$ which is obtained from

$$f_A(x_1, \ldots, x_n) = \prod_{i \in L} (\boldsymbol{x} \cdot \boldsymbol{\chi}_A - i)$$

by reducing each exponent to 1, as usual. Clearly,

$$\bar{f}_A(\boldsymbol{\chi}_A) = f_A(\boldsymbol{\chi}_A) = \prod_{i \in L} (p^2 - 1 - i) \neq 0.$$

On the other hand, if A, B are not connected in $\overline{G}$, then $|A \cap B| \equiv p - 1 \pmod{p}$, which implies that $|A \cap B| \in L$ and $\bar{f}_A(\boldsymbol{\chi}_B) = 0$. By Theorem 27 we conclude that

$$\Theta(\overline{G}) \leq \sum_{j=0}^{p-1} \binom{n}{j}. \tag{34}$$

And now we are done. The calculations of Section 9 say that we can choose appropriate p and n so that, by (33) and (34), we have $\Theta(G), \Theta(\overline{G}) < t$, while $v(G) \geq t^{(1+o(1))\ln t/4\ln\ln t}$. Now from Lemma 25 we conclude the following.

Theorem 28 (Alon [Alo98]) *For any t there is a graph G such that $\Theta(G) < t$ and $\Theta(\overline{G}) < t$ but*

$$\Theta(G \sqcup \overline{G}) \geq t^{(1+o(1))\frac{\ln t}{8\ln\ln t}}. \blacksquare$$

10.3 Upper Bounds via Linear Programming

To obtain upper bounds on the Shannon capacity, we have to find a method for proving upper bounds on the independence number. A basic observation is that if we cover the vertex set of G by k cliques, then $\alpha(G) \leq k$ because an independent set can share at most one vertex with a clique. (A *clique* is a set of vertices that spans a complete graph.) Good news is that clique coverings always exist (e.g. by one-vertex cliques). Bad news is that for some graphs they produce weak bounds. For example, the smallest number of cliques covering C_5 is three, while $\alpha(C_5) = 2$.

One idea is to take 'fractional cliques' as follows. Let $\mathcal{C}(G)$ be the set of all cliques of G. A collection of cliques can be naturally defined by specifying for each $D \in \mathcal{C}(G)$ the number $y_D \in \{1, 0\}$ which indicates whether D is included or not. Now, the property that each vertex lies in at least one clique amounts to

$$\sum_{\substack{D\in\mathcal{C}(G)\\ D\ni v}} y_D \geq 1, \quad \text{for every vertex } v \in V(G). \tag{35}$$

Let us see what happens if we make a *relaxation* of this property by not requiring the y_D's to be integers. Namely, an *FCC* (a *fractional clique covering*) of a graph G is a set of non-negative reals $\boldsymbol{y} = (y_D)_{D\in\mathcal{C}(G)}$ such that (35) holds.

It is plausible that for any FCC $\boldsymbol{y}$ we have $\alpha(G) \leq \sum_{D\in\mathcal{C}(G)} y_D$, that is,

$$\alpha(G) \leq \gamma(G) := \inf\Big\{ \sum_{D\in\mathcal{C}(G)} y_D : \boldsymbol{y} \in FCC(G)\Big\}. \tag{36}$$

The infimum above is in fact minimum because $FCC(G)$ is a compact set.

We see that $\gamma(G)$ is the minimum of a certain linear programming problem so we can use the standard techniques and results of the area. For example, the linear programming duality (see e.g. [Chv83]) tells us that the right-hand side of (36) equals $\alpha'(G) := \max_{\boldsymbol{x}} \sum_{v\in V} x_v$, where the maximum is taken over all $\boldsymbol{x} \geq \mathbf{0}$ such that

$$\sum_{v\in D} x_v \leq 1, \quad \text{for any } D \in \mathcal{C}(G). \tag{37}$$

Given the FCC-notation, we call an $\boldsymbol{x} \geq \mathbf{0}$ satisfying (37) an *FIS* (a *fractional independent set*); $\alpha'(G)$ is the *FIN* (the *fractional independence number*) of G.

Thus it is known that $\gamma(G) = \alpha'(G)$, but we keep separate symbols. (The α'-notation is the standard one.)

So, perhaps we can try to prove that $\alpha(G) \le \alpha'(G)$ and, indeed, it is trivial: choose a maximum independent set $U \subset V$ and let $\boldsymbol{x}$ be the characteristic vector of U.

However, it is not immediately clear how we can prove $\Theta(G) \le \alpha'(G)$. On the other hand, the definition of γ makes the following claim easy.

Lemma 29 *$\gamma(G^n) \le (\gamma(G))^n$.*

Proof. Let $\boldsymbol{y} \in FCC(G)$ be arbitrary. Define $\boldsymbol{z}$ be to zero except $z_{D_1 \times \cdots \times D_n} = \prod_{i=1}^n y_{D_i}$ for $D_1, \dots, D_n \in \mathcal{C}(G)$. (Of course, $D_1 \times \cdots \times D_n \in \mathcal{C}(G^n)$.) It is easy to see that $\boldsymbol{z}$ is an FCC:

$$\sum_{\substack{D \in \mathcal{C}(G^n) \\ D \ni \boldsymbol{a}}} z_D = \prod_{i=1}^n \sum_{\substack{D_i \in \mathcal{C}(G) \\ D_i \ni a_i}} y_{D_i} \ge 1, \quad \boldsymbol{a} \in V(G^n).$$

Hence,

$$\gamma(G^n) \le \sum_{D \in \mathcal{C}(G^n)} z_D = \prod_{i=1}^n \sum_{D_i \in \mathcal{C}(G)} y_{D_i} = \left(\sum_{D \in \mathcal{C}(G)} y_D \right)^n .$$

As $\boldsymbol{y} \in FCC(G)$ was arbitrary, the lemma follows. ▮

Now, taking for granted that $\alpha' = \gamma$, we can easily deduce that $\Theta(G) \le \gamma(G)$ because by Lemma 29 we have

$$\alpha(G^n) \le \alpha'(G^n) = \gamma(G^n) \le (\gamma(G))^n.$$

Thus to give a self-contained proof of $\Theta(G) \le \gamma(G)$, it remains to show that for any G we have $\gamma(G) \ge \alpha'(G)$. The linear programming duality consists of two opposite inequalities, one being trivial and the other hard. The notorious Murphy's Laws would suggest that we need the harder inequality, but this case is a fortunate exception! The proof of $\gamma(G) \ge \alpha'(G)$ is trivial: for any FCC $\boldsymbol{y}$ and any FIS $\boldsymbol{x}$ we have

$$\sum_{D \in \mathcal{C}(G)} y_D \ge \sum_{D \in \mathcal{C}(G)} y_D \sum_{v \in D} x_v = \sum_{v \in V} x_v \sum_{\substack{D \in \mathcal{C}(G) \\ D \ni v}} y_D \ \ge \sum_{v \in V} x_v.$$

Now we can discard α' altogether especially that it usually easier to prove an upper bound on $\gamma(G)$ (just give an example of a FCC $\boldsymbol{y}$) than on $\alpha'(G)$. Excluding α' from the proofs of $\alpha \le \alpha' \le \gamma$ we obtain the following.

Lemma 30 *For any graph G we have $\alpha(G) \le \gamma(G)$.*

Proof. Let $U \subset V(G)$ be an independent set and $\boldsymbol{y}$ be an FCC. We have

$$\sum_{D \in \mathcal{C}(G)} y_D \ge \sum_{D \in \mathcal{C}(G)} |U \cap D| \, y_D = \sum_{u \in U} \sum_{\substack{D \in \mathcal{C}(G) \\ D \ni u}} y_D \ \ge |U|,$$

which clearly implies the claim. ▮

In particular, by Lemma 29 we deduce the following.

Corollary 31 (Shannon [Sha56]) *For any graph G, $\Theta(G) \leq \gamma(G)$.* ▮

Let us see what our findings say about the capacity of the m-cycle C_m, for example. Trivially, $\Theta(C_3) = 1$, so let $m \geq 4$. The fractional clique covering which is $1/2$ on the edges (which are cliques of size 2) and zero otherwise shows that $\Theta(C_m) \leq m/2$. As $\alpha(C_m) = \lfloor m/2 \rfloor$ we deduce the following.

Theorem 32 *For even $m \geq 4$ we have $\Theta(C_m) = m/2$.* ▮

However, difficulties arise already in the case $m = 5$, which we consider in detail in the next section.

Notes

The results we presented in this section are due to Shannon [Sha56].

10.4 Pentagon

Up to now we know that $\sqrt{5} \leq \Theta(C_5) \leq 5/2$. These bounds stood for more than 20 years until Lovász [Lov79] showed that $\Theta(C_5) = \sqrt{5}$. Lovász' main idea was to represent the vertices of a graph by real vectors $\boldsymbol{v}_i$ of norm 1 such that any two vectors which correspond to non-adjacent vertices in G are orthogonal. Let us call such a set of vectors an *orthonormal presentation* of G.

Clearly, at least one orthonormal representation exists: just take an orthonormal basis in dimension $|V|$. However, sometimes a more 'economical' representation can be found. Let us look at C_5. For this graph we can obtain a nice orthonormal representation in $\mathbb{R}^3$ by considering an 'umbrella' with five ribs $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_5$ of unit length. Now open the umbrella (with the tip at the origin) to the position where the angles between alternate ribs are 90°.

In an outline our approach is standard: construct a representation of G^n given one for G and derive an upper bound on $\alpha(G^n)$ from this.

The tensor product of vectors seems a good (and obvious) candidate for the first task. The realisation of this idea is just the matter of notation; the proof is straightforward and takes care of itself.

Let graphs G and H have orthonormal representations R and S in $\mathbb{R}^r$ and $\mathbb{R}^s$, respectively, To the vertex of $G \times H$ corresponding to the pair $(\boldsymbol{v}, \boldsymbol{w})$, $\boldsymbol{v} \in R$, $\boldsymbol{w} \in S$ we associate the vector

$$\boldsymbol{v}\boldsymbol{w}^T := (v_1w_1, \ldots, v_1w_s, v_2w_1, \ldots, v_2w_s, \ldots, v_rw_1, \ldots, v_rw_s) \in \mathbb{R}^{rs}.$$

The notation $\boldsymbol{v}\boldsymbol{w}^T$ means the matrix multiplication of the $r \times 1$-matrix $\boldsymbol{v}$ by the $1 \times s$-matrix $\boldsymbol{w}^T$. Let us denote the obtained system of vectors by $R \otimes S$.

Lemma 33 *$R \otimes S$ is an orthonormal representation for $G \times H$.*

Proof. It is readily checked that for $\boldsymbol{v}, \boldsymbol{x} \in \mathbb{R}^r$ and $\boldsymbol{w}, \boldsymbol{y} \in \mathbb{R}^s$ we have

$$\boldsymbol{v}\boldsymbol{w}^T \cdot \boldsymbol{x}\boldsymbol{y}^T = \sum_{i=1}^{r}\sum_{j=1}^{s} v_i w_j x_i y_j = (\boldsymbol{v} \cdot \boldsymbol{x}) \times (\boldsymbol{w} \cdot \boldsymbol{y}). \tag{38}$$

For example, let us verify that each vector $\boldsymbol{v}\boldsymbol{w}^T \in R \times S$ has norm 1:

$$\|\boldsymbol{v}\boldsymbol{w}^T\|^2 = \boldsymbol{v}\boldsymbol{w}^T \cdot \boldsymbol{v}\boldsymbol{w}^T = (\boldsymbol{v} \cdot \boldsymbol{v}) \times (\boldsymbol{w} \cdot \boldsymbol{w}) = 1. \blacksquare$$

The second part of our programme is to bound $\alpha(G)$ given a representation T of G. Assume that $V(G) = [m]$ and $T = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} \subset \mathbb{R}^t$.

Let U be an independent set in G. The most obvious observation is that $|U| \le t$: the vectors corresponding to U are orthogonal and hence linearly independent. This approach works but for C_5 it gives $\Theta(C_5) \le 3$ only. A less obvious observation is that the vector $\frac{1}{|U|}\sum_{i\in U}\boldsymbol{v}_i$ has norm $1/\sqrt{|U|}$ — the main point is that the norm is small if $|U|$ is large. Thus,

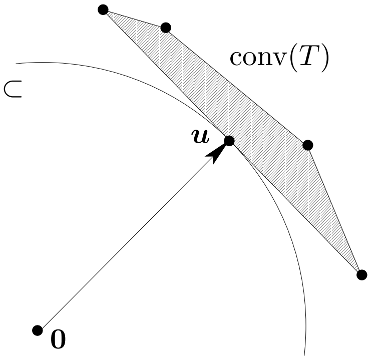

Fig. 3: The minimum norm vector.

$$\alpha(G) \le \frac{1}{\mu_T}, \tag{39}$$

where we define μ_T to be the infimum (in fact, minimum) of

$$\mu(\boldsymbol{x}) := \|x_1\boldsymbol{v}_1 + \ldots + x_m\boldsymbol{v}_m\|^2$$

over all *probability distributions* $\boldsymbol{x}$ on $[m]$ (that is, the reals x's are non-negative and their total sum is precisely 1). In other words,

$$\mu_T = \min\{\|\boldsymbol{u}\|^2 : \boldsymbol{u} \in \text{conv}(T)\},$$

where conv(T) denotes the *convex hull* of T.

Now we try to prove that $\Theta(G) \le 1/\mu_T$ by showing that $\mu_{R\otimes S} \ge \mu_R\mu_S$ for any representations R and S. At first, it is not quite clear how to prove this inequality although the opposite one $\mu_{R\otimes S} \le \mu_R\mu_S$ is straightforward to show: pick vectors $\boldsymbol{v} \in \text{conv}(R)$, $\boldsymbol{w} \in \text{conv}(S)$ of the minimum norm; then $\boldsymbol{v}\boldsymbol{w}^T \in \text{conv}(R \otimes S)$ has norm $\|\boldsymbol{v}\| \times \|\boldsymbol{w}\| = \mu_R\mu_S$ by (38).

The reason for this is that an example of a vector $\boldsymbol{u} \in \text{conv}(T)$ shows only $\mu_T \le \|\boldsymbol{u}\|^2$. What other information do we need to conclude that in fact $\mu_T = \|\boldsymbol{u}\|^2$? A rough picture (Figure 3) suggests that no $\boldsymbol{v} \in T$ lies in the same open half-space as $\boldsymbol{0}$ with respect to the affine hyperplane that contains $\boldsymbol{u}$ and is orthogonal to $\boldsymbol{u}$. In the linear algebra language,

$$\boldsymbol{v} \cdot \boldsymbol{u} \ge \boldsymbol{u} \cdot \boldsymbol{u} \text{ for each } \boldsymbol{v} \in T. \tag{40}$$

Suppose that (40) does not hold for $\boldsymbol{v} \in T$. Figure 3 suggest that moving from $\boldsymbol{u}$ to $\boldsymbol{v}$ along the line $(1-t)\boldsymbol{u} + t\boldsymbol{v}$ we hit the interior of the sphere through $\boldsymbol{u}$. The formal proof is as easy: the vector $\boldsymbol{w} := (1-t)\boldsymbol{u} + t\boldsymbol{v}$ lies in $\operatorname{conv}(T)$ and contradicts the minimality of $\boldsymbol{u}$ for all small $t > 0$:

$$\|\boldsymbol{w}\|^2 = (1-t)^2\|\boldsymbol{u}\|^2 + 2t\boldsymbol{v}\cdot\boldsymbol{u} + t^2\|\boldsymbol{v}\|^2 = (1-2t)\|\boldsymbol{u}\|^2 + 2t\boldsymbol{v}\cdot\boldsymbol{u} + O(t^2) < \|\boldsymbol{u}\|^2.$$

Let us show that the converse is also true. Looking at Figure 3 again we see that any $\mu(\boldsymbol{x}) \in \operatorname{conv}(T)$ lies in the correct half-space and hence has norm at least $\|\boldsymbol{u}\|$, as required. Expressing this argument analytically we obtain

$$\mu(\boldsymbol{x})\cdot\boldsymbol{u} = \sum_{i=1}^{m} x_i\,(\boldsymbol{v}_i\cdot\boldsymbol{u}) \ge \sum_{i=1}^{m} x_i\,(\boldsymbol{u}\cdot\boldsymbol{u}) = \|\boldsymbol{u}\|^2;$$

and hence

$$\|\mu(\boldsymbol{x})\| \ge \frac{\mu(\boldsymbol{x})\cdot\boldsymbol{u}}{\|\boldsymbol{u}\|} \ge \frac{\boldsymbol{u}\cdot\boldsymbol{u}}{\|\boldsymbol{u}\|} = \|\boldsymbol{u}\|.$$

Here we applied the *Cauchy–Schwarz inequality* to $\mu(\boldsymbol{x})$ and $\boldsymbol{u}$; let us give its proof for the sake of completeness.

Lemma 34 (Cauchy–Schwarz Inequality) *For any $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ we have*

$$|\boldsymbol{a}\cdot\boldsymbol{b}| \le \|\boldsymbol{a}\|\,\|\boldsymbol{b}\|. \tag{41}$$

We have equality in (41) *if and only if $\boldsymbol{a}$ and $\boldsymbol{b}$ are collinear.*

Proof. Consider the vector $\boldsymbol{c} := \boldsymbol{a} + \lambda\boldsymbol{b}$, where are free to choose $\lambda \in \mathbb{R}$. We have

$$0 \le \boldsymbol{c}\cdot\boldsymbol{c} = \|\boldsymbol{a}\|^2 + 2\lambda\,\boldsymbol{a}\cdot\boldsymbol{b} + \lambda^2\|\boldsymbol{b}\|^2.$$

This (quadratic in λ) inequality is valid for any λ, so let us choose the 'worst' λ. Namely we let $\lambda = -\boldsymbol{a}\cdot\boldsymbol{b}/\|\boldsymbol{b}\|^2$, minimising the right-hand side which now equals $\|\boldsymbol{a}\|^2 - (\boldsymbol{a}\cdot\boldsymbol{b})^2/\|\boldsymbol{b}\|^2$. This should be non-negative, implying the lemma. ▮

It is easy to see that the $\otimes$-products of representations preserves (40):

Lemma 35 *Let $R = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ and $S = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$. Let $\mu_R = \|\boldsymbol{y}\|^2$ and $\mu_S = \|\boldsymbol{z}\|^2$ with $\boldsymbol{y} = \sum_{i=1}^m \alpha_i\boldsymbol{v}_i \in \operatorname{conv}(R)$ and $\boldsymbol{z} = \sum_{j=1}^n \beta_j\boldsymbol{w}_j \in \operatorname{conv}(S)$. Then $\boldsymbol{y}\boldsymbol{z}^T$ satisfies* (40)*. In particular,*

$$\mu_{R\otimes S} = \|\boldsymbol{y}\boldsymbol{z}^T\|^2 = \|\boldsymbol{y}\|^2\,\|\boldsymbol{z}\|^2 = \mu_R\mu_S.$$

Proof. We already know that $\mu_{R\otimes S} \le \mu_R\mu_S$. On the other hand,

$$\boldsymbol{y}\boldsymbol{z}^T = \sum_{i=1}^{m}\sum_{j=1}^{n}(\alpha_i\beta_j)\boldsymbol{v}_i\boldsymbol{w}_j^T \in \operatorname{conv}(R\otimes S).$$

(Note that $\sum_{i=1}^m\sum_{j=1}^n \alpha_i\beta_j = (\sum_{i=1}^m \alpha_i)(\sum_{j=1}^n \beta_j) = 1$.) Also, for any $\boldsymbol{v}_i\boldsymbol{w}_j \in R\otimes S$ we have by (38)

$$\boldsymbol{v}_i\boldsymbol{w}_j^T\cdot\boldsymbol{y}\boldsymbol{z}^T = (\boldsymbol{v}_i\cdot\boldsymbol{y})\times(\boldsymbol{w}_j\cdot\boldsymbol{z}) \ge (\boldsymbol{y}\cdot\boldsymbol{y})\times(\boldsymbol{z}\cdot\boldsymbol{z}) = \boldsymbol{y}\boldsymbol{z}^T\cdot\boldsymbol{y}\boldsymbol{z}^T.$$

The lemma is proved. ▮

Corollary 36 (Lovasz [Lov79]) *If T is an orthonormal representation for a graph G, then $\Theta(G) \leq 1/\mu_T$.* ▮

Now let us look at our 'umbrella' representation of C_5. Here the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_5$ lie on a plane, forming a regular pentagon. Clearly, $\boldsymbol{u} := \frac{1}{5}(\boldsymbol{v}_1 + \cdots + \boldsymbol{v}_5)$ is the minimum-norm vector. It is not hard to calculate that $\|\boldsymbol{u}\|^2 = 1/\sqrt{5}$. (A detailed proof of this can be found in [AZ98, page 177].)

Theorem 37 (Lovasz [Lov79]) $\Theta(C_5) = \sqrt{5}$. ▮

10.5 Sperner Capacity

Sperner capacity is an extension of Shannon capacity to directed confusion graphs and is therefore more difficult to compute. Rather then giving general definition let us consider one concrete example.

Suppose that the alphabet is $\mathbb{F}_3 = \{0, 1, 2\}$ and when we transmit a letter $i \in \mathbb{F}_3$ through our channel, then the output is either i (i.e. no error) or $i - 1$. Thus sending 0 we may receive 2 at the output but not vice versa. This can be represented by the *directed confusion graph* which is, in this particular case, a *directed triangle* consisting of the arcs $(0, 2)$, $(1, 0)$ and $(2, 1)$.

Let $t(n)$ be the maximal number of length-n messages such that any transmission error can be detected. It is easy to see that $t(n)$ is the maximum size of $X \subset (\mathbb{F}_3)^n$ such that for every distinct $\boldsymbol{x}, \boldsymbol{y} \in X$ we have $x_i - y_i = 1$ for some $i \in [n]$. Thus the information rate is $\frac{1}{n} \log_2 t(n) = \log_2 \sqrt[n]{t(n)}$ and the so called *Sperner capacity* of the directed triangle is $\sup_{n \geq 1} \sqrt[n]{t(n)}$.

Theorem 38 *The Sperner capacity of the directed triangle is* 2.

Proof. The lower bound is easy: take for X the subset of $\{0, 1\}^n$ consisting of the sequences having precisely $\lfloor n/2 \rfloor$ ones, which gives $\binom{n}{\lfloor n/2 \rfloor} = 2^{1-o(1)}$ sequences by Stirling's formula.

On the other hand, let $X \subset (\mathbb{F}_3)^n$ be as above. For each $\boldsymbol{u} \in X$ consider the multilinear polynomial

$$F_{\boldsymbol{u}}(x_1, \ldots, x_n) = \prod_{i=1}^{n} (x_i - u_i - 1).$$

We have $F_{\boldsymbol{u}}(\boldsymbol{u}) = (-1)^n \neq 0$, for every $\boldsymbol{u} \in X$, but if $\boldsymbol{u}, \boldsymbol{v}$ are different elements of X, then $F_{\boldsymbol{u}}(\boldsymbol{v}) = 0$. So the polynomials $F_{\boldsymbol{u}}$, $\boldsymbol{u} \in X$, form an independent set in the vector space V of multilinear polynomials in n variables. It follows that $|X| \leq \dim(\mathsf{V}) = 2^n$. ▮

Notes

Gargano, Körner and Vaccaro [GKV92] were apparently first to define the Sperner capacity. This capacity is related to so-called Sperner subsets, whose theory was largely developed by (as you have guessed) Sperner.

Theorem 38 was independently discovered by Calderbank, Frankl, Graham, Li and Shepp [CFG$^+$93] and Blokhuis [Blo93]; the polynomial trick comes from the latter paper.

11 Combinatorial Nullstellensatz

Combinatorial Nullstellensatz of Alon [Alo99] gives certain conditions on a multivariate polynomial f and a set S which ensure that $f(\boldsymbol{s}) \neq 0$ for some $\boldsymbol{s} \in S$. Historically, the statement of the theorem came as an attempt to look at known polynomial proofs of combinatorial results and extract their common essence. Alon came up with a genuinely useful definition: a spectacular array of combinatorial results (old and new) can be proved with his Nullstellensatz by devising a polynomial f so that the existence of $\boldsymbol{s}$ with $f(\boldsymbol{s}) \neq 0$ can be interpreted combinatorially.

11.1 Cauchy–Davenport Theorem

Let us start with the following problem from combinatorial number theory. Let $A, B \subset \mathbb{Z}_n$. Find $r_n(a,b)$, the smallest possible size of

$$A + B := \{x + y : x \in A,\ y \in B\}$$

given $a = |A|$ and $b = |B|$. The obvious example of $A = [0, a-1]$ and $B = [0, b-1]$ shows that

$$r_n(a,b) \le \min(n, a+b-1). \tag{42}$$

Another easy observation is that if $a + b > n$, then for any $x \in \mathbb{Z}_n$ the sets A and $x - B := \{x - y : y \in B\}$ of cardinalities a and b must intersect, that is, $A + B = \mathbb{Z}_n$ and we conclude that

$$r_n(a,b) = n, \quad \text{if } a + b > n. \tag{43}$$

Playing further with the problem we discover that for a non-prime n, there are in some cases better constructions. For example, if $n = 2k$ is even, then letting A, B consist of all even residues modulo n we infer than $r(2k, k, k) \le k$. Things get too complicated, so let us restrict our consideration to a prime $n = p$ only, when it seems, after a few experiments, that (42) is sharp.

Suppose that $a + b \le p$. How could we prove a lower bound on $|C|$, where $C = A + B$? A possible approach is to consider the polynomial

$$f(x,y) = \prod_{i \in C} (x + y - i) \tag{44}$$

over $\mathbb{F}_p$. Note that the degree of f is precisely $|C|$ (e.g. the coefficient at $x^{|C|}$ is 1). Also, and this is the main point of the definiton, $f(x,y)$ is zero for any $(x,y) \in A \times B$. Can we deduce from this information a lower bound on $\deg(f)$?

The result that comes immediately to the mind is the well-known fact that the number of roots (taken with multiplicities) of a *univariate polynomial* (i.e. depending on a single variable) is at most its degree. It is not obvious how to interpret multiple roots combinatorially so let us state a slightly weaker but more combinatorial version.

Lemma 39 *Let $\mathbb{F}$ be a field (finite or infinite). For any non-zero $f \in \mathbb{F}[x]$ the number of distinct roots is at most* $\deg(f)$.

Proof. Let $f(a) = 0$. Divide $f(x)$ by $x-a$ using the Euclidean Algorithm and conclude that $f(x) = (x-a)g(x)$, where $\deg(g) = \deg(f) - 1$. Apply induction to g. ∎

It is not obvious how to extend this result to multivariate polynomials: for example, the degree-1 polynomial $g(x, y) = x-y$ has $|\mathbb{F}|$ roots $\{(a, a) : a \in \mathbb{F}\}$. One of the reasons is that it is not clear how a multivariate division algorithm should be defined. The most plausible way is to view all variables but one as fixed and apply the Euclidean Algorithm with respect to the selected variable.

But what do we divide and by what? The divident should apparently be our polynomial $f(x, y)$ which is zero on $A \times B$. To choose a divisor g, we should have a clearer idea what we are to achieve by the division. We will obtain a representation $f = gh + f_1$; it is not clear how we can ensure that $f_1 = 0$ so maybe we should just try to have f_1 'simpler' than f. On the other hand, we would prefer f_1 to retain at least some properties of f. If f_1 is to be zero on $A \times B$, this can be most easily ensured by choosing g which is zero on $A \times B$. Given this, there are not many options for g and the choice $g(y) = \prod_{v\in B}(y - v)$, for example, is an obvious one.

How do we divide f by g? The highest-degree term in $g(y)$ is y^b: $g(y) = y^b + \sum_{i=0}^{b-1} g_i y^i$. If we expand f by powers of y, that is, write

$$f(x, y) = \sum_{i=0}^{t} p_i(x)y^i,$$

then (if $t \geq b$) the y-degree of $f(x, y) - p_t(x)g(y)y^{t-b}$ is strictly smaller than t. So we can iterate the procedure until the y-degree drops below t, obtaining a representation

$$f(x, y) = g(y)h(x, y) + f_1(x, y), \tag{45}$$

with $\deg_y(f_1) < b$ and, as it is easy to see, $\deg(f_1) \leq \deg(f)$. It is going well: we have reduced the y-degree to at most $b - 1$ without increasing the total degree. As f_1 is zero on $A \times B$, we can conclude by Lemma 39 that for any fixed $u \in A$ the univariate polynomial $f_1(u, y) \in \mathbb{F}[y]$ is identically zero. This sounds good but this does not imply yet that $f_1(x, y) \in \mathbb{F}[x, y]$ is identically zero: we may have, for example, $f_1(x, y) = k(x)$, where

$$k(x) := \prod_{u\in A}(x - u) = x^a + \sum_{i=0}^{a-1} k_i x^i.$$

So, why don't we apply the same trick again, this time with respect to x? It is probably better to leave gh as it is but divide f_1 by k, obtaining

$$f_1(x, y) = k(x)l(x, y) + f_2(x, y)$$

with $\deg_x(f_2) < a$ and $\deg(f_2) \leq \deg(f_1)$. Now, let us see, very carefully (hold the breath!) if we have $\deg_y(f_2) < b$. At each step we replace some term $x^t p_t(y)$

by $x^{t-a}p_t(y)k(x) - x^{t-a}p_t(y)\sum_{i=0}^{a-1} k_i x^i$, which clearly does not increase the y-degree; therefore, we indeed have $\deg_y(f_2) \le \deg_y(f_1) < b$.

Now the identity $f_2 = 0$ is implied by the following easy lemma which we can, at no extra expense, state in the general multivariate case (and for any field $\mathbb{F}$).

Lemma 40 *Let $f \in \mathbb{F}[x_1, \dots, x_n]$, $t_i = \deg_{x_i}(f)$ and $S_i \in \binom{\mathbb{F}}{t_i+1}$, $i \in [n]$. If $f(\boldsymbol{s}) = 0$ for any $\boldsymbol{s} \in S_1 \times \cdots \times S_n$, then $f = 0$.*

Proof. We apply induction on n with Lemma 39 implying the case $n = 1$. Let $n \ge 2$. Write f as

$$f(\boldsymbol{x}) = \sum_{i=0}^{t_n} p_i(x_1, \dots, x_{n-1})\, x_n^i.$$

Fix an arbitrary $(n-1)$-tuple $\boldsymbol{s} \in S_1 \times \cdots \times S_{n-1}$. The polynomial $g(x_n) = f(\boldsymbol{s}, x_n) \in \mathbb{F}[x_n]$ vanishes on the set S_n. As this set has more than $t_n = \deg(g)$ elements, we conclude by Lemma 39 that g is identically 0.

This means that $p_i(\boldsymbol{s}) = 0$ for all $\boldsymbol{s} \in S_1 \times \cdots \times S_{n-1}$. By the induction hypothesis $p_i = 0$ for all i, implying that $f = 0$. ▮

Thus, we obtain that if $f(x, y)$ is zero on $A \times B$, then

$$f(x, y) = g(y)h(x, y) + k(x)l(x, y). \tag{46}$$

What can we say about h and l? Analysing the division algorithm we can only conclude that $\deg(h) \le \deg(f) - b$, $\deg_x(h) \le \deg_x(f)$, $\deg_y(h) \le \deg_y(f) - b$, and some similar inequalities for l. The identity (46) looks interesting but it does not seem to imply any bound on $\deg(f)$. Have we made all this path in vain? No! Looking more carefully at the right-hand side of (46) we observe that it is the sum of $x^a l(x, y)$, $y^b h(x, y)$, and terms of degree strictly smaller than $\deg(f)$. Hence, we can conclude that every non-zero monomial of f of degree $\deg(f)$ contains either x^a or y^b as a factor.

This does not sound very impressive but let us check what our findings imply about the original problem of estimating $r_p(a, b)$. Can we derive a contradiction by assuming that $a + b \le p$ and that $c \le a + b - 2$, where $c = |C| = |A + B|$? Then the highest-degree terms of f are $\binom{c}{d} x^d y^{c-d}$, $d \in [0, c]$: observe that $\binom{c}{d}$ is non-zero (modulo p) in the view of $c < p$. But then the monomial $x^{a-1}y^{c-a+1}$ contains neither x^a nor y^b, a contradiction!

Putting all together, we have computed $r_p(a, b)$:

Theorem 41 (Cauchy–Davenport Theorem [Dav35]) *For any prime p we have $r_p(a, b) = \min(p, a + b - 1)$.* ▮

Now that our algebraic result seems useful let us try to state (with a sketchy proof) its general version. Alon [Alo99] calls it *Combinatorial Nullstellensatz* as its assumptions and conclusions bear some resemblance to the Hilbert Nullstellensatz.

Theorem 42 (Combinatorial Nullstellensatz, Alon [Alo99]) *Let $\mathbb{F}$ be an arbitrary field, and let $f \in \mathbb{F}[x_1, \dots, x_n]$. Suppose that for some non-negative integers $t_1, \dots, t_n$ we have $\deg(f) = \sum_{i=1}^n t_i$ and the coefficient at $\prod_{i=1}^n x_i^{t_i}$ in f is non-zero. Then, if $S_1, \dots, S_n$ are subsets of $\mathbb{F}$ with $|S_i| > t_i$, there is $(s_1, \dots, s_n) \in S_1 \times \cdots \times S_n$ such that*

$$f(s_1, \dots, s_n) \neq 0.$$

Proof. Clearly we may assume that $|S_i| = t_i + 1$ for all $i \in [n]$. Suppose that the result is false, and define $g_i(x_i) = \prod_{s \in S_i}(x_i - s)$.

Let $f_0 = f$. For $i = 1, 2, \dots, n$ divide $f_{i-1}(\boldsymbol{x})$ by $g_i(x_i)$ to obtain

$$f_{i-1}(\boldsymbol{x}) = g_i(x_i)h_i(\boldsymbol{x}) + f_i(\boldsymbol{x}).$$

with $\deg(h_i) \leq \deg(f_{i-1}) - t_i - 1$, $\deg(f_i) \leq \deg(f_{i-1})$ and $\deg_{x_i}(f_i) \leq t_i$. Show by induction on i that $\deg(h_i) \leq \deg(f) - t_i - 1$, $\deg_{x_j}(f_i) \leq t_j$ for any $j \in [i]$ and that f_i is zero on any $\boldsymbol{s} \in S_1 \times \cdots \times S_n$. (*Exercise!*)

By Lemma 40 we conclude that $f_n = 0$. Hence, we have obtained a representation

$$f(\boldsymbol{x}) = \sum_{i=1}^n g_i(x_i)h_i(\boldsymbol{x}),$$

with $\deg(h_i) \leq \deg(f) - t_i - 1$. Thus

$$f(\boldsymbol{x}) = \sum_{i=1}^n x_i^{t_i+1}h_i(\boldsymbol{x}) + (\text{terms of degree} < \deg(f)), \tag{47}$$

By assumption, the coefficient at $\prod_{i=1}^n x_i^{t_i}$ in the left-hand side of (47) is non-zero, while it is impossible to have such a monomial in the right-hand side, a contradiction. ▮

Notes

Apparently, Theorem 41 was proved by Cauchy in 1813 and rediscovered by Davenport [Dav35]; their proofs use some induction on $|B|$. Further related results, obtained essentially via Combinatorial Nullstellensatz were proved by Alon, Nathanson and Rusza [ANR95, ANR96].

11.2 Regular Subgraphs

A graph is called *k-regular* if its every vertex has degree k. There are many problems related to the existence of regular subgraphs in a graph. For example, Erdős and Sauer asked for the value of $\mathrm{ex}(n, k\text{-reg})$, the largest number of edges in a graph of order n without a k-regular subgraph. It is trivial to see that $\mathrm{ex}(n, 1\text{-reg}) = 0$ and $\mathrm{ex}(n, 2\text{-reg}) = n - 1$. The first non-trivial case is $k = 3$. The difficulty is that there are very few known conditions on a graph G ensuring a k-regular subgraph $H \subset G$ for $k \geq 3$.

Let us investigate if Combinatorial Nullstellensatz can give any such conditions on G. To define $H \subset G$, up to isolated vertices, we have to specify a subset of $E(G)$, which can be done by specifying numbers $x_D \in \{0,1\}$, $D \in E(G)$. Can we devise a polynomial f depending on $e(G)$ variables so that, for a $(0,1)$-vector $\boldsymbol{x}$, the inequality $f(\boldsymbol{x}) \neq 0$ implies that the subgraph $H \subset G$ determined by $\boldsymbol{x}$ is k-regular?

The H-degree of a vertex v can be computed by the linear function $L_v(\boldsymbol{x}) = \sum_{u\in\Gamma(v)} x_{uv}$. (For simplicity of notation, we write x_{uv} instead of $x_{\{u,v\}}$, etc.) Unfortunately, the polynomial (over the reals)

$$f(\boldsymbol{x}) := \sum_{v\in V(G)} L_v(\boldsymbol{x})^2 (L_v(\boldsymbol{x}) - k)^2 \tag{48}$$

does not work: f is zero iff each degree is 0 or k while we need it rather otherwise.

Well, maybe we can find some polynomial h so that $h(f(\boldsymbol{x}))$ is zero whenever f is non-zero. A non-zero polynomial $h(x)$ with $h(x) = 0$ for any $x \neq 0$ does not exists... unless we live inside a finite field. For example, for a prime p we can define $h(x) = \prod_{i=1}^{p-1}(x-i) \in \mathbb{F}_p$.

But then the definition (48) loses its properties when considered over a finite field. (There are lots of non-zero $\boldsymbol{a}$'s with $\sum a_i^2 = 0$.) Our next attempt is to consider

$$f(\boldsymbol{x}) := \prod_{v\in V(G)} \prod_{i\in\mathbb{F}_p\setminus\{0,k\}} (L_v(\boldsymbol{x}) - i) \in \mathbb{F}_p, \tag{49}$$

as $f(\boldsymbol{x}) \neq 0$ implies that each $L_v(\boldsymbol{x})$ is 0 or k. But this does not imply that H is k-regular: the equality $L_v(\boldsymbol{x}) = 0$ could mean, for example, that $d_H(v) = p$. We rectify this drawback by assuming that no vertex of G has degree p or larger (assuming also that $k < p$). But then we cannot apply Combinatorial Nullstellensatz for sets $S_{uv} = \{0,1\}$, $uv \in E(G)$: one of the necessary conditions says that

$$\deg(f) = (p-2)v(G) \leq \sum_{uv\in E(G)} (|S_{uv}| - 1) = e(G),$$

which is not compatible with the inequality $\Delta(G) \leq p-1$ (unless $p \leq 3$; but this is not interesting as $k \leq p-1$).

This seems a dead end... but why don't we take $k = p$? Then $d_H(v) = p$ is acceptable while we can prevent other bad degrees by assuming that $\Delta(G) < 2p$. The new definition is

$$f(\boldsymbol{x}) := \prod_{v\in V(G)} h(L_v(\boldsymbol{x})), \tag{50}$$

where (as before) $h(x) = \prod_{i=1}^{p-1}(x-i) \in \mathbb{F}_p$.

It is not crucial, but more aesthetically pleasing, if we observe that, over $\mathbb{F}_p$, $h(x) = x^{p-1} - 1$. Indeed, the polynomial $h_0(x) := h(x) - x^{p-1} + 1$ has degree at most $p-2$ while by Fermat's Little Theorem (Theorem 43) each $a \in \mathbb{F}_p \setminus \{0\}$ is a root of h_0; now Lemma 39 implies that $h_0 = 0$.

Theorem 43 (Fermat's Little Theorem) *If p is a prime and $a \in \mathbb{F}_p$, then $a^p = a$. In particular, $a^{p-1} = 1$ for any $a \in \mathbb{F}_p \setminus \{0\}$.*

Proof. Note that $(a+1)^p = \sum_{i=0}^{p} a^i \binom{p}{i} = a^p + 1$ as p divides $\binom{p}{i}$ for $i \in [1, p-1]$. Apply induction on a. ▮

The definition (50) works fine ($f(\boldsymbol{x}) \neq 0 \;\Rightarrow\; \forall v\, L_v(\boldsymbol{x}) = 0$) except the all-zero vector $\mathbf{0}$ causes us a problem by corresponding to the empty subgraph. (Of course, when we talk about 'regular subgraphs' we mean the non-empty ones.)

The final (promise!) modification comes from combining f with the polynomial $\prod_{uv\in E(G)}(1 - x_{uv})$ which is zero unless $\boldsymbol{x} = \mathbf{0}$.

Theorem 44 (Alon, Friedland and Kalai [AFK84]) *Let p be a prime and G be a graph of* average degree $d(G) := 2e(G)/v(G)$ *larger than* $2p-2$ *and maximum degree at most* $2p-1$. *Then G contains a p-regular subgraph.*

Proof. Consider the polynomial

$$f(\boldsymbol{x}) = \prod_{v\in V(G)} \left(1 - \Big(\sum_{u\in\Gamma(v)} x_{uv} \Big)^{p-1} \right) - \prod_{uv\in E(G)} (1 - x_{uv}),$$

in variables x_{uv}, $uv \in E(G)$, over $\mathbb{F}_p$. Notice that the coefficient at $\prod_{uv\in E(G)} x_{uv}$ in f is $(-1)^{e(G)+1} \neq 0$. Thus, $\deg(f) = e(G)$, since the degree of the first product is at most $(p-1)v(G) < e(G)$, by the assumption on the average degree of G.

Therefore, we can apply Combinatorial Nullstellensatz (Theorem 42 for $S_{uv} = \{0,1\}$ and $t_{uv} = 1$) which gives us a $(0,1)$-vector $\boldsymbol{x}$ with $f(\boldsymbol{x}) \neq 0$. Let

$$E(H) = \{uv \in E(G) : x_{uv} = 1\}$$

and let $V(H)$ consist of the vertices spanned by $E(H)$. As $f(\mathbf{0}) = 0$, we conclude that $\boldsymbol{x} \neq \mathbf{0}$ and H is non-empty. The second summand in $f(\boldsymbol{x})$ is therefore zero and it follows from Fermat's Little Theorem (Theorem 43) that $\sum_{u\in\Gamma(v)} x_{uv}$ is zero (modulo p) for every v. As $\Delta(G) < 2p$, we conclude that H is p-regular, as required. ▮

Let us return to ex(n, k-reg). The following lemmas are not difficult and come as an attempt to find iteratively a subgraph which is closer and closer to being regular.

As we do not even know the order of magnitude of ex(n, k-reg), we do not worry about multiplicative constants. In situations like this, it is always a good idea to restrict our consideration to bipartite graphs (by losing at most half of the edges):

Lemma 45 *Every graph G contains a bipartite subgraph H with $e(H) \geq \frac{1}{2}\, e(G)$.*

Proof. Let $G(A,B) = \{\{x,y\} \in E(G) : x \in A, y \in B\}$. If we want a partition $V(G) = A \cup B$ with $|G(A,B)|$ being large, let us take a partition maximising this quantity. It is easy to see that no $x \in A$, for example, can have more neighbours in A than in B: otherwise we can increase $|G(A,B)|$ by moving x to B. Hence, at least half of all edges go across. ▮

Our next observation is that we can make the minimal degree big by consecutively removing vertices of small degree. To avoid the danger of ending up with the empty graph, let us remove a vertex $x \in V(G)$ if this does not decrease the average degree $d(G)$. Thus, we should have

$$d(G-x) = \frac{2(e(G)-d(x))}{v(G)-1} \geq d(G) = \frac{2e(G)}{v(G)},$$

which is equivalent to $d(x) \geq d(G)/2$. So, when we get stuck, each vertex has degree larger than $d(G)/2$. Let $\sigma = \delta(G) > d(G)/2$ be the current minimum degree. By deleting edges we can achieve that all vertices in one part of our bipartite graph G have degree σ. We should probably do this for the bigger part because then the average degree on the other (smaller) half is guaranteed to be at least σ. Let us call a bipartite graph H *σ-half-regular* if for some bipartition $V(H) = A \cup B$ we have $d(x) = \sigma$ for all $x \in A$ and $|A| \geq |B|$. We have proved the following claim.

Lemma 46 *Every bipartite graph G contains a σ-half-regular subgraph H with $\sigma > \frac{1}{2}d(G)$.* ▮

A *k-factor* of G is a *spanning* (i.e. $V(H) = V(G)$) k-regular subgraph $H \subset G$. In particular, a 1-factor is a set of disjoint edges covering all vertices.

Lemma 47 *Every σ-half-regular graph G contains a σ-half-regular subgraph H with a 1-factor.*

Proof. For the colour classes $V(G) = A \cup B$ we have $|A| \geq |B|$ by the definition. Let X be a minimal non-empty subset of A with $|\Gamma_G(X)| \leq |X|$, where

$$\Gamma_G(X) := \{v \in V(G) : \exists x \in X \ \{x, v\} \in E(G)\}.$$

As $|A| \geq |\Gamma_G(A)|$, such X exists. We have in fact $|\Gamma_G(X)| = |X|$ (otherwise $|X| \geq 2$ and the removal of any vertex from X contradicts the minimality of X). Again, by the minimality of X, we have $|\Gamma_G(Y)| \geq |Y|$ for any $Y \subset X$. By Hall's Marriage Theorem the graph H spanned by $X \cup \Gamma_G(X)$ has a 1-factor. ▮

Armed with the above results we can now attack $\mathrm{ex}(n, k\text{-reg})$. Let G be a graph of order n. We aim at showing that G contains a k-regular subgraph provided $e(G)$ is sufficiently large. By Lemmas 45, 46 and 47 we can find a bipartite, σ-half regular subgraph $G_0 \subset G$ with a 1-factor F_0, where $\sigma \geq \frac{1}{4}d(G)$.

Now we repeat the following for $i = 1, \ldots, \sigma - 1$. Given a $(\sigma - i + 1)$-half-regular graph G_{i-1} with a 1-factor F_{i-1}, remove $E(F_{i-1})$ from $E(G_{i-1})$ and apply Lemma 47 to obtain a $(\sigma - i)$-half-regular graph $G_i \subset G_{i-1}$ with a 1-factor F_i.

What we have obtained is a sequence of edge-disjoint matchings $F_0, \ldots, F_{\sigma-1}$ with nested vertex sets:

$$V(F_0) \supset V(F_1) \supset \cdots \supset V(F_{\sigma-1}) \neq \emptyset. \tag{51}$$

If σ is large, then we can find many matchings of nearly equal size. As their union F is 'nearly regular', we could try to apply Theorem 44. The fact that it gives regular graphs of prime valency only is not an obstacle by the following lemma.

Lemma 48 *Any bipartite l-regular graph G has a k-factor for every $k \le l$.*

Proof. It is enough to show that G has a 1-factor as we can remove these edges and repeat the argument for the obtained $(l-1)$-regular graph.

Let $V(G) = A \cup B$. We have

$$l\,|X| = |G(X, \Gamma(X))| \le l\,|\Gamma(X)|, \quad X \subset A \text{ or } X \subset B,$$

which implies that $|A| = |B|$ and, by Hall's theorem, that there is a perfect matching. ∎

So, we fix some prime $p \ge k$ and try to satisfy the assumptions of Theorem 44. To ensure that $\Delta(F) \le 2p-1$, we just take some $2p-1$ machings (but not more). By (51) we have $V(F) = V(F_j)$ for some $j \in [0, \sigma-1]$. Given this, the size of F is maximised if we take $2p-1$ consecutive matchings: $E(F) = \cup_{i=j}^{j+2p-2} E(F_i)$. Then we have the following bound:

$$d(F) = \frac{2e(F)}{v(F)} \ge \frac{2(2p-1) \times \frac{1}{2} v(F_{j+2p-2})}{v(F_j)}.$$

(Note that $e(F_i) = \frac{1}{2} v(F_i)$.) Thus we are done if $d(F) > 2p-2$, which is the case if we can find $j \in [0, \sigma - 2p + 1]$ with $v(F_{j+2p-2}) > \frac{2p-2}{2p-1} v(F_j)$.

If such j does not exist, it means in particular that $v(F_{i(2p-2)}) \le v(F_0) \left(\frac{2p-2}{2p-1}\right)^i$. Hence,

$$2 \le v(F_{\sigma-1}) \le v(F_0) \left(\frac{2p-2}{2p-1}\right)^{\lfloor \frac{\sigma-1}{2p-2} \rfloor} \le n \left(\frac{2p-2}{2p-1}\right)^{\lfloor \frac{\sigma-1}{2p-2} \rfloor}.$$

We obtain a contradiction if $\sigma \ge c_p \log n$, where c_p is a sufficiently large constant (depending on p only). Recalling that $\sigma \ge \frac{1}{4} d(G) = 2e(G)/v(G)$, we derive the following upper bound on $\mathrm{ex}(n, k\text{-reg})$.

Theorem 49 (Pyber [Pyb85]) $\mathrm{ex}(n, k\text{-reg}) = O(n \log n)$ *for any fixed k.* ∎

Notes

The problem of Erdős and Sauer to compute $\mathrm{ex}(n, k\text{-reg})$ was not published by these authors but is mentioned in e.g. [Bol78, page 399] or [Erd81].

Apparently, Lemma 45 appears first in Erdős [Erd67].

Pyber, Rödl and Szemerédi [PRS95], proved via probabilistic arguments, that there are graphs on n vertices with at least $\Omega(n \log\log n)$ edges that contains no 3-regular subgraph. Thus ths estimate in Theorem 49 is not far from being best possible.

11.3 Covering Cube by Affine Hyperplanes

Here is one result whose proof via Combinatorial Nullstellensatz is very short.

Theorem 50 (Alon and Füredi [AF93]) *Let $\mathsf{H}_1, \mathsf{H}_1, \dots, \mathsf{H}_m$ be a family of affine hyperplanes in $\mathbb{R}^n$ that cover all vertices of the unit cube $\{0,1\}^n$ except $\mathbf{0}$ (which is uncovered). Then $m \ge n$.*

Proof. As $\mathsf{H}_i \not\ni \mathbf{0}$, we have $\mathsf{H}_i = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}_i \cdot \boldsymbol{x} = 1\}$ for some $\boldsymbol{a}_i \in \mathbb{R}^n$, $i \in [m]$.

Assume that the assertion is false, that is, $m < n$, and consider the polynomial

$$f(\boldsymbol{x}) = \prod_{i=1}^{m}(1 - \boldsymbol{a}_i \cdot \boldsymbol{x}) - \prod_{i=1}^{n}(1 - x_i).$$

The degree of this polynomial is clearly n and the coefficient at $x_1 \dots x_n$ is $(-1)^{n+1} \neq 0$. Therefore, by Combinatorial Nullstellensatz (Theorem 42 for $S_i = \{0,1\}$ and $t_i = 1$) there is a point $\boldsymbol{x} \in \{0,1\}^n$ for which $f(\boldsymbol{x}) \neq 0$. We have $\boldsymbol{x} \neq \mathbf{0}$, as $f(\mathbf{0}) = 1 - 1 = 0$. But then $\boldsymbol{a}_i \cdot \boldsymbol{x} = 1$ for some $i \in [m]$ (as $\boldsymbol{x}$ is covered by some H_i), implying that f vanishes on this point: a contradiction. ▮

Notes

The bound on m in Theorem 50 is clearly tight. The paper [AF93] contains several extensions of the result.

11.4 Chevalley–Waring Theorem

Here we deduce the following version of the Chevalley–Waring Theorem.

Theorem 51 (Chevalley–Waring Theorem) *Let p be a prime, and let*

$$P_1, \dots, P_m \in \mathbb{F}_p[x_1, \dots, x_n]$$

be some polynomials in n variables. If $n > \sum_{i=1}^{m} \deg(P_i)$ and the polynomials P_i's have a common zero $(c_1, \dots, c_n) \in (\mathbb{F}_p)^n$, then they have another common zero.

Proof. Suppose that the claim is false. Define

$$f(x_1, \dots, x_n) = \prod_{i=1}^{m}(1 - (P_i(x_1, \dots, x_n))^{p-1}) - \prod_{j=1}^{n}(1 - (x_j - c_j)^{p-1}).$$

Observe that $f(\boldsymbol{c}) = 1 - 1 = 0$. Let $\boldsymbol{x} \in (\mathbb{F}_p)^n \setminus \{\boldsymbol{c}\}$. There is, by our assumption, a polynomial P_i that does not vanish on $\boldsymbol{x}$, implying by Fermat's Little Theorem (Theorem 43) that $(P_i(\boldsymbol{x}))^{p-1} = 1$. For some $j \in [n]$ we have $x_j \neq c_j$; then $(x_j - c_j)^{p-1} = 1$, again by Theorem 43. We conclude that the both summands are zero. Thus we have shown that

$$f(x_1, \dots, x_n) = 0, \quad \text{for any } \boldsymbol{x} \in (\mathbb{F}_p)^n. \tag{52}$$

The total degree of the first summand of f is at most $(p-1)\sum_{i=1}^{m} \deg(P_i) < (p-1)n$. On the other hand, the coefficient at $x_1^{p-1} \dots x_n^{p-1}$ in f is $(-1)^{n+1} \neq 0$, so $\deg(f) = (p-1)n$.

By Combinatorial Nullstellensatz (Theorem 42 for $S_i = \mathbb{F}_p$, $t_i = p-1$) we conclude that there is $\boldsymbol{x} \in (\mathbb{F}_p)^n$ for which $f(\boldsymbol{s}) \neq 0$, contradicting (52) and completing the proof. ▮

Notes

In fact, a stronger version of Theorem 51 is true: if there is one common root of $P_1, \ldots, P_m$, then there are at least p common roots. But for our application in Section 11.5 the presented version suffices.

11.5 Sets Meeting Every Affine Hyperplane

A *blocking set* in a hypergraph $\mathcal{H}$ is a set of vertices which intersects every edge.

For a prime p let $\mathcal{H}_{n,p}$ be the hypergraph on $(\mathbb{F}_p)^n$ consisting of all $(n-1)$-dimensional affine hyperplanes. Thus, $\mathcal{H}$ has p^n vertices and its every edge has p^{n-1} vertices.

Lemma 52 *Let $B \subset (\mathbb{F}_p)^n$ consist of all n-vectors with at most one non-zero coordinate. (Thus, for example, $|B| = 1 + n(p-1)$.) Then B is a blocking set for $\mathcal{H}_{n,p}$.*

Proof. We prove the claim by induction on n with the case $n = 1$ being trivially true (then $B = \mathbb{F}_p$). Take any $E \in \mathcal{H}_{n,p}$. Let

$$D := \{\boldsymbol{x} \in (\mathbb{F}_p)^n : x_n = 0\} \in \mathcal{H}_{n,p}.$$

If E is *parallel* to D (that is, $E \cap D = \emptyset$ or $E = D$), then $E \cap B \ni (0, \ldots, 0, c)$ for some $c \in \mathbb{F}_p$ and we are home. Otherwise apply induction to the $(n-2)$-dimensional hyperplane $E \cap D$ that lives inside D which can be naturally identified with $(\mathbb{F}_p)^{n-1}$. ∎

The above lemma is sharp in the following sense.

Theorem 53 (Jamison [Jam77]; Brouwer & Schrijver [BS78]) *Let p be a prime. If B is a blocking set in $\mathcal{H}_{n,p}$, then $|B| \geq n(p-1) + 1$.*

Proof. By translating B we may assume that $\mathbf{0} \in B$. Let $A := B \setminus \{\mathbf{0}\}$. Then A intersects all $(n-1)$-dimensional hyperplanes not containing $\mathbf{0}$. Thus, for every $\boldsymbol{x} \in (\mathbb{F}_p)^n \setminus \{\mathbf{0}\}$ the equation $\boldsymbol{x} \cdot \boldsymbol{a} = 1$ has a solution $\boldsymbol{a} \in A$. Let

$$f(\boldsymbol{x}) := \prod_{\boldsymbol{a} \in A} (1 - \boldsymbol{x} \cdot \boldsymbol{a}).$$

We have $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in (\mathbb{F}_p)^n \setminus \{\mathbf{0}\}$ and $f(\mathbf{0}) = 1$.

Define the polynomial

$$F(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}) := -p + 1 + \sum_{i=1}^{p-1} f(\boldsymbol{x}_i), \quad \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1} \in (\mathbb{F}_p)^n,$$

in the $n(p-1)$ variables x_{ij}, $i \in [p-1]$, $j \in [n]$. As f assumes only values 0 and 1, we conclude that F is zero iff each $f(\boldsymbol{x}_i) = 1$. This implies that $\mathbf{0}$ is the only root of F.

By the Chevalley–Waring Theorem (Theorem 51 with $m = 1$), $\deg(F)$ is at least $n(p-1)$, the number of variables. We have

$$\deg(F) \leq \deg(f) \leq |A| = |B| - 1,$$

which implies the theorem. ∎

Notes

Theorem 53 was proved independently by Jamison [Jam77] and by Brouwer and Schriver [BS78]. The proof we present is due to Alon [Alo95].

12 Desarguesian Projective Plane $PG(2,q)$

The below material is EXAMINABLE. As it has only partly been covered in the lectures, it is for the independent study.

Let q be a prime or a prime power. Define

$$V := \{[x,y,z] : (x,y,z) \in (\mathbb{F}_q)^3 \setminus \{\mathbf{0}\}\},$$

where $[x,y,z] := \{(\lambda x, \lambda y, \lambda z) : \lambda \in \mathbb{F}_q \setminus \{0\}\}$. Alternatively, one could identify V with the set of 1-dimensional linear subspaces of $(\mathbb{F}_q)^3$. We will refer to the elements of V as *points*. Also we define the set

$$L := \{\langle a,b,c\rangle : (a,b,c) \in (\mathbb{F}_q)^3 \setminus \{\mathbf{0}\}\},$$

consisting of *lines*

$$\langle a,b,c\rangle := \{[x,y,z] : ax+by+cz=0\}.$$

Alternatively, one could identify L with the set of 2-dimensional linear subspaces of $(\mathbb{F}_q)^3$, but we do not pursue this point of view here.

The *Desarguesian projective plane* of order q is $PG(2,q) := (V,L)$. Let us derive some properties of $PG(2,q)$.

Claim 1 $|V| = q^2+q+1$.

Proof of Claim. Vertices $(x,y,z) \in (\mathbb{F}_q)^3 \setminus \{\mathbf{0}\}$ with $x \neq 0$ give us the following q^2 points: $[1,y,z]$, $(x,y) \in (\mathbb{F}_q)^2$. Vertices with $x=0$ but $y \neq 0$ give us q points: $[0,1,z]$, $z \in \mathbb{F}_q$. Finally, vertices with $x=y=0$ make the point $[0,0,1]$. ▌

Claim 2 Any line $l = \langle a,b,c\rangle \in L$ has precisely $q+1$ points.

Proof of Claim. Assume, for example, that $c \neq 0$. In the equation $ax+by+cz=0$ the value of z is uniquely determined by specifying any $(x,y) \in (\mathbb{F}_q)^2$, which gives us a point of l except for $(x,y)=(0,0)$. However, every point $v \in l$ is counted $q-1$ times. Hence, l has $\frac{q^2-1}{q-1} = q+1$ points. ▌

Similarly we show that every point belong to precisely $q+1$ lines. Counting the number of incident pairs $(v \in l)$ we obtain $|V|(q+1) = |L|(q+1)$, that is, $|L| = |V| = q^2+q+1$.

If we delete, for example, the line $l_0 := \langle 0,0,1\rangle$ from V, then the remaining vertices can be identified with $(\mathbb{F}_q)^2$ via the bijections

$$\begin{aligned} V \setminus l_0 \ni [x,y,z] &\mapsto (x/z, y/z) \in (\mathbb{F}_q)^2 \\ (\mathbb{F}_q)^2 \ni (x,y) &\mapsto [x,y,1] \in V \setminus l_0. \end{aligned}$$

Under this correspondence a line $\langle a,b,c\rangle \neq l_0$ corresponds to the affine 1-dimensional subspace $\{(x,y) \in (\mathbb{F}_q)^2 : ax+by+c=0\}$ and vice versa. In other words, we get the hypergraph $\mathcal{H}_{2,q}$ from Section 11.5; usually $\mathcal{H}_{2,q}$ is denoted by $AG(2,q)$ and is called the *Desarguesian affine plane.*

Claim 3 For any two distinct points $v=[x,y,z]$ and $v'=[x',y',z']$ there is at least one line l containing both.

Proof of Claim. If $z=z'=0$, we can take l_0. If, for example, $z=0$ but $z'\neq 0$, then we can choose c such that $v'\in\langle y,-x,c\rangle$; this line also contains $v=[x,y,0]$. Finally, if $z\neq 0$ and $z'\neq 0$, then we take the 1-dimensional affine subspace in $AG(2,q)$ through v and v' (using the above identification). ▎

Let us count the number m of pairs (A,l) where l is a line and A is a 2-subset of l. Clearly, $m=|L|\binom{q+1}{2}$. On the other hand, by Claim 3, each pair of points $\{v,v'\}\in\binom{V}{2}$ contributes at least 1 to m, that is, $m\geq\binom{|V|}{2}$. Thus we obtain

$$(q^2+q+1)\,\frac{(q+1)q}{2}=m\geq\frac{(q^2+q+1)(q^2+q)}{2},$$

and we have in fact equality. This means that no pair of points can belong to more than one line. Hence we have shown the following, very useful property.

Claim 4 There is exactly one line through any pair of distinct points of V. ▎

In the similar fashion we count the number n of pairs $(\{l,l'\},v)$, where $l,l'\in L$ are distinct and $v\in l\cap l'$, and obtain

$$\binom{|L|}{2}\geq n=|V|\binom{q+1}{2},$$

which implies the following.

Claim 5 Every two distinct lines intersect in precisely one point. ▎

Let us present two constructions involving $PG(2,q)$.

For a bipartite graph F let $\mathrm{ex}(n,n,F):=\max\{e(G):G\subset K_{n,n},\ G\not\supset F\}$. The question to determine this function for a complete bipartite graph F is usually known the *problem of Zarankiewicz*, see [Zar51]. Here we consider the 4-cycle $C_4=K_{2,2}$.

Theorem 54 (Kővari, Sós & Turán [KST54]) $\mathrm{ex}(n,n,C_4)=(1+o(1))\,n^{3/2}$.

Proof. We show the upper bound. Let G be a bipartite C_4-free graph with parts V_1,V_2, each of size n. There are $\sum_{x\in V_1}\binom{d(x)}{2}$ pairs $(x,\{y,z\})$, where $x\in V_1$ and $y,z\in V_2$ are distinct neighbours of x. On the other hand, for any distinct $y,z\in V_2$ there is at most one such x (otherwise we obtain C_4). Hence,

$$\sum_{x\in V_1}\binom{d(x)}{2}\leq\binom{n}{2}. \tag{53}$$

The *Arithmetic–Quadratic Mean Inequality* says that

$$\frac{d_1+\cdots+d_n}{n}\leq\sqrt{\frac{d_1^2+\cdots+d_n^2}{n}}. \tag{54}$$

(Proof: Apply the Cauchy–Schwarz Inequality (Lemma 34) to the vectors $\boldsymbol{d}, \mathbf{1} \in \mathbb{R}^n$.)

It is straightforward to deduce from (53) and (54) that $e(G) = \sum_{x \in V_1} d(x) \le (1 + o(1))\, n^{3/2}$, which proves the upper bound.

For the lower bound choose the largest prime p with $p^2 + p + 1 \le n$. By the Prime Distribution Theorem (Corollary 18) we know that $p = (1 + o(1))\sqrt{n}$. Consider $PG(2, p) = (V, L)$. Define the bipartite graph G on $V \cup L$ by connecting $v \in V$ to $l \in L$ iff $v \in l$. Claim 4 (or Claim 5) implies that $G \not\supset C_4$. On the other hand, $e(G) = |V|(p + 1) = (p^2 + p + 1)(p + 1)$, which gives us the required lower bound. ▮

The *Turán function* of a graph F is $\mathrm{ex}(n, F) := \max\{e(G) : v(G) = n,\ G \not\supset F\}$.

Theorem 55 (Brown [Bro66]; Erdős, Rényi & Sós [ERS66]) $\mathrm{ex}(n, C_4) = (\frac{1}{2} + o(1))\, n^{3/2}$.

Proof. The upper bound is proved in the similar way as in Theorem 54: given a C_4-free graph G of order n we conclude that

$$\sum_{x \in V(G)} \binom{d(x)}{2} \le \binom{n}{2},$$

and applying (54) we obtain the required upper bound

$$2e(G) = \sum_{x \in V(G)} d(x) \le (1 + o(1))\, n^{3/2}.$$

Let us prove the lower bound. If we take the construction of Theorem 54, then we obtain $\mathrm{ex}(2n, C_4) \ge (1 + o(1))\, n^{3/2}$, which, after scaling, gives only $\mathrm{ex}(n, C_4) \ge (2^{-3/2} + o(1))\, n^{3/2}$.

The idea is to 'squash' the parts V_1 and V_2 into one. This trick does not work in general (a copy of C_4 may be created) but in our case we have a very nice symmetry between lines and points. Namely, let the bijection $P : V \longleftrightarrow L$ be defined by $[x, y, z] \longleftrightarrow \langle x, y, z \rangle$. Clearly, the composition $P \circ P$ is the identity map and the map P 'reverses' the incidence relation: $v \in l$ iff $P(v) \ni P(l)$.

So we choose the largest prime p with $p^2 + p + 1 \le n$. Let $V(G) = V$ consist of the points of $PG(2, p)$. We connect $v, v' \in V$ iff $v \ne v'$ and $v \in P(v')$. Can we have a 4-cycle on $v_1, v_2, v_3, v_4 \in V$? No, because then the lines $P(v_2)$ and $P(v_4)$ would intersect in two points v_1, v_3, a contradiction.

Each vertex $v \in G$ has degree either $p + 1$ or p (depending on whether $v \in P(v)$). Thus, $e(G) \ge p|V|$, which gives the required lower bound. ▮

Notes

Füredi [Für96a] shows that the construction in Theorem 55 is in fact best possible. More precisely, for any prime power $q \ge 15$ (including $q = 2^k$), we have

$$\mathrm{ex}(q^2 + q + 1, C_4) = \frac{q(q + 1)^2}{2}.$$

Clearly, $\mathrm{ex}(2n, K_{t,t}) \geq \mathrm{ex}(n, n, K_{t,t})$ while, using a version of Lemma 45, one can show that $\mathrm{ex}(n, n, K_{t,t}) \geq \frac{1}{2}\,\mathrm{ex}(2n, K_{t,t})$. Thus, in a sense, these functions behave similarly. Kővari, Sós and Turán [KST54] proved that for fixed t we have $\mathrm{ex}(n, K_{t,t}) = O(n^{2-1/t})$ which is conjectured to be the correct magnitude. The conjecture has been verified for $t = 2$ and $t = 3$ only.

Execises and Further Reading

[Für96b]: A very nice paper proving $\mathrm{ex}(n, K_{3,3}) = (\frac{1}{2} + o(1))\, n^{5/3}$.

[Bol95]: Chapter 1.3 has an overview of $\mathrm{ex}(n, m, K_{s,t})$ and $\mathrm{ex}(n, K_{s,t})$.

[KRS96]: An important paper with an algebraic construction of a $K_{s,t}$-free graph of order n with $\Theta(n^{2-1/t})$ edges, for any fixed $s \geq t!$. (Unfortunately, the method does not work for $s = t$, say.)

13 Designs

One use of combinatorial objects, called *designs*, originates from statistical applications. Let us assume that the wine committee wants to compare v varieties of wines. In order to make the testing procedure as fair as possible it is natural to require that

1. each member of the committee tests the same number (say k) of varieties so that each person's opinion has the same weight;

2. each pair of of varieties is compared by the same number (say λ) of persons so that each variety gets the same treatment.

One possibility would be to let everyone taste all the varieties. But if v is large, this is very impractical (if not dangerous, as in the case of wine). Thus, one would wish to design a tasting with small k.

Here is the formal definition. Let $V = [v]$ be the vertex set. A (v, k, λ)-*design* is a hypergraph $\mathcal{H} \subset \binom{V}{k}$ such that every pair of vertices is contained in exactly λ edges. Usually, the edges of $\mathcal{H}$ are called *blocks*.

For example, $PG(2, q)$ is a $(q^2+q+1, q+1, 1)$-design, while $AG(2, q)$ is a $(q^2, q, 1)$-design.

It is easy to see that every vertex x in a (v, k, λ)-design belongs to the same number r of blocks. Indeed, removing x from these blocks, we obtain by the definition a collection of $(k-1)$-sets covering each of the remaining $v-1$ vertices exactly λ times, that is, we have $r = \frac{\lambda(v-1)}{k-1}$. Also, it is easy to see that the total number of blocks is $b = \frac{vr}{k} = \frac{\lambda v(v-1)}{k(k-1)}$.

Clearly, a necessary condition for the existence of a (v, k, λ)-design is that $\frac{\lambda(v-1)}{k-1}$ and $\frac{\lambda v(v-1)}{k(k-1)}$ are integers. Wilson, in a series of papers [Wil72a, Wil72b, Wil75], proved that these trivial necessary conditions are also sufficient if v is sufficiently large, $v \geq v_0(\lambda, k)$. Here we try to present some partial results in this direction.

The first step of Wilson's programme was to construct (v, k, λ)-designs at least for some large v. Here we concentrate on the case $\lambda = 1$.

13.1 Difference Families

Our approach is to take some Abelian group $(G, +)$ and try to find a *difference family* $\mathcal{D} \subset \binom{G}{k}$, when we require that for any element $x \in G \setminus \{0\}$ there is the unique triple (A, a, b) with $A \in \mathcal{D}$, $a, b \in A$ and $x = a - b$. If this is the case, then the set system $\{A + a : A \in \mathcal{D}, a \in G\}$, made of the translates of $\mathcal{D}$, is obviously a $(q, k, 1)$-design.

To keep things as simple as possible, let us take $(\mathbb{F}_q, +)$ as the group and try to construct our difference family by taking multiples of some set A. Namely, let $\mathcal{D} = \{sA : s \in C\}$ for some $C, A \subset \mathbb{F}_q$, $C \not\ni 0$. To keep things even simpler, let us assume that C is a subgroup of $\mathbb{F}_q^* := (\mathbb{F}_q \setminus \{0\}, \times)$, the multiplicative group of $\mathbb{F}_q$. Then, for any $a, b \in A$, the differences $sa - sb$, $s \in C$, span the C-coset containing $a - b$. As the cosets of C partition $\mathbb{F}_q^*$, all we have to ensure is that there are precisely $k(k-1)$ cosets (or, equivalently, that $|C| = \frac{q-1}{k(k-1)}$) and all possible $k(k-1)$ differences $a - b$, over ordered pairs $a, b \in A$, fall bijectively into different cosets.

However, if $-1 \in C$, then this will not work because $a - b$ and $b - a$ always fall into the same coset. There are two ways around this problem. The first one is of course to require that $-1 \notin C$. The second option is to take only half the elements of C. As we will see, both solutions work. Let us state what to do in the case $-1 \in C$.

Lemma 56 *Let $C \subset \mathbb{F}_q^*$ be a subgroup of* index $|\mathbb{F}_q^*|/|C| = \binom{k}{2}$ *containing -1. Suppose that there exists a k-set $A = \{a_1, \ldots, a_k\} \subset \mathbb{F}_q$ such that the $\binom{k}{2}$ differences $a_j - a_i$, $1 \le i < j \le k$, fall in distinct C-cosets. (Thus every coset of C receives precisely one of these differences.) Then there exists a $(q, k, 1)$-design.*

Proof. Let S consist of half of the elements of C, one element from each pair $\{x, -x\} \subset C$. Our design is

$$\mathcal{H} = \{sA + t : s \in S, t \in \mathbb{F}_q\}.$$

Let us see how many edges of $\mathcal{H}$ contain some two distinct $x, y \in \mathbb{F}_q$. Choose the (unique) $\{a, b\} \in \binom{A}{2}$ such that $a - b$ and $x - y$ lie in the same C-coset. The elements $\pm\frac{x-y}{a-b}$ belong to C; let s be the one in S, say $s = \frac{x-y}{a-b}$. Let $t = x - sa$. Then $x = sa + t$ and $y = sb + t$, that is, we have found a block containing both x and y. On the other hand it is easy to argue that such a block is unique by, for example, observing that we have the right number $\frac{q-1}{2m} \cdot q = \binom{q}{2}/\binom{k}{2}$ of blocks. ∎

13.2 Cyclotomic Classes

Let us now look at the structure of $\mathbb{F}_q^*$.

Lemma 57 *The multiplicative group $\mathbb{F}_q^*$ of any finite field $\mathbb{F}_q$ is cyclic. In other words, there is an element $\gamma \in \mathbb{F}_q^*$, called* primitive*, such that $\mathbb{F}_q^* = \{1, \gamma, \gamma^2, \ldots, \gamma^{q-2}\}$.*

Proof. For $a \in \mathbb{F}_q^*$ let its *order* be $d(a) := \min\{d \ge 1 : a^d = 1\}$. Clearly, $d(a) \le |\mathbb{F}_q^*|$ is well-defined.

Claim 1 If $a^s = 1$, then $d(a) \mid s$ (i.e. $d(a)$ divides s).

Proof of Claim. Divide s by $d(a)$: $s = td(a) + r$, $r \in [0, d(a) - 1]$. We have

$$1 = a^s = (a^{d(a)})^t a^r = a^r,$$

which implies that $r = 0$. ▎

Claim 2 If $k := d(a)$ and $l := d(b)$ are coprime, then $d(ab) = kl$.

Proof of Claim. Let $s := d(ab)$. As $(ab)^{kl} = 1$, we conclude that $s \le kl$. On the other hand, the identity $1 = ((ab)^s)^k = b^{sk}$ implies by Claim 1 that $l \,|\, s$. Similarly, $k \,|\, s$; hence $s = kl$. ▎

Now, let $a \in \mathbb{F}_q^*$ be an element of the maximum order $d(a) =: k$.

Claim 3 For any $b \in \mathbb{F}_q^*$ we have $l \,|\, k$, where $l := d(b)$.

Proof of Claim. Suppose that the claim is not true. This means that there is a prime p such that $k = p^\alpha k'$ and $l = p^\beta l'$ with $\beta > \alpha$, $p \nmid k'$ and $p \nmid l'$.

It is easy to see that $d(a^{p^\alpha}) = k'$ and $d(b^{l'}) = p^\beta$. As these numbers are coprime, we conclude by Claim 2 that $d(a^{p^\alpha} b^{l'}) = k' p^\beta > k$, which contradicts the definition of k. ▎

Thus the polynomial $x^k - 1$ has $q - 1$ roots, so by Lemma 39 we must have $k \ge q - 1$. As $k \le |\mathbb{F}_q^*|$, it follows $k = q - 1$. Thus the elements $1, a, a^2, \ldots, a^{q-2}$ are distinct, so every element of $\mathbb{F}_q^*$ appears in this list exactly once. ▮

In particular, Lemma 57 implies the following result which can be viewed as a generalisation of Fermat's Little Theorem (Theorem 43).

Corollary 58 *Let q be a prime power. Then $a^{q-1} = 1$ for any $a \in \mathbb{F}_q \setminus \{0\}$.* ▮

Let $q = mf + 1$ be a prime power. By Lemma 57 we have $\mathbb{F}_q^* \cong (\mathbb{Z}_{mf}, +)$, so $\mathbb{F}_q^*$ has the unique subgroup C_0 of index m (and order f):

$$C_0 = \{x \in \mathbb{F}_q^* : x^f = 1\}. \tag{55}$$

The multiplicative cosets $C_0, C_1, \ldots, C_{m-1}$ of C_0 are called the *cyclotomic classes* of index m. They partition $\mathbb{F}_q^*$. For $x \in \mathbb{F}_q^*$ define the *cyclotomic index* $c(x) \in \mathbb{Z}_m$ by $x \in C_{c(x)}$.

Observe by (55) that the cyclotomic class C_0 contains -1 if and only if f is even. If $-1 \in C_0$, then $c(x - y) = c(y - x)$ for any distinct $x, y \in \mathbb{F}_q$. If $-1 \notin C_0$, then all cyclotomic classes split into the pairs $(C, -C)$ and, again, $c(x - y)$ determines $c(y - x)$.

It seems that, apart from these, there are no other obvious necessary relations between the additive structure and cyclotomic indexes. Given this meek and unconcrete statement, the following bold question comes as a surprise.

Problem 59 *Can we always find k elements in $\mathbb{F}_q$ so that the cyclotomic indexes of their differences have any, beforehand specified, values compatible with the relation between $c(x)$ and $c(-x)$?*

Let us formalise the problem. Let r be an integer. Let $\boldsymbol{l}$ be a $\mathbb{Z}_m$-*bisequence*, that is, a sequence $l_{ij} \in \mathbb{Z}_m$ indexed by pairs $i < j$ of positive integers. Let $X^{(r)}$ be the set of all ordered r-tuples of distinct elements of X. For example, the size of $[n]^{(r)}$ is $n^{(r)} := n(n-1)\dots(n-r+1)$. Define

$$\mathcal{N}(r, \boldsymbol{l}) := \{\boldsymbol{a} \in \mathbb{F}_q^{(r)} : \forall\, 1 \le i < j \le r \;\; c(a_j - a_i) = l_{ij}\}. \tag{56}$$

Note that as $c(0)$ is undefined, we restrict the consideration to $\mathbb{F}_q^{(r)}$. Now, the Problem 59 asks whether $\mathcal{N}(r, \boldsymbol{l})$ is non-empty for any r and $\boldsymbol{l}$.

To prevent cases like $r > q$, when the answer is clearly in the negative, let us assume that q is sufficiently large while r and m are fixed. The straightforward approach to Problem 59 of constructing $\boldsymbol{a}$ by using induction on r does not work: there can be partial inextensible sequences. Having said this, the next thing we say is that... the induction does work: the trick is to prove a stronger claim! Namely, let us try to prove that $|\mathcal{N}(r, \boldsymbol{l})| = \Theta(q^r)$, that is, $\mathcal{N}(r, \boldsymbol{l})$ contains a fixed proportion of elements of $\mathbb{F}_q^{(r)}$.

Our proof will use induction on r with the case $r = 1$ being trivially true: $\mathcal{N}(1, \boldsymbol{l}) = \mathbb{F}_q$. Suppose it holds for some r. An $(r+1)$-tuple in $\mathcal{N}(r+1, \boldsymbol{l})$ can be constructed by taking an r-tuple $\boldsymbol{a} \in \mathcal{N}(r, \boldsymbol{l})$ and adding an extra element. In other words, we have the following representation

$$\mathcal{N}(r+1, \boldsymbol{l}) = \bigcup_{\boldsymbol{a} \in \mathcal{N}(r, \boldsymbol{l})} \mathcal{M}_{l_{1,r+1}, l_{2,r+1}, \dots, l_{r,r+1}}(\boldsymbol{a}), \tag{57}$$

where for $\boldsymbol{i} = (i_1, \dots, i_r) \in \mathbb{Z}_m^r$ and $\boldsymbol{a} = (a_1, \dots, a_r) \in \mathbb{F}_q^{(r)}$ we denote

$$\mathcal{M}_{\boldsymbol{i}}(\boldsymbol{a}) = \{x \in \mathbb{F}_q \setminus \{a_1, \dots, a_r\} : \forall j \in [r] \; c(x - a_j) = i_j\}.$$

Also, let $m_{\boldsymbol{i}}(\boldsymbol{a}) = |\mathcal{M}_{\boldsymbol{i}}(\boldsymbol{a})|$. The identity (57) tells us that $|\mathcal{N}(r+1, \boldsymbol{l})|$ is the sum of $|\mathcal{N}(r, \boldsymbol{l})|$ quantities $m_{\boldsymbol{l}}(\boldsymbol{a})$.

So we would like to have a result which states that if you take many (a positive proportion) of elements from the sequence $\boldsymbol{m}$, then their sum is still big. One of the standard tools for doing this is to compute the variance of $\boldsymbol{m}$: if it is small, then $\boldsymbol{m}$ is closely concentrated around its mean and we might be able to deduce some estimates.

13.3 Tools: Mean and Variance

Let us give the corresponding definitions and results. Given a sequence $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ of real numbers, its *mean* is

$$\mathrm{E}(\boldsymbol{\alpha}) := \frac{1}{n}(\alpha_1 + \dots + \alpha_n)$$

and its *variance* is

$$\mathrm{Var}(\boldsymbol{\alpha}) := \frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \mathrm{E}(\boldsymbol{\alpha}))^2$$

These definitions and terminology are motivated by the following observation: if $i \in [n]$ is the uniformly distributed random variable, then the (probabilistic) mean and variance of α_i are given by the above formulae.

The abstract problem is: given a sequence $\alpha_1, \ldots, \alpha_n$ of small variance can we deduce that the sum of any l members is large by showing that this sum is not far away from $l\,\mathrm{E}(\boldsymbol{\alpha})$? Without loss of generality, assume that we are interested in estimating

$$d := |(\alpha_1 + \ldots + \alpha_l) - l\,\mathrm{E}(\boldsymbol{\alpha})|.$$

If we add some constant to each α_i, then d and the variance do not change, so without loss of generality we can assume that $\mathrm{E}(\boldsymbol{\alpha}) = 0$.

Observe that $\alpha_1 + \cdots + \alpha_l = \boldsymbol{\alpha} \cdot \boldsymbol{\chi}_{[l]}$. Thus we have to bound $|\boldsymbol{\alpha} \cdot \boldsymbol{\chi}_{[l]}|$ in terms of $\mathrm{Var}(\boldsymbol{\alpha}) = \frac{1}{n}(\alpha_1^2 + \cdots + \alpha_n^2) = \frac{1}{n}\,\|\boldsymbol{\alpha}\|^2$. This is easy by applying the Cauchy–Schwarz Inequality (Lemma 34) to vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\chi}_{[l]}$:

$$(\alpha_1 + \cdots + \alpha_l)^2 = (\boldsymbol{\alpha} \cdot \boldsymbol{\chi}_{[l]})^2 \le \|\boldsymbol{\alpha}\|^2 \cdot \|\boldsymbol{\chi}_{[l]}\|^2 = \|\boldsymbol{\alpha}\|^2 \cdot l \tag{58}$$

It is not crucial for our purposes but let us be perfectionists (at least here) and improve the bound by noting that $\boldsymbol{\alpha} \cdot \mathbf{1} = 0$, so the left-hand side of (58) does not change if we replace $\boldsymbol{\chi}_{[l]}$ by $\boldsymbol{\chi}_{[l]} + \beta\mathbf{1}$, where we are free to choose $\beta \in \mathbb{R}$.

$$(\alpha_1 + \cdots + \alpha_l)^2 = (\boldsymbol{\alpha} \cdot (\boldsymbol{\chi}_{[l]} + \beta\mathbf{1}))^2 \le \|\boldsymbol{\alpha}\|^2 \cdot \|\boldsymbol{\chi}_{[l]} + \beta\mathbf{1}\|^2$$

To make $\|\boldsymbol{\chi}_{[l]} + \beta\mathbf{1}\|^2 = l(1+\beta)^2 + (n-l)\beta^2$ as small as possible, we let $\beta = -l/n$ and obtain

$$(\alpha_1 + \cdots + \alpha_l)^2 \le (\alpha_1^2 + \cdots + \alpha_n^2)\,\frac{l(n-l)}{n} = l(n-l)\mathrm{Var}(\boldsymbol{\alpha}).$$

Now it remains to restate the obtained inequality for arbitrary $\mathrm{E}(\boldsymbol{\alpha})$.

Lemma 60 *For any sequence $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and $l \in [0, n]$, we have*

$$((\alpha_1 + \ldots + \alpha_l) - l \cdot \mathrm{E}(\boldsymbol{\alpha}))^2 \le l(n-l)\,\mathrm{Var}(\boldsymbol{\alpha}) \le \frac{n^2}{4}\,\mathrm{Var}(\boldsymbol{\alpha}). \blacksquare \tag{59}$$

13.4 Computing Mean and Variance for $\boldsymbol{m}$

Let us try to compute the mean and variance of $\boldsymbol{m}$.

It is immediate that for $\boldsymbol{a} \in \mathbb{F}_q^{(r)}$ we have $\sum_{\boldsymbol{i}} m_{\boldsymbol{i}}(\boldsymbol{a}) = q - r$: any $x \in \mathbb{F}_q \backslash \{a_1, \ldots, a_r\}$ belongs to precisely one of $\mathcal{M}_{\boldsymbol{i}}(\boldsymbol{a})$, $\boldsymbol{i} \in \mathbb{Z}_m^r$. Thus

$$\sum_{\boldsymbol{a} \in \mathbb{F}_q^{(r)}} \sum_{\boldsymbol{i} \in \mathbb{Z}_m^r} m_{\boldsymbol{i}}(\boldsymbol{a}) = q^{(r)}(q-r) = q^{(r+1)}.$$

The mean over all $q^{(r)}m^r$ pairs $(\boldsymbol{a}, \boldsymbol{i})$ is

$$\mathrm{E}(\boldsymbol{\alpha}) = \frac{1}{q^{(r)}m^r} \times q^{(r+1)} = \frac{q-r}{m^r}$$

For the variance of $\boldsymbol{\alpha}$ we use the standard trick of computing $Q(\boldsymbol{\alpha}) := \sum_{i=1}^n \alpha_i^{(2)}$ first. (There seems to be no special name for $Q(\boldsymbol{a})$, unfortunately.) This is often an easier way than computing $\mathrm{Var}(\boldsymbol{\alpha})$ directly as $\alpha_i^{(2)} = \alpha_i(\alpha_i - 1)$ can be interpreted combinatorially. The formula that expresses the variance as a function of $Q(\boldsymbol{\alpha})$ and $\mathrm{E}(\boldsymbol{\alpha})$ is easy to find. First we expand the definition of variance:

$$\mathrm{Var}(\boldsymbol{\alpha}) = \frac{n-1}{n^2}\sum_{i=1}^{n}\alpha_i^2 - \frac{2}{n^2}\sum_{1\le i<j\le n}\alpha_i\alpha_j.$$

Then we choose consecutively a multiple of $(\mathrm{E}(\boldsymbol{\alpha}))^2$, $Q(\boldsymbol{\alpha})$, and $\mathrm{E}(\boldsymbol{\alpha})$ equalising the coefficients at $\sum_{i<j}\alpha_i\alpha_j$, $\sum_i \alpha_i^2$, and $\sum_i \alpha_i$. We obtain

$$\mathrm{Var}(\boldsymbol{\alpha}) = -(\mathrm{E}(\boldsymbol{\alpha}))^2 + \frac{1}{n}\,Q(\boldsymbol{\alpha}) + \mathrm{E}(\boldsymbol{\alpha}). \tag{60}$$

Clearly, $(m_{\boldsymbol{i}}(\boldsymbol{a}))^{(2)}$ is the number of pairs $(x,y)\in\mathbb{F}_q^{(2)}$ such that

$$c(x-a_j) = c(y-a_j) = i_j \text{ for all } j\in[r].$$

Thus, for fixed $\boldsymbol{a}\in\mathbb{F}_q^{(r)}$, $\sum_{\boldsymbol{i}}(m_{\boldsymbol{i}}(\boldsymbol{a}))^{(2)}$ is the number of $(x,y)\in\mathbb{F}_q^{(2)}$ such that $x-a_j$ and $y-a_j$ belong to the same cyclotomic class for all $j\in[r]$; and

$$\sum_{\boldsymbol{a}\in\mathbb{F}_q^{(r)}}\sum_{\boldsymbol{i}\in\mathbb{Z}_n^r}(m_{\boldsymbol{i}}(\boldsymbol{a}))^{(2)}$$

counts the number of $(r+2)$-tuples $(a_1,\ldots,a_r;x,y)\in\mathbb{F}_q^{(r+2)}$ with $c(x-a_j)=c(y-a_j)$, $j\in[r]$. For fixed $(x,y)\in\mathbb{F}_q^{(2)}$ all such $(r+2)$-tuples are obtained by choosing $a_1,\ldots,a_r$ as distinct elements of the set

$$Z(x,y) := \{z\in\mathbb{F}_q : c(x-z)=c(y-z)\}.$$

Note that $x,y\notin Z(x,y)$ as $c(0)$ is undefined.

Now $x-z$ and $y-z$ are in the same cyclotomic class if and only if $x-z=b(y-z)$ for some $b\in C_0\setminus\{1\}$. But for each $b\in C_0\setminus\{1\}$, there is the unique such solution z; that is,

$$|Z(x,y)| = |C_0|-1 = \frac{q-m-1}{m}$$

which is independent of x,y. Now we finish this double (or rather triple?) counting

$$\sum_{\boldsymbol{a}\in\mathbb{F}_q^{(r)}}\sum_{\boldsymbol{i}\in\mathbb{Z}_n^r}(m_{\boldsymbol{i}}(\boldsymbol{a}))^{(2)} = \sum_{(x,y)\in\mathbb{F}_q^{(2)}}\left(\frac{q-m-1}{m}\right)^{(r)} = q(q-1)\left(\frac{q-m-1}{m}\right)^{(r)}.$$

Finally, we compute the variance $\mathrm{Var}(\boldsymbol{m})$ using (60).

Lemma 61 *Fix r and m. The mean value of $\boldsymbol{m}$ over the $m^r q^{(r)}$ choices of $\boldsymbol{i}\in\mathbb{Z}_m^r$ and $\boldsymbol{a}\in\mathbb{F}_q^{(r)}$ is $\mathrm{E}(\boldsymbol{m})=(q-r)m^{-r}$ and the variance is*

$$\mathrm{Var}(\boldsymbol{m}) = -\frac{(q-r)^2}{m^{2r}} + \frac{q(q-1)}{m^r q^{(r)}}\left(\frac{q-m-1}{m}\right)^{(r)} + \frac{q-r}{m^r} = O(q). \ \blacksquare$$

Now we are able to answer Problem 59 in the affirmative.

Lemma 62 *Let m, r be fixed integers and $\boldsymbol{l}$ be any $\mathbb{Z}_m$-bisequence. Then $|\mathcal{N}(r, \boldsymbol{l})| = \Theta(q^r)$.*

Proof. We use induction on r. The claim is clearly true for $r = 1$ as $\mathcal{N}(1, \boldsymbol{l}) = \mathbb{F}_q$. Suppose that it holds for some r. By (57) we see that $|\mathcal{N}(r+1, \boldsymbol{l})|$ is the sum of $|\mathcal{N}(r, \boldsymbol{l})|$ quantities $m_{\boldsymbol{l}}(\boldsymbol{a})$ and we can apply Lemma 60 to conclude that

$$\Big| |\mathcal{N}(r+1, \boldsymbol{l})| - |\mathcal{N}(r, \boldsymbol{l})| \times \mathrm{E}(\boldsymbol{m}) \Big| \leq \frac{m^r q^{(r)}}{2} \sqrt{\mathrm{Var}(\boldsymbol{m})}.$$

We know that $\mathrm{E}(\boldsymbol{m}) = \Theta(q)$ and $\mathrm{Var}(\boldsymbol{m}) = O(q)$, which implies the claim. ▮

13.5 Putting Everything Together

We are finally on the finishing line! So, let k be given. Define $m = \binom{k}{2}$.

Lemma 63 *For any m there are arbitrarily large, prime powers $q \equiv 1 \pmod{2m}$.*

Proof. Take any prime p not dividing $2m$. By the Pigeon-Hole Principle there are $i < j$ such that $p^i \equiv p^j \pmod{2m}$. Then for any integer $l \geq 1$ we have $p^{l(j-i)} \equiv 1 \pmod{2m}$. ▮

Dirichlet's Theorem (see e.g. Landau [Lan74, pp. 422-446]) states that any arithmetic progression $\{ai + b : i \in \mathbb{N}\}$ with coprime a and b contains infinitely many primes. However, it is a very difficult theorem with a complicated proof.

I do not know how to prove (without appealing to Dirichlet's Theorem) that there is a prime power q of the form $m(2f+1)+1$. To be self-contained, we have to satisfy ourselves with the case $q = 2mf + 1$, dealt with by Lemma 63.

So, let $q = 2mf + 1$ be sufficiently large. Let $C_0, \ldots, C_{m-1}$ be the cyclotomic classes of index m. As $|C_0| = 2f$ is even, we have $\pm 1 \in C_0$. By Lemma 62 we can find a k-tuple $\boldsymbol{a} \in \mathbb{F}_q^{(k)}$ such that the $m = \binom{k}{2}$ differences $a_j - a_i$, $1 \leq i < j \leq k$, fall one-to-one into the m cyclotomic classes. Now by Lemma 56 we deduce the following theorem.

Theorem 64 (Wilson [Wil74]) *For any k and v_0 there is $v \geq v_0$ such that there exists a $(v, k, 1)$-design.* ▮

Remark. Our methods can be extended, with little extra work, to arbitrary λ. (See the Notes.)

Notes

Modifying these methods it is possible to prove the existence of a (v, k, λ)-design, given any fixed k and λ, for a series of values of v.

Here is an outline. Let $m = \binom{k}{2}$ and $q = 2\lambda mf + 1$ be a prime power. Let $C_0, \ldots, C_{\lambda m-1}$ be the cyclotomic classes of index λm. We have $\pm 1 \in C_0$. We can find

λ sets $A_1, \ldots, A_\lambda \in \binom{\mathbb{F}_q}{k}$ such that all $\lambda\binom{k}{2}$ possible differences fall one-to-one into the cyclotomic classes. Let $C'_0, \ldots, C'_{m-1}$ be the cyclotomic classes of index m. It is easy to see that C'_0 is the union of λ cosets of C_0. Now let S consist of 'half' the elements of C'_0 and consider the following hypergraph

$$\mathcal{H} = \{sA_i + t : s \in S,\, i \in [\lambda],\, t \in \mathbb{F}_q\}.$$

14 Eigenvalues and Expanders

In the whole of this section let $G = (V, E)$ be a graph with the vertex set $V = [n]$ and let A be the adjacency $n \times n$-matrix of G (over $\mathbb{R}$). We will study the relations between the eigenvalues of A and the properties of G.

Recall that a scalar λ is an *eigenvalue* of a square $n \times n$-matrix A if the equation $\mathsf{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ has a solution $\boldsymbol{x} \neq \mathbf{0}$, which is the case if and only if the *characterictic polynomial* $p_{\mathsf{A}}(y) := \det(y\mathsf{I}_n - \mathsf{A})$ has λ as a root. A non-zero $\boldsymbol{x}$ with $\mathsf{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ is called an *eigenvector* corresponding to the eigenvalue λ.

We were imprecise in the above definition by not specifying the underlying field in each case, which leads to some ambiguity. For example, how do we interpret complex roots of p_{A}? However, we do not have to worry about this because the following lemma guarantees that all eigenvalues (and eigenvectors) of A are real.

Lemma 65 *Let* M *be a real* $n \times n$*-matrix which is* symmetric *(that is,* $\mathsf{M}^T = \mathsf{M}$*). Then* M *has* n *real eigenvalues. Also, we can find an orthonormal basis of* $\mathbb{R}^n$ *made of corresponding eigenvectors.*

Proof. Clearly, $\deg(p_{\mathsf{M}}) = n$, so p_{M} has n (complex) roots. Take any root $\mu \in \mathbb{C}$ and choose $\boldsymbol{x} \in \mathbb{C}^n$ with $\mathsf{M}\boldsymbol{x} = \mu\boldsymbol{x}$.

Let $\overline{a + ib} := a - ib \in \mathbb{C}$ be the *complex adjoint.* On vectors it acts componentwise: $\overline{\boldsymbol{x}} = (\overline{x_1}, \ldots, \overline{x_n})$. Using the facts that M is real and that M is symmetric we obtain

$$\overline{\overline{\boldsymbol{x}} \cdot \mathsf{M}\boldsymbol{x}} = \boldsymbol{x} \cdot \mathsf{M}\overline{\boldsymbol{x}} = \overline{\boldsymbol{x}} \cdot \mathsf{M}\boldsymbol{x}.$$

We conclude that $\overline{\boldsymbol{x}} \cdot \mathsf{M}\boldsymbol{x} \in \mathbb{R}$. Now the identity

$$\overline{\boldsymbol{x}} \cdot \mathsf{M}\boldsymbol{x} = \overline{\boldsymbol{x}} \cdot \mu\boldsymbol{x} = \mu\, \|\boldsymbol{x}\|^2$$

implies that $\mu \in \mathbb{R}$ and proves the first claim of the lemma.

We do not prove the second claim, whose proof can be found in Horn and Johnson [HJ85, Theorem 2.5.6], for example. ▮

Thus, from now on, let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of A and $(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)$ be an orthonormal basis of $\mathbb{R}^n$ with $\mathsf{A}\boldsymbol{f}_i = \lambda_i\boldsymbol{f}_i$.

14.1 First Eigenvalue

The following formula computes the largest eigenvalue of A. (Also, one can compute the smallest eigenvalue by applying this formula to $-\mathsf{A}$.)

Lemma 66 (Raleigh–Ritz Formula) *Let* M *be a real symmetric* $n \times n$*-matrix with eigenvalues* $\mu_1 \geq \cdots \geq \mu_n$*. (Recall that, by Lemma 65, the eigenvalues of* M *are real.) Then*

$$\mu_1 = \max \left\{ \frac{\boldsymbol{x} \cdot \mathsf{M}\boldsymbol{x}}{\|\boldsymbol{x}\|^2} : \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\} \right\}. \tag{61}$$

Proof. Let $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n$ be corresponding orthonormal eigenvectors, cf. Lemma 65.

By the definition, $\boldsymbol{m}_1 \cdot \mathsf{M}\boldsymbol{m}_1 = \boldsymbol{m}_1 \cdot (\mu_1 \boldsymbol{m}_1) = \mu_1 \|\boldsymbol{m}_1\|^2$, which proves the upper bound in (61).

On the other hand, let $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$. Represent $\boldsymbol{x} = \sum_{i=1}^n c_i \boldsymbol{m}_i$. We have

$$\boldsymbol{x} \cdot \mathsf{M}\boldsymbol{x} = \Big(\sum_{i=1}^n c_i \boldsymbol{m}_i \Big) \cdot \Big(\sum_{i=1}^n \mu_i c_i \boldsymbol{m}_i \Big) = \sum_{i=1}^n \mu_i c_i^2 \leq \mu_1 \sum_{i=1}^n c_i^2 = \mu_1 \|\boldsymbol{x}\|^2, \tag{62}$$

which establishes the desired lower bound on μ_1. ▮

Even with the Raleigh–Ritz Formula it is not clear how to compute the largest eigenvalue of the adjacency matrix A. There are a few options for our further investigation. For example, we could substitute $\boldsymbol{x} := \boldsymbol{\chi}_A$, $A \subset V$, into the right-hand side of (61) and see which bounds on λ_1 this would imply.

But for simplicity (and with a foresight) let us now assume that G is d-regular. Then $\mathsf{A}\boldsymbol{1} = d\boldsymbol{1}$, so d is an eigenvalue and, as it is easy to see, it is the largest eigenvalue, $\lambda_1 = d$.

If we normalise $\boldsymbol{1} \in \mathbb{R}^n$, we obtain the vector $\frac{1}{\sqrt{n}}\boldsymbol{1} = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$. We can now assume, without loss of generality, that $\boldsymbol{f}_1 = \frac{1}{\sqrt{n}}\boldsymbol{1}$. Indeed, if k eigenvalues of A are equal to λ_1, then analysing (62), we see that

$$\boldsymbol{A}\boldsymbol{x} = \lambda_1 \boldsymbol{x} \text{ if and only if } \boldsymbol{x} \in \mathsf{H}, \tag{63}$$

where $\mathsf{H} := \mathrm{Span}\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k\}$ is the linear subspace spanned by $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k\}$. So, $(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k)$ can be replaced by any other orthonormal basis of H, in particular, by one containing $\frac{1}{\sqrt{n}}\boldsymbol{1} \in \mathsf{H}$.

14.2 Second Eigenvalue

It is surprising that, given these difficulties with the first eigenvalue of A, we now move to studying the even more obscure second eigenvalue. Let $\lambda = \lambda_2$. How can we compute λ? The key observation is that λ becomes the largest eingevalue if we restrict the action of A to the invariant subspace $\mathrm{Span}\{\boldsymbol{f}_2, \ldots, \boldsymbol{f}_n\} = \boldsymbol{f}_1^\perp$. A moment's thought reveals that, given $\boldsymbol{f}_1 = \frac{1}{\sqrt{n}}\boldsymbol{1}$, we have

$$\lambda = \max \left\{ \frac{\boldsymbol{x} \cdot \mathsf{A}\boldsymbol{x}}{\|\boldsymbol{x}\|^2} : \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\},\ \boldsymbol{x} \cdot \boldsymbol{1} = 0 \right\}. \tag{64}$$

What we can do now is to plug into (64) some concrete $\boldsymbol{x} \in \boldsymbol{1}^\perp$. A most simple example is to take a partition $V = B \cup C$ and define $\boldsymbol{x} := b\boldsymbol{\chi}_C - c\boldsymbol{\chi}_B$, where $b := |B|$

and $c := |C| = n-b$. (In other words, $x_i = -c$ if $i \in B$ and $x_i = b$ otherwise.) Clearly, $\boldsymbol{x} \cdot \mathbf{1} = 0$. Thus, by (64),

$$\boldsymbol{x} \cdot \mathsf{A}\boldsymbol{x} \leq \lambda \|\boldsymbol{x}\|^2 = \lambda(bc^2 + cb^2) = \lambda bcn. \tag{65}$$

On the other hand, $\boldsymbol{x} \cdot \mathsf{A}\boldsymbol{x}$ equals

$$\sum_{i \in V} x_i \Big(\sum_{j \in \Gamma(i)} x_j \Big) = 2 \sum_{\{i,j\} \in E} x_i x_j = 2 \left(c^2 e(G[B]) + b^2 e(G[C]) - e_{BC} bc \right),$$

where e_{BC} denotes the number of edges between B and C. We can simplify the last expression by using the relations $2e(G[B]) + e_{BC} = bd$ and $2e(G[C]) + e_{BC} = cd$, where we just count the edges incident to B and C. We obtain

$$\boldsymbol{x} \cdot \mathsf{A}\boldsymbol{x} = (bd - e_{BC})c^2 + (cd - e_{BC})b^2 - 2e_{BC}\, bc = cbdn - e_{BC}\, n^2. \tag{66}$$

Putting (65) and (66) together we obtain the following result.

Lemma 67 *For any partition $V = B \cup C$ we have*

$$e_{BC} \geq \frac{(d-\lambda)\,|B|\,|C|}{n}. \ \blacksquare \tag{67}$$

In particular we see that if $\lambda < d$, then $e_{BC} > 0$ for any $B \cup C = V$, that is, G is connected. In fact, the converse holds as well.

Lemma 68 *The graph G is connected if and only if $\lambda < d$.*

Proof. Let G be connected. Suppose on the contrary that there is $\boldsymbol{v}$ not collinear to $\mathbf{1}$ with $\mathsf{A}\boldsymbol{v} = d\boldsymbol{v}$. It is easy to see that there is $c \in \mathbb{R}$ such that $\boldsymbol{u} := \mathbf{1} + c\boldsymbol{v} \geq \mathbf{0}$ and, for some indexes $i, j \in [n]$, we have $u_j > 0$ and $u_i = 0$. We have

$$\mathsf{A}\boldsymbol{u} = \mathsf{A}\mathbf{1} + c\mathsf{A}\boldsymbol{v} = d\mathbf{1} + cd\boldsymbol{v} = d\boldsymbol{u},$$

that is, $\mathsf{A}^l\boldsymbol{u} = d^l\boldsymbol{u}$. It follows in particular that $(\mathsf{A}^l)_{i,j} = 0$ for any l, that is, there is no walk joining i to j (see Section 14.3). This is clearly a contradiction. $\blacksquare$

There are yet more interesting consequences of Lemma 67 to exploit. As G is d-regular, we get an estimate for the number of neighbours of any $B \subset V$:

$$d\,|\Gamma(B) \setminus B| \geq e_{B,\overline{B}} \geq \frac{(d-\lambda)b(n-b)}{n}. \tag{68}$$

This ‘expansion’ property (each set having many neighbours) plays an important role in various areas of Discrete Mathematics. There are a few, slightly inequivalent, quantitative defitions. We will use the following version: our (d-regular, order-n) graph G is called an (n,d,c)-*expander* if for any $B \subset V$ with $|B| \leq \frac{1}{2}|V|$ we have $|\Gamma(B) \setminus B| \geq c\,|B|$. Observe that some restriction of the form $|B| \leq \frac{1}{2}|V|$ is necessary to prevent cases like $B = V$.

In this terminology, (68) clearly implies the following.

Theorem 69 *G is an $(n, d, \frac{d-\lambda}{2d})$-expander.* ▮

Remark. On the other hand the expansion property of G implies some bounds on λ but we will not need this result (whose proof is more difficult).

Of course, if G is connected, it is an (n, d, c)-expander for some $c > 0$. For applications we typically need an *explicitly* constructed family of (n, d, c)-expanders where $n \to \infty$ while d and c are some fixed constants. While its existence can be routinely established via probabilistic arguments, it is very difficult to come up with an explicit construction, especially to prove that it has the expansion property.

We will describe an application of expanders to derandomisation of algorithms and then present an explicit construction. We need some preliminaries.

14.3 Walks on Expanders

An *l-walk* on G is a sequence $(v_0, \dots, v_l)$ of $l + 1$ vertices of V with $v_{i-1}v_i \in E$ for $i \in [l]$. Note that a walk can visit a vertex more than once. We have chosen this definition because $(\mathsf{A}^l)_{u,v}$ is then precisely the number of l-walks from u to v and we can apply linear algebra to analyse walks.

Let us try to prove something along the lines that a random l-walk is likely to visit $C \subset V$ if $|C|$ and l are large.

The obvious definition of a *random l-walk* is to choose $v_0 \in V$ and, inductively, $v_i \in \Gamma(v_{i-1})$, each vertex having the uniform distribution on the available choices. It is clear that this procedure gives a uniformly distributed l-walk on the set of all nd^l l-walks.

Thus to estimate the probability that a random l-walk avoids C we have to count the number of l-walks on $G[B]$, where $B := \overline{C}$, say $B = [b]$. This number clearly equals $\mathbf{1} \cdot \mathsf{B}^l \mathbf{1}$, where B is the adjacency $b \times b$-matrix of $G[B]$.

By Lemma 65 we know that B has b real eigenvalues $\beta_1 \ge \dots \ge \beta_b$ with corresponding orthonormal basis $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_b)$. Of course, in this basis B (and B^l) looks simplest. Express $\mathbf{1} \in \mathbb{R}^b$ in this basis: $\mathbf{1} = \sum_{i=1}^b a_i \boldsymbol{b}_i$. Then we have

$$\mathbf{1} \cdot \mathsf{B}^l \mathbf{1} = \Big(\sum_{i=1}^b a_i \boldsymbol{b}_i \Big) \cdot \Big(\sum_{i=1}^b \beta_i^l a_i \boldsymbol{b}_i \Big) = \sum_{i=1}^b \beta_i^l a_i^2. \tag{69}$$

To estimate this sum we can use the following result.

Lemma 70 (Perron's Theorem) *Let M be a real symmetric $n \times n$-matrix with eigenvalues $\mu_1 \ge \dots \ge \mu_n$. If $\mathsf{M} \ge 0$ (i.e. each entry is non-negative), then $|\mu_n| \le \mu_1$.*

Proof. Let $\boldsymbol{m}_1, \dots, \boldsymbol{m}_n$ be corresponding orthonormal eigenvectors, cf. Lemma 65. For $\boldsymbol{x} \in \mathbb{R}^n$ let $|\boldsymbol{x}| := (|x_1|, \dots, |x_n|)$. (Do not confuse $|\boldsymbol{x}|$ with the norm $\|\boldsymbol{x}\|$.) We clearly have

$$|\mu_n|\,|\boldsymbol{m}_n| = |\mu_n\,\boldsymbol{m}_n| = |\mathsf{M}\boldsymbol{m}_n| \le \mathsf{M}\,|\boldsymbol{m}_n|.$$

Using this and the Raleigh–Ritz Formula (Lemma 66) we obtain

$$|\mu_n|\,|\boldsymbol{m}_n| \cdot |\boldsymbol{m}_n| \le \mathsf{M}\,|\boldsymbol{m}_n| \cdot |\boldsymbol{m}_n| \le \mu_1\,|\boldsymbol{m}_n| \cdot |\boldsymbol{m}_n|,$$

which implies the claim as $\boldsymbol{m}_n \neq \mathbf{0}$. ▮

Perron's Theorem and (69) imply that

$$\mathbf{1} \cdot \mathsf{B}^l \mathbf{1} \leq \beta_1^l \sum_{i=1}^{b} a_i^2 = \beta_1^l \, \|\mathbf{1}\|^2 = \beta_1^l b. \tag{70}$$

Remark. We could have avoided using Perron's Theorem altogether. As $\mathbf{1} \cdot \mathsf{B}^l \mathbf{1}$ is non-negative (it counts some walks), we conclude that $a_i = 0$ whenever $|\beta_i| > \beta_1$ (analyse (69) for odd $l \to \infty$). Now (70) follows.

Our task now is to bound β_1 from above. By the Raleigh–Ritz Formula we have to find an upper bound on $\boldsymbol{x} \cdot \mathsf{B}\boldsymbol{x}$ over all $\boldsymbol{x} \in \mathbb{R}^b$ with $\|\boldsymbol{x}\| = 1$. Clearly, $\boldsymbol{x} \cdot \mathsf{B}\boldsymbol{x} = \boldsymbol{y} \cdot \mathsf{A}\boldsymbol{y}$ where $\boldsymbol{y} \in \mathbb{R}^n$ is obtained from $\boldsymbol{x}$ by adding $n-b$ zeros. Of course, $\boldsymbol{y} \cdot \mathsf{A}\boldsymbol{y} \leq d$, but this does not imply anything more than the number of l-walks on $G[B]$ is at most bd^l, which is trivial to see directly.

Let us use the now familiar technique of writing $\boldsymbol{y} = \sum_{i=1}^n c_i \boldsymbol{f}_i$ when $\boldsymbol{y} \cdot \mathsf{A}\boldsymbol{y} = \sum_{i=1}^n \lambda_i c_i^2$. If we knew that $c_1 = 0$, for example, then we would be able to deduce that $\boldsymbol{y} \cdot \mathsf{A}\boldsymbol{y} \leq \lambda$. Perhaps $c_1 = 0$ is too much to hope, but even an upper bound on $|c_1|$ will be useful.

Clearly, $c_1 = \boldsymbol{y} \cdot \boldsymbol{f}_1 = \sum_{i=1}^b x_i \frac{1}{\sqrt{n}}$; also we know that $\sum_{i=1}^b x_i^2 = 1$. We can now apply the Cauchy–Schwarz Inequality (Lemma 34):

$$c_1^2 = \Big(\sum_{i=1}^{b} x_i \, \frac{1}{\sqrt{n}} \Big)^2 \leq \Big(\sum_{i=1}^{b} x_i^2 \Big) \times b \, \frac{1}{n} = \frac{b}{n}.$$

Now it sounds obvious that $\sum_{i=1}^n \lambda_i c_i^2$ is maximised when $c_1^2 = b/n$, $c_2^2 = 1 - b/n$ and all other c_i's are zero. It is easy to find a formal proof:

$$\sum_{i=1}^{n} \lambda_i c_i^2 \leq d c_1^2 + \lambda \sum_{i=2}^{b} c_i^2 = (d-\lambda) c_1^2 + \lambda \leq (d-\lambda) \, \frac{b}{n} + \lambda = \frac{bd + (n-b)\lambda}{n}.$$

We are through! Putting everything together we obtain:

Theorem 71 *Let $C \subset V$, $|C| = cn$. Then the number of l-walks avoiding C is at most $(1-c)n((1-c)d + c\lambda)^l$. In particular, the probability that a random l-walk avoids C is at most $(1-c)(1-c+\frac{c\lambda}{d})^l$.* ▮

As we see, the bound on probability is exponential in l, so it tends to 0 very quickly when $l \to \infty$ (and $\lambda < d$ are fixed).

14.4 Derandomisation

Here is one application of expanders. Suppose that our aim is to find some $w \in W$ where W is some subset of $[n]$ about which we know nothings except that, for example, $|W| \geq \frac{n}{2}$. The deterministic algorithm of taking elements of $[n]$ one by one until we hit W may require as many as $\lceil n/2 \rceil$ steps. However, if n is large, this sounds extremely unlikely.

Suppose we set some δ and will be satisfied is our algorithm finds $w \in W$ with probability at least $1 - \delta$.

Algorithm 1: take l independent, uniformly distributed elements of $[n]$. Clearly, the probability of missing W is at most 2^{-l}, so we take $l = \lceil \log_2 \frac{1}{\delta} \rceil$ points. To generate a random point we need m random bits, where we assume for simplicity that $n = 2^m$. Thus our algorithm uses $\Theta(\log \frac{1}{\delta}\ \log n)$ random bits.

However, for various reasons it is preferred to reduce the required number b of random bits. For example, if b gets really small, then it may be feasible to run the deterministic algorithm which checks all possible 2^b binary strings. A general procedure, when we reduce the number of random bits by modifying a probabilistic algorithm, is called *derandomisation*.

So here is our **Algorithm 2**. Suppose we have an explicitly constructed expander G on $[n]$. Take a random l-walk, where l is chosen such that the probability of missing C is at most δ. By Theorem 71 $l = \Theta(\log \frac{1}{\delta})$ suffices (assuming that d and λ are fixed; or rather that d is fixed while $\lambda < (1-\varepsilon)d$ for some constant $\varepsilon > 0$). We need $\Theta(\log n)$ random bits to generate an initial vertex of a walk and $k = O(1)$ random bits per each step (assume $d = 2^k$). Thus the required number of random bits is $\Theta(\log \frac{1}{\delta} + \log n)$, which is smaller than that in Algorithm 1 when $\delta \to 0$.

Although the above setting (to find $w \in W$) seems somewhat artificial, it does appear in real-life computational problems. A notable example of the above situation is primality testing for which so far no deterministic algorithm is known.

Notes

The fact that a graph with a small second eigenvalue has some expansion properties was independently discovered by Tanner [Tan84] and by Alon and Milman [AM85]. Alon and Milman [AM85] proved a stronger version of Theorem 69: G is an $(n, d, \frac{2(d-\lambda)}{3d-2\lambda})$-expander.

The converse correspondence was established by Alon [Alo86b] who showed that $\lambda \le d - \frac{c^2}{4+2c^2}$ for any (n, d, c)-expander.

The first results along the lines of Theorem 71 were discovered by Ajtal, Komlós ans Szemeredi [AKS87].

15 Explicit Constant-Degree Expanders

As we have already mentioned, it is very difficult to find an explicit construction of an (n, d, c)-expander for fixed d, c and large n. Margulis [Mar73] was first to give such a construction. Other constructions appeared as well: [GG81, AM85, LPS88, Mar88], to name a few.

Here we present a recent explicit construction of expanders due to Reingold, Vadham and Wigderson [RVW02] where also a new graph product, the zig-zag product, is introduced. Besides giving a neat way of constructing expanders, this product may have interesting applications to other combinatorial problems.

15.1 Operations on Graphs

Let us call G an *(n, d, μ)-graph* if it has n vertices, is d-regular (so its largest eigenvalue $\lambda_1 = d$), and $|\lambda_i| \le \mu d$ for any $i \in [2, n]$. By Theorem 69 the expansion coefficient of G is at least $(1 - \mu)/2$. So, it is enough to construct, for some fixed d and $\mu < 1$, (n, d, μ)-graphs for infinitely many n.

We will need two operations on graphs. The first one will correspond to squaring the adjacency matrix of a graph. Unfortunately, the class of $(0, 1)$-matrices is not invariant under matrix multiplication. Therefore, *in this section we will allow our graphs to have multiple edges and loops (and even multiple loops).*

To define the adjacency matrix A of a such graph G, let A_{ii} be the number of loops at the vertex i and let A_{ij} be the number of edges between i and j. We extend the other definitions of Section 14 to these settings so that the algebraic properties of A preserve their combinatorial meaning. (Usually, the modifications are the obvious ones.) For example, we want d to be an eigenvalue of A for a d-regular graph. So we say that G is *d-regular* if the sum of the entries of A in each row (or column) equals d; hence, the degree of a vertex v is defined as the number of loops at v plus the number of edges incident to v taken with their multiplicities.

The *square* $G^{\wedge}$ of a graph G is the graph (on the same vertex set) that corresponds to the square of the adjacency matrix of G. In other words, edges in $G^{\wedge}$ correspond to 2-walks in G. (The standard notation is G^2, but we tried to avoid the collision with the definition of G^n from Section 10 and chose '$\wedge$' as this symbol resembles a 2-walk.)

Lemma 72 *If G is an (n, d, μ)-graph, then $G^{\wedge}$ is an (n, d^2, μ^2)-graph.*

Proof. G is d^2-regular as there are precisely d^2 2-walks in G starting at any one vertex. If $\mathsf{A}\boldsymbol{x} = \alpha\boldsymbol{x}$, then $\mathsf{A}^2\boldsymbol{x} = \alpha^2\boldsymbol{x}$, so the eigenvalues of A get squared. The lemma clearly follows. ▮

It is does not come as a surprise that taking 2-walks improves on the expansion property. Unfortunately, we pay a price for this: the degree increases.

Another operation which we will need is the *zig-zag product* $G \textcircled{z} H$ of two graphs. We will give its definition and prove the following lemma a bit later.

Lemma 73 *If G is an (n, d, μ)-graph and H is an (d, l, γ)-graph, then $G \textcircled{z} H$ is an $(nd, l^2, \mu + \gamma + \gamma^2)$-graph.*

In our applications H will be a small fixed graph. The remarkable property of the zig-zag product is that the resulting graph has small degree (even if the degree of G is big) while its expansion coefficient is not much worse than those of the factors. Combining the square and zig-zag products one can get a family of expanders as follows.

First, one argues that for some d there exists a $(d^4, d, 1/5)$-graph H. Two different explicit examples of such H are given in [RVW02]. Due to the lack of time, we do not describe H.

Given H define $G_1 = H^\wedge$ and $G_{i+1} = G_i^\wedge \textcircled{z} H$.

Theorem 74 *For every i, G_i is a $(d^{4i}, d^2, 2/5)$-graph.*

Proof. We use induction with the case $i = 1$ being clearly true. If the claim is true for some i, then $G_i^\wedge$ is a $(d^{4i}, d^4, 4/25)$-graph by Lemma 72 and $G_i^\wedge \textcircled{z} H$ is a $(d^{4i+4}, d^2, 2/5)$-graph by Lemma 73, as required. (Note that $\frac{4}{25} + \frac{1}{5} + (\frac{1}{5})^2 = \frac{2}{5}$.) ▮

Now by Theorem 69 we conclude that $(G_i)_{i \in \mathbb{N}}$ is a family of expanders!

Corollary 75 *For every i, G_i is a $(d^{4i}, d^2, 3/10)$-expander.* ▮

15.2 Zig-Zag Product

Here we define the zig-zag product of an (n, d, μ)-graph G and a (d, l, γ)-graph H. Note that the number of vertices of H must be equal to the regularity degree of G because in order to apply the zig-zag product we should have, for every $v \in V(G)$, a bijective labelling of the edges of G incident to v by the vertices of $V(H)$. An edge of G is allowed to have two distinct labels at its endpoints. (But a loop has only one label.) Although $G \textcircled{z} H$ depends on the choice of labelling, the properties stated in Lemma 73 hold for any labelling.

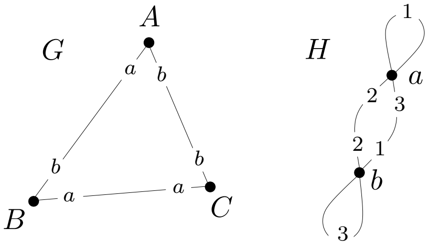

Fig. 4: Labelled Graphs

Now we have to decide how to represent such a labelling. Following [RVW02] a labelling of a d-regular graph $G = (V, E)$ by labels from a d-set D can be specified by a *rotation map* as follows. For $(v, i) \in V \times D$ define $\mathrm{Rot}(v, i) := (u, j)$ where $u \in \Gamma(v)$ is the vertex which we reach if we start at v and move along the label-i edge E; j is the label of E at u (which may be different from i, the label of E at v). Clearly, Rot is a bijection of $V \times D$ onto itself such that $\mathrm{Rot} \circ \mathrm{Rot}$ is the identity map.

We illustrate this with Figure 4:

$$\mathrm{Rot}_G : \begin{bmatrix} (A,a) & \longleftrightarrow & (B,b) \\ (A,b) & \longleftrightarrow & (C,b) \\ (B,a) & \longleftrightarrow & (C,a) \end{bmatrix} \qquad \mathrm{Rot}_H : \begin{bmatrix} (a,1) & \longleftrightarrow & (a,1) \\ (a,2) & \longleftrightarrow & (b,2) \\ (a,3) & \longleftrightarrow & (b,1) \\ (b,3) & \longleftrightarrow & (b,3) \end{bmatrix}$$

Now we are ready to define the zig-zag product of graphs G and H. Let $G = (V, E)$ be an (n, d, μ)-graph and $H = (D, E')$ be a (d, l, γ)-graph. We assume that G comes with a D-labelling, i.e., we have a rotation map $\mathrm{Rot}_G : V \times D \to V \times D$. Assume also that H is labelled by an l-set L via $\mathrm{Rot}_H : D \times L \to D \times L$. Then their *zig-zag product* $G \circledz H$ is an l^2-regular graph on $V \times D$ with the L^2-labelling whose rotation map can be computed as follows. Given $((v, k), (i, j)) \in (V \times D) \times (L \times L)$:

1. Let $(k', i') := \mathrm{Rot}_H(k, i)$.
2. Let $(w, l') := \mathrm{Rot}_G(v, k')$.
3. Let $(l, j') := \mathrm{Rot}_H(l', j)$.
4. The result is $((w, l), (j', i')) \in (V \times D) \times (L \times L)$.

Given the rotation map on $(V \times D) \times (L \times L)$ it is obvious how to define the corresponding l^2-regular graph.

It will be more instructive to illustate this with Figure 4, where $V = \{A, B, C\}$, $D = \{a, b\}$, and $L = \{1, 2, 3\}$. For example, take vertex (B, a) and the label $(3, 1)$. First we move in H along the edge number 3 and get to (B, b). Now we move in G using b and obtain (A, a). Finally, we take the label 1 in H (which is a loop at a) and arrive back at (A, a). Thus the edge labelled $(3, 1)$ at at the vertex (B, a) leads to (A, a) at which end its label is $(1, 1)$. (*Check that if you start at* (A, a) *and move along the edge labelled* $(1, 1)$ *you arrive at* (B, a). *Compute a few more cases.*) We do not draw the whole graph $G \circledz H$ which has little resemblance to G or H anyway. (Well, this is the point of the definition: expanders should look 'random-like'.)

15.3 Properties of the Zig-Zag Product

Here we prove Lemma 73. Let M be the adjacency matrix of $G \circledz H$. We have only to show that the second eigenvalue of M is at most $(\mu + \gamma + \gamma^2)l^2$. By (64) it is enough to to show that for any $\boldsymbol{\alpha} \in \mathbb{R}^{nd}$ with $\boldsymbol{\alpha} \cdot \mathbf{1}_{nd} = 0$ we have

$$\boldsymbol{\alpha} \cdot \mathsf{M}\boldsymbol{\alpha} \le (\mu + \gamma + \gamma^2)\|\boldsymbol{\alpha}\|^2. \tag{71}$$

We view vectors in $\mathbb{R}^{nd}$ as indexed by $V \times D$, $V := [n]$, $D := [d]$. Thus, for $\boldsymbol{x} \in \mathbb{R}^{nd}$ and $v \in V$, we define $\boldsymbol{x}_v \in \mathbb{R}^d$ by $(\boldsymbol{x}_v)_k = x_{(v,k)}$. Also, define a linear map $C : \mathbb{R}^{nd} \to \mathbb{R}^n$ by

$$C(\boldsymbol{x})_v := \sum_{k=1}^{d} x_{(v,k)}, \quad \boldsymbol{x} \in \mathbb{R}^{nd}.$$

On the other hand, for $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^d$, define the *tensor product* $\boldsymbol{x} \otimes \boldsymbol{y} \in \mathbb{R}^{nd}$ by $(\boldsymbol{x} \otimes \boldsymbol{y})_{(v,k)} = x_v y_k$, $(v, k) \in V \times D$. For $\boldsymbol{\alpha} \in \mathbb{R}^{nd}$ we have $\boldsymbol{\alpha} = \sum_{v \in V} \boldsymbol{e}_v \otimes \boldsymbol{\alpha}_v$, where $\boldsymbol{e}_v$ is the standard v-th basis vector in $\mathbb{R}^n$.

The *tensor product* of an $n \times n$-matrix N and a $d \times d$-matrix D is the $(nd) \times (nd)$-matrix $\mathsf{N} \otimes \mathsf{D}$ defined by

$$(\mathsf{N} \otimes \mathsf{D})_{(u,j),(v,k)} = \mathsf{N}_{u,v}\, \mathsf{D}_{j,k}.$$

It is routine to verify that $(\mathsf{N} \otimes \mathsf{D})\,(\boldsymbol{x} \otimes \boldsymbol{y}) = (\mathsf{N}\boldsymbol{x}) \otimes (\mathsf{D}\boldsymbol{y})$. (*Check this.*)

So, let $\boldsymbol{\alpha} \in \mathbb{R}^{nd}$. For $v \in V$ let $\boldsymbol{\alpha}_v^{\|} := \frac{1}{d}\,(\boldsymbol{\alpha}_v \cdot \mathbf{1}_d)\,\mathbf{1}_d$ and $\boldsymbol{\alpha}_v^{\perp} := \boldsymbol{\alpha}_v - \boldsymbol{\alpha}_v^{\|}$. We have $\boldsymbol{\alpha}_v^{\perp} \cdot \mathbf{1} = 0$; in other words, we represented $\boldsymbol{\alpha}_v = \boldsymbol{\alpha}_v^{\|} + \boldsymbol{\alpha}_v^{\perp}$ as sum of two vectors, one parallel and the other perpendicular to $\boldsymbol{\alpha}_v$. This gives us the following representation

$$\boldsymbol{\alpha} = \sum_{v \in V} \boldsymbol{e}_v \otimes \boldsymbol{\alpha}_v = \sum_{v \in V} \boldsymbol{e}_v \otimes \boldsymbol{\alpha}_v^{\|} + \sum_{v \in V} \boldsymbol{e}_v \otimes \boldsymbol{\alpha}_v^{\perp} =: \boldsymbol{\alpha}^{\|} + \boldsymbol{\alpha}^{\perp}.$$

Let A and B the be adjacency matrices of G and H correspondently. We now decompose M into the product of three matrices, corresponding to the three steps in the definition of $G \textcircled{z} H$.

Define $\tilde{\mathsf{B}} := \mathsf{I}_n \otimes \mathsf{B}$. What is $(\tilde{\mathsf{B}})_{(u,j),(v,k)}$? First of all, it is zero unless $u = v$ in which case it equals $\mathsf{B}_{j,k}$. So, $\tilde{\mathsf{B}}$ encodes the l-regular graph on $N \times D$ which is used in the first and third step of the zig-zag construction. In other words,

$$\mathsf{M} = \tilde{\mathsf{B}}\tilde{\mathsf{A}}\tilde{\mathsf{B}},$$

where $\tilde{\mathsf{A}}$ is the permutation matrix on $N \times D$ corresponding to the second step. By the symmetry of B we have

$$\boldsymbol{\alpha} \cdot \mathsf{M}\boldsymbol{\alpha} = \boldsymbol{\alpha} \cdot \tilde{\mathsf{B}}\tilde{\mathsf{A}}\tilde{\mathsf{B}}\boldsymbol{\alpha} = \tilde{\mathsf{B}}\boldsymbol{\alpha} \cdot \tilde{\mathsf{A}}\tilde{\mathsf{B}}\boldsymbol{\alpha}. \tag{72}$$

Observe that

$$\boldsymbol{\alpha}^{\|} = \sum_{v \in V} \boldsymbol{e}_v \otimes \boldsymbol{\alpha}_v^{\|} = \frac{1}{d} \sum_{v \in V} (\boldsymbol{\alpha}_v \cdot \mathbf{1}_d)\,\boldsymbol{e}_v \otimes \mathbf{1}_d = \frac{1}{d}\,C(\boldsymbol{\alpha}) \otimes \mathbf{1}_d.$$

We have

$$\tilde{\mathsf{B}}\boldsymbol{\alpha}^{\|} = (\mathsf{I}_n \otimes \mathsf{B})\,(\tfrac{1}{d}\,C(\boldsymbol{\alpha}) \otimes \mathbf{1}_d) = \frac{1}{d}\,C(\boldsymbol{\alpha}) \otimes (l\mathbf{1}_d) = l\,\boldsymbol{\alpha}^{\|}.$$

Substituting this into (72) we obtain

$$\boldsymbol{\alpha} \cdot \mathsf{M}\boldsymbol{\alpha} = (l\,\boldsymbol{\alpha}^{\|} + \tilde{\mathsf{B}}\boldsymbol{\alpha}^{\perp}) \cdot \tilde{\mathsf{A}}(l\,\boldsymbol{\alpha}^{\|} + \tilde{\mathsf{B}}\boldsymbol{\alpha}^{\perp}).$$

Using the Cauchy–Schwarz Inequality (Lemma 34) and the fact that $\|\tilde{\mathsf{A}}\boldsymbol{x}\| = \|\boldsymbol{x}\|$ for any $\boldsymbol{x} \in \mathbb{R}^{nd}$ (because $\tilde{\mathsf{A}}$ is a permutation matrix), we obtain

$$|\boldsymbol{\alpha} \cdot \mathsf{M}\boldsymbol{\alpha}| \le l^2\,|\boldsymbol{\alpha}^{\|} \cdot \tilde{\mathsf{A}}\boldsymbol{\alpha}^{\|}| + 2l\,\|\boldsymbol{\alpha}^{\|}\|\,\|\tilde{\mathsf{B}}\boldsymbol{\alpha}^{\perp}\| + \|\tilde{\mathsf{B}}\boldsymbol{\alpha}^{\perp}\|^2.$$

Let $p := \|\boldsymbol{\alpha}^{\|}\|$ and $q := \|\boldsymbol{\alpha}^{\perp}\|$. We have $p^2 + q^2 = \|\boldsymbol{\alpha}\|^2$, so $pq \le \frac{1}{2}\,\|\boldsymbol{\alpha}\|^2$. The Claims 1 and 2 (below) imply that

$$l^{-2}\,|\boldsymbol{\alpha} \cdot \mathsf{M}\boldsymbol{\alpha}| \le \mu p^2 + 2\gamma pq + \gamma^2 q^2 \le (\mu + \gamma + \gamma^2)\,\|\boldsymbol{\alpha}\|^2, \tag{73}$$

which implies Lemma 73. It remains to prove the two claims.

Claim 1 $\|\tilde{\mathsf{B}}\boldsymbol{\alpha}^{\perp}\| \le \gamma l\,\|\boldsymbol{\alpha}^{\perp}\|$.

Proof of Claim. We have

$$\tilde{\mathsf{B}}\boldsymbol{\alpha}^{\perp} = (\mathsf{I}_n \otimes \mathsf{B})\Big(\sum_{v \in V} \boldsymbol{e}_v \otimes \boldsymbol{\alpha}_v^{\perp}\Big) = \sum_{v \in V} \boldsymbol{e}_v \otimes (\mathsf{B}\boldsymbol{\alpha}_v^{\perp}).$$

The vector $\boldsymbol{\alpha}_v^\perp$ is orthogonal to $\mathbf{1}_d$ which collinear to the eigenvector $\boldsymbol{b}_1$ of B corresponding to the largest eingevalue $\beta_1 = l$. Hence, $\boldsymbol{\alpha}_v^\perp$ lies in the linear span of the remaining eigenvectors, say $\boldsymbol{\alpha}_v^\perp = \sum_{i=2}^d c_i \boldsymbol{b}_i$. But then

$$\|\tilde{\mathsf{B}}\boldsymbol{\alpha}_v^\perp\|^2 = \left\| \sum_{i=2}^d \beta_i c_i \boldsymbol{b}_i \right\|^2 = \sum_{i=2}^d \beta_i^2 c_i^2 \le \gamma^2 l^2 \sum_{i=2}^d c_i^2 = \gamma^2 l^2 \, \|\boldsymbol{\alpha}_v^\perp\|^2.$$

Now,

$$\|\tilde{\mathsf{B}}\boldsymbol{\alpha}^\perp\|^2 = \sum_{v\in V} \|\mathsf{B}\boldsymbol{\alpha}_v^\perp\|^2 \le \gamma^2 l^2 \sum_{v \in V} \|\boldsymbol{\alpha}_v^\perp\|^2 = \gamma^2 l^2 \, \|\boldsymbol{\alpha}^\perp\|^2,$$

which proves the claim. ▮

Claim 2 $|\boldsymbol{\alpha}^\| \cdot \tilde{\mathsf{A}}\boldsymbol{\alpha}^\|| \le \mu \, \|\boldsymbol{\alpha}^\|\|^2$.

Proof of Claim. Observe that $C\tilde{\mathsf{A}}(\boldsymbol{e}_v \otimes \mathbf{1}_d) = \mathsf{A}\boldsymbol{e}_v$, for any $v \in V$, because its value on $u \in V$ equals the number of edges between u and v:

$$(C\tilde{\mathsf{A}}(\boldsymbol{e}_v \otimes \mathbf{1}_d))_u = \sum_{k=1}^d \tilde{\mathsf{A}}(\boldsymbol{e}_v \otimes \mathbf{1}_d)_{u,k} = \sum_{k=1}^d \sum_{(w,i)\in N\times D} \tilde{\mathsf{A}}_{(w,i),(u,k)} \, (\boldsymbol{e}_v \otimes \mathbf{1}_d)_{(w,i)} = \sum_{k=1}^d \sum_{i=1}^d \tilde{\mathsf{A}}_{(v,i),(u,k)}$$

Because the $\boldsymbol{e}_v$'s form a basis, this is true for any $\boldsymbol{\beta} \in \mathbb{R}^n$:

$$C(\tilde{\mathsf{A}}(\boldsymbol{\beta} \otimes \mathbf{1}_d)) = \mathsf{A}\boldsymbol{\beta}.$$

Recalling that $\boldsymbol{\alpha}^\| = \frac{1}{d} C(\boldsymbol{\alpha}) \otimes \mathbf{1}_d$ we obtain

$$\boldsymbol{\alpha}^\| \cdot \tilde{\mathsf{A}}\boldsymbol{\alpha}^\| = \frac{1}{d} (C(\boldsymbol{\alpha}) \otimes \mathbf{1}_d) \cdot \tilde{\mathsf{A}}\boldsymbol{\alpha}^\| = \frac{1}{d} C(\boldsymbol{\alpha}) \cdot C\tilde{\mathsf{A}}\boldsymbol{\alpha}^\| = \frac{1}{d^2} C(\boldsymbol{\alpha}) \cdot \mathsf{A}\, C(\boldsymbol{\alpha}).$$

By the definition, $\boldsymbol{\alpha} \cdot \mathbf{1}_{nd} = 0$, implying that $C(\boldsymbol{\alpha})$ is orthogonal to $\mathbf{1}_n$ which is an eigenvector of A corresponding to the largest eigenvalue d. Hence (cf. the proof of Claim 1) we have

$$C(\boldsymbol{\alpha}) \cdot \mathsf{A}\, C(\boldsymbol{\alpha}) \le \mu d \, \|C(\boldsymbol{\alpha})\|^2 = \mu \, \|(C(\boldsymbol{\alpha})) \otimes \mathbf{1}_d\|^2 = \mu d^2 \, \|\boldsymbol{\alpha}^\|\|^2.$$

Putting all together we prove the claim. ▮

Execises and Further Reading

[RVW02]: A very well and clearly written paper, where the zig-zag product is introduced.

16 Example Sheet

16.1 Basic Problems

Problem 76 Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ be vectors with n rational entries. Prove that the $\boldsymbol{v}_i$'s are linearly independent over $\mathbb{Q}$ if and only if they are linearly independent over $\mathbb{R}$.

Problem 77 Let $v_1, \ldots, v_m$ be $(0,1)$-vectors, each of length n. Prove that if these vectors are linearly independent over $\mathbb{F}_p$ then they are linearly independent over $\mathbb{Q}$.

Show that the converse is not in general true but if a set of $(0,1)$-vectors is linearly independent over $\mathbb{Q}$, then it is linearly independent over $\mathbb{F}_q$ for every sufficiently large prime p.

Problem 78 Given an integer d construct a square $(0,1)$-matrix with determinant d.

Problem 79 Construct a set of $(0,1)$-vectors, linearly independent over $\mathbb{Q}$ and $\mathbb{F}_3$ but linearly dependent over $\mathbb{F}_2$ and $\mathbb{F}_5$.

Problem 80 (Graham & Pollak [GP71]) Prove that any connected graph G admits a $\{0, 1, *\}$-addressing.

Problem 81 (Alon [Alo98]) Let p and q be two primes, put $s = pq - 1$ and let $r > s$ be an integers. Let G be the graph whose vertices are the subsets of $[r]$ of cardinality s, where two are adjacent iff the cardinality of their intersection is -1 modulo p. Prove the following bounds on the Shannon capacity: $\Theta(G) \leq \sum_{i=0}^{p-1} \binom{r}{i}$ and $\Theta(\overline{G}) \leq \sum_{i=0}^{q-1} \binom{r}{i}$.

Problem 82 Recall that $\mathcal{H}_{2,q}$ consists of all 1-dimensional affine subspaces of $(\mathbb{F}_q)^2$. Show that $\mathcal{H}_{2,q}$ is a $(q^2, q, 1)$-design.

Problem 83 For any $c > 0$ there is n_0 such that there is no $(n, 2, c)$-expander with $n > n_0$.

Problem 84 Let G be an (n, d, c)-expander. Show that any two points $u, v \in V(G)$ are connected by a path of length at most $\frac{2 \log n}{\log(1+c)}$. *[Hint: Build the desired path from both ends.]*

Problem 85 Let G be a d-regular connected graph. Show that $-d$ is its eigenvalue if and only if G is bipartite.

16.2 Harder Problems

The following problems may require a considerable amount of thought or work.

Problem 86 Suppose there are m red clubs $R_1, \ldots, R_m$ and m blue blubs $B_1, \ldots, B_m$ involving n people. Assume that these clubs satisfy the following rules.

1. $|R_i \cap B_i|$ is odd for every i;

2. $|R_i \cap B_j|$ is even for any $i \neq j$.

Prove that $m \leq n$.

Problem 87 Weaken Assumption 2 of the preceding exercise to $|R_i \cap B_j|$ is even for $1 \leq i < j \leq m$.

Problem 88 (Babai & Frankl [BF80]) Let p be a prime and $k \geq 1$. Let $A_1, \ldots, A_m \subset [n]$ satisfy the following: the sizes of the A_i's are not divislble by p^k but their pairwise intersection are divisible by p^k. Prove that $m \leq n$.

Problem 89 Let G be a graph on $\binom{[n]}{k}$ with two k-sets being adjacent iff their intersection is non-empty. Prove that the Shannon capacity $\Theta(G) \leq n/k$.

Problem 90 (Haemers [Hae79]) Let p be a prime not dividing k. Let G be the graph on $\binom{[n]}{k}$ with two vertices A and B being adjacent iff $|A \cap B| \not\equiv 0 \pmod{p}$. Then the Shannon capacity $\Theta(G) \leq n$.

Problem 91 (Raigorodski [Rai99]) Let K be a subset of the unit sphere $S^{n-1} \subset \mathbb{R}^{n-1}$ with $\operatorname{diam}(K) \geq \sqrt{2}$. Then there exists $L \subset S^n$ with $\operatorname{diam}(L) \geq \sqrt{2}$ and $f(L) \geq f(K) + 1$, where $f(X)$ is the minimal number of parts of smaller diameter partitioning X. *[Hint: We can additionally require that $|L| = |K| + 1$.]*

Problem 92 (Alon [Alo00]) Let p be an odd prime and let $k < p$. Applying Combinatorial Nullstellensatz prove that for any two k-tuples $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{F}_p^{(k)}$ there is a permutation π of $[k]$ such that the sums $a_i + b_{\pi(i)}$ (in $\mathbb{F}_p$) are pairwise distinct.

Problem 93 (Mixing Time of Random Walks) Let $v_0, v_1, \ldots, v_l$ be a random walk on an (n, d, μ)-graph (d and $\mu < 1$ are fixed). Prove that there is constant $C = C(d, \mu) > 1$ such that for any vertex v

$$|\mathrm{Prob}(v_l = v) - 1/n| < C^{-l}.$$

Problem 94 (Pinsker [Pin73]) Using probabilistic methods show that there exists a constant $c > 0$ for which, for all sufficiently large even n, there exists an $(n, 3, c)$-expander. *[Hint: Take random 3-regular graph of order n, see Bollobás [Bol01, Chapter 2.4], and estimate the probability that it is an expander (messy).]*

16.3 Yet Harder!

Tough questions...

Problem 95 (Kézdy and Snevily [KS02]) Show that the conjecture in Problem 101 is true when $2k \leq n+1$. *[Hint: Having done Problem 92, move to the field $\mathbb{R}$, modifying your polynomial in a certain way.]*

Problem 96 Construct an $(n, 3, 1)$-design for any $n \equiv 3 \pmod{6}$. (In fact, an $(n, 3, 1)$-design exists iff $n \equiv 1$ or $3 \pmod{6}$, see [LW92, Chapter 19] for constructions.)

Problem 97 (Wilson [Wil74]) Let H be an arbitrary graph. Show that for any v_0 there is $v > v_0$ such that the complete graph K_v decomposes into edge-disjoint copies of H.

16.4 Some Important Research Problems

Problem 98 We know the following bounds on the chromatic number of $\mathbb{R}^2$: $4 \leq \chi(\mathbb{R}^2) \leq 7$. Improve.

Problem 99 (Alon [Alo98]) Let G be a random graph on n vertices (in which each edge is included in G independently of others and with probability 1/2). It is conjectured that there is $C > 0$ such that

$$\mathrm{Prob}(\Theta(G) > C \log n) \to 0 \quad \text{as } n \to \infty.$$

Problem 100 It is conjectured that if an $(n^2+n+1, n+1, 1)$-design exists, then n is a prime power. The first open case is $n = 12$.

Problem 101 (Snevily, see [KS02]) Snevily conjectured that for any integers $k < n$ and any sequence $a_1, \ldots, a_k$ of not necessarily distinct elements of $\mathbb{Z}_n$, there exists a permutation π of $[k]$ such that the elements $a_{\pi(i)} + i$ are all distinct modulo n. Compare with Problems 92 and 95; note that n is not required to be prime. (If true this conjecture would imply very important results on latin squares.)

16.5 Application: Primality Testing

Here we describe a primality testing algorithm, where expanders can be used for derandomisation. To define an integer n we need $\Theta(\log n)$ bits in input. An algorithm is called *polynomial-time* if its running time is bounded by some polynomial in $\log n$.

Let n be an integer of the form $n = 4t+3$. (For $n = 4t+1$ things are slightly more complicated.)

Problem 102 If n is a prime power, then for any $a \in \mathbb{Z}_n$ the equation $x^2 = a$ has either none or 2 roots. If n is not a prime power, then any such equation has either none or at least 4 roots.

Problem 103 (Euler Criteria) If n is a prime and $a \in \mathbb{F}_n$, then a is a square if and only if $a^{\frac{n-1}{2}} = 1$. *[Hint: Use Lemma 57.]* Moreover, if a is a square, then a square root of a can be computed by the formula $x = a^{t+1}$.

Problem 104 Choose a random $x \in \mathbb{Z}_n$. Let $a = x^2$ and compute $y = a^{t+1}$. If $y^2 \neq a$ or if $y \neq \pm x$, return "composite" otherwise return "prime". Show that if n is not a prime power, then this procedure outputs "composite" with probability at least 1/2.

Problem 105 There is a (deterministic) polynomial-time algorithm checking if $n = m^k$ for some $k \geq 2$. *[Hint: There are at most $\log_2 n$ possible values of k.]*

Problem 106 Devise a polynomial-time algorithm which detects if an integer $n = 4t+3$ is a prime with the probability of mistake at most $\frac{1}{n^{1000000}}$. Decribe how a family of expanders can be used to derandomise your algorithm.

References

[AF93] N. Alon and Z. Füredi. Covering the cube by affine hyperplanes. *Europ. J. Combin.*, 14:79–83, 1993. ⇑ 45, 46

[AFK84] N. Alon, S. Friedland, and G. Kalai. Regular subgraphs on almost regular graphs. *J. Combin. Theory* (B), 37:79–91, 1984. ⇑ 43

[AKS87] M. Ajtal, J. Komlós, and E. Szemeredi. Deterministic simulation in LOGSPACE. In *Proc. 19-th Annual ACM STOC*, pages 132–140, New York, 1987. ⇑ 63

[Alo86a] N. Alon. Decomposition of the complete r-graph into complete r-partite r-graphs. *Graphs Combin.*, 2:95–100, 1986. ⇑ 7

[Alo86b] N. Alon. Eigenvalues, geometric expanders, sorting in rounds and Ramsey theory. *Combinatorica*, 6:207–219, 1986. ⇑ 63

[Alo95] N. Alon. Tools from higher algebra. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, pages 1749–1784. Elsevier Science B.V., 1995. ⇑ 48

[Alo98] N. Alon. The Shannon capacity of a union. *Combinatorica*, 18:301–310, 1998. ⇑ 29, 32, 69, 71

[Alo99] N. Alon. Combinatorial Nullstellensatz. *Combin. Prob. Computing*, 8:7–29, 1999. ⇑ 38, 40, 41

[Alo00] N. Alon. Additive latin transversals. *Israel J. Math.*, 117:125–130, 2000. ⇑ 70

[AM85] N. Alon and V. D. Milman. Isoperimetric inequalities for graphs and superconcentrators. *J. Combin. Theory* (B), 1985:73–88, 1985. ⇑ 63, 64

[ANR95] N. Alon, M. B. Nathanson, and I. Z. Ruzsa. Adding dinstinct congruence classes modulo a prime. *Amer. Math. Monthly*, 102:250–255, 1995. ⇑ 41

[ANR96] N. Alon, M. B. Nathanson, and I. Z. Ruzsa. The polynomial method and restricted sums of congruence classes. *J. Number Theory*, 56:404–417, 1996. ⇑ 41

[AZ98] M. Aigner and G. M. Ziegler. *Proofs from The Book.* Springer, 1998. ⇑ 18, 29, 37

[BF80] L. Babai and P. Frankl. On set intersections. *J. Combin. Theory* (A), 28:103–105, 1980. ⇑ 70

[BF92] L. Babai and P. Frankl. *Linear Algebra Methods in Combinatorics.* Dept. Comput. Sc., Univ. Chicago, 1992. Preliminary Version 2. Moore Library classmark: QA184.B33 1992. ⇑ 7, 8, 10, 12

[BG85] V. G. Bolyanski and I. T. Gohberg. *Results and Problems in Combinatorial Geometry.* Cambridge Univ. Press, 1985. ⇑ 20

[Blo93] A. Blokhuis. On the Sperner capacity of the cyclic triangle. *J. Algebraic Comb.*, 2:123–124, 1993. ⇑ 38

[Bol78] B. Bollobás. *Extremal Graph Theory.* Academic Press, London, 1978. ⇑ 45

[Bol95] B. Bollobás. Extremal graph theory. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, pages 1231–1292. Elsevier Science B.V., 1995. ⇑ 51

[Bol01] B. Bollobás. *Ramdom Graphs.* Cambridge Univ. Press, 2d edition, 2001. ⇑ 70

[Bor33] K. Borsuk. Drei Sätze über die n-dimensionale euklisische Sphäre. *Fund. Math.*, 20:177–190, 1933. ⇑ 19

[Bro66] W. G. Brown. On graphs that do not contain a Thomsen graph. *Can. Math. Bull.*, 9:281–289, 1966. ⇑ 50

[BS78] A. E. Brouwer and A. Schrijver. The blocking number of an affine space. *J. Combin. Theory* (A), 24:251–253, 1978. ⇑ 47, 48

[CFG$^+$93] A. R. Calderbank, P. Frankl, R. L. Graham, W.-C. W. Li, and L. A. Shepp. The Sperner capacity of linear and nonlinear codes for the cyclic triangle. *J. Algebraic Comb.*, 2:31–48, 1993. ⇑ 38

[Chv83] V. Chvatál. *Linear Programming.* Freeman, New York, 1983. ⇑ 32

[CS99] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups.* Springer Verlag, 3d edition, 1999. ⇑ 23

[Dav35] H. Davenport. On the addition of residue classes. *J. Lond. Math. Soc.*, 10:30–32, 1935. ⇑ 40, 41

[DFS83] M. Deza, P. Frankl, and N. M. Singhi. On functions of strength t. *Combinatorica*, 3:331–339, 1983. ⇑ 12

[Erd32] P. Erdős. Beweis eines Satzes von Tchebyshef. *Acta Sci. Math. (Szeged)*, 5:194–198, 1932. ⇑ 18

[Erd47] P. Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, 53:292–294, 1947. ⇑ 25

[Erd67] P. Erdős. On bipartite subgraphs of graphs. *Mat. Lapok*, 18:283–288, 1967. In Hungarian. ⇑ 45

[Erd81] P. Erdős. On the combinatorial problems I would most like to see solved. *Combinatorica*, 1:25–42, 1981. ⇑ 45

[ERS66] P. Erdős, A. Rényi, and V. T. Sós. On a problem in the theory of graphs. *Stud. Sci. Math. Hungar.*, 1:215–235, 1966. ⇑ 50

[Fis40] R. A. Fisher. An examination of the possible different solutions of a problem in incomplete blocks. *Ann. Eugenics (London)*, 10:52–75, 1940. ⇑ 8, 12

[Fra77] P. Frankl. A constructive lower bound for Ramsey numbers. *Ars Combinatoria*, 3:297–302, 1977. ⇑ 27

[Für96a] Z. Füredi. On the number of edges of quadrilaterial-free graphs. *J. Combin. Theory* (B), 68:1–6, 1996. ⇑ 50

[Für96b] Z. Füredi. An upper bound on Zarankiewicz' problem. *Combin. Prob. Computing*, 5:29–33, 1996. ⇑ 51

[FW81] P. Frankl and R. M. Wilson. Intersection theorems with geometric consequences. *Combinatorica*, 1:357–368, 1981. ⇑ 12, 16, 27

[GG81] O. Gabber and Z. Galil. Explicit constructions of linear-sized superconcentrators. *J. Computer Syst. Sci.*, 22:407–420, 1981. ⇑ 64

[GKV92] L. Gargano, J. Körner, and U. Vaccaro. Qualitative independence and Sperner problems for directed graphs. *J. Combin. Theory* (A), 61:173–192, 1992. ⇑ 37

[GP71] R. L. Graham and H. O. Pollak. On the addressing problem for loop switching. *Bell Syst. Tech. J.*, 50:2495–2515, 1971. ⇑ 7, 69

[Gra94] R. L. Graham. Recent trends in Euclidean Ramsey theory. *Discrete Math.*, 136:119–127, 1994. ⇑ 18

[GS01] V. Golmusz and B. Sudakov. On k-wise set-intersections and k-wise hamming-distances, 2001. Manuscript. ⇑ 10

[Had44] H. Hadwiger. Ein überdeckingssatz für den Euklidischen Raum. *Potrugaliae Math.*, 4:140–144, 1944. ⇑ 18

[Hae79] W. Haemers. On some problems of Lovász concerning the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25:231–232, 1979. ⇑ 70

[Hin02] A. Hinrichs. Spherical codes and Borsuk's conjecture. *Discrete Math.*, 243:253–256, 2002. ⇑ 20, 22

[HJ85] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, 1985. ⇑ 58

[HR02] A. Hinrichs and C. Richter. New sets with large Borsuk numbers. Manuscript, 2002. ⇑ 20

[Jam77] R. E. Jamison. Covering finite fields with cosets of subspaces. *J. Combin. Theory* (A), 22:253–266, 1977. ⇑ 47, 48

[KK93] J. Kahn and G. Kalai. A counterexample to Borsuk's conjecture. *Bull. Amer. Math. Soc.*, 29:60–62, 1993. ⇑ 19

[KRS96] J. Kollár, L. Rónyai, and T. Szabó. Norm graphs and bipartite Turán numbers. *Combinatorica*, 16:399–406, 1996. ⇑ 51

[KS02] A. E. Kézdy and H. S. Snevily. Distinct sums modulo n and tree embeddings. Submitted, 2002. ⇑ 70, 71

[KST54] P. Kővari, V. T. Sös, and P. Turán. On a problem of K. Zarankiewicz. *Colloq. Math.*, 3:50–57, 1954. ⇑ 49, 51

[Lan74] E. Landau. *Handbuch der Lehre von der Verteilung der Primzahlen*. New York: Chelsea, 3 edition, 1974. ⇑ 57

[Lin93] B. Lindstrom. Another theorem on families of sets. *Ars Combinatoria*, 35:123–124, 1993. ⇑ 4, 5

[Lov79] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Information Theory*, 25:1–7, 1979. ⇑ 34, 37

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8:261–277, 1988. ⇑ 64

[LR72] D. G. Larman and C. A. Rogers. The realization of distances within sets in Euclidean space. *Mathematika*, 19:1–24, 1972. ⇑ 18

[LW92] J. H. van Lint and R. M. Wilson. *A Course in Combinatorics*. Cambridge Univ. Press, 1992. ⇑ 7, 8, 70

[LY92] S. Lou and Q. Yao. A Chebyshev's type of prime number theorem in a short interval II. *Hardy-Ranamujan J.*, 15:1–33, 1992. ⇑ 19

[Mar73] G. A. Margulis. Explicit constructions of concentrators. *Probl. Information Transmission*, 9:325–332, 1973. ⇑ 64

[Mar88] G. A. Margulis. Explicit group-theoretical constructions of combinatorial schemes and their applications to the design of expanders and superconcentrators. *Probl. Information Transmission*, 24:39–46, 1988. ⇑ 64

[Nag72] Zs. Nagy. A certain constructive estimate of the Ramsey number. *Matematikai Lapok*, 23:301–302, 1972. ⇑ 25

[Pin73] M. Pinsker. On the complexity of a concentrator. In *7-th Internat. Teletraffic Conf.*, pages 318/1–318/4. Stokholm, 1973. ⇑ 70

[Pri86] D. Pritikin. Applying a proof of Tverberg to complete bipartite decompositions of digraphs and multigraphs. *J. Graph Theory*, 10:197–201, 1986. ⇑ 7

[PRS95] L. Pyber, V. Rödl, and E. Szemerédi. Dense graphs without 3-regular subgraphs. *J. Combin. Theory* (B), 53:41–54, 1995. ⇑ 45

[Pyb85] L. Pyber. Regular subgraphs of dense graphs. *Combinatorica*, 6:347–349, 1985. ⇑ 45

[Rai99] A. M. Raigorodski. On a bound in the Borsuk problem. *Russian Math. Surveys*, 54:453–454, 1999. ⇑ 70

[Rai01] A. M. Raigorodski. The Borsuk problem and the chromatic numbers of some metric spaces. *Russian Math. Surveys*, 56:103–139, 2001. ⇑ 18, 20

[RCW75] D. K. Ray-Chaudhuri and R. M. Wilson. On t-designs. *Osaka J. Math*, 12:737–744, 1975. ⇑ 12

[RVW02] O. Reingold, S. P. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. To appear in *Annals Math.*, 2002. ⇑ 64, 65, 68

[Sch88] O. Schramm. Illuminating sets of constant width. *Mathematika*, 35:180–189, 1988. ⇑ 20

[Sha56] C. E. Shannon. The zero-error capacity of a noisy channel. *IRE Trans. Information Theory*, 3:3–15, 1956. ⇑ 27, 29, 34

[Tan84] R. M. Tanner. Explicit construction of concerntrators from generalized N-gons. *SIAM J. Alg. Disc. Meth.*, 5:287–293, 1984. ⇑ 63

[Tve82] H. Tverberg. On the decomposition of K_n into complete bipartite graphs. *J. Graph Theory*, 6:493–494, 1982. ⇑ 7

[Wil72a] R. M. Wilson. An existence theory for pairwise balanced designs, I: Composition theorems and morphisms. *J. Combin. Theory*, 13:220–245, 1972. ⇑ 51

[Wil72b] R. M. Wilson. An existence theory for pairwise balanced designs, II. *J. Combin. Theory*, 13:246–273, 1972. ⇑ 51

[Wil74] R. M. Wilson. Constructions ans uses of pairwise balanced designs. In M. Hall Jr. and J. H. van Lint, editors, *Combinatorics, Part 1*, volume 55 of *Math. Centre Tracts*, pages 18–41. Mathematisch Centrum, Amsterdam, 1974. ⇑ 57, 70

[Wil75] R. M. Wilson. An existence theory for pairwise balanced designs, III. *J. Combin. Theory*, 18:71–79, 1975. ⇑ 51

[Win83] P. M. Winkler. Proof of the squashed cube conjecture. *Combinatorica*, 3:135–139, 1983. ⇑ 5, 7

[Zar51] K. Zarankiewicz. Problem p. 101. *Colloq. Math.*, 2:116–131, 1951. ⇑ 49

Index