5.1.3 Almost independent sample spaces

In this section we relax the independence requirement of sample spaces that each bit-vector should appear with the exact same probability, and allow that they appear with *roughly* the same probability, up to an error of ϵ.

Definition: A sample space $S \subseteq \mathbb{F}_2^N$ is called *ϵ-close to independent* if for any vector $a \in \{0,1\}^N$, we have

$$\left| \mathbb{P}[s = a] - \frac{1}{2^N} \right| \leq \epsilon.$$

Note that being 0-close to independent is equivalent to being independent. For our application we of course need the extension of the definition to almost d-wise independence.

Definition: The sample space $S \subseteq \{0,1\}^N$ is called *ϵ-close to d-wise independent* if for any subset $J \in \binom{[N]}{d}$ of the coordinates, the sample space $S|_J \subseteq \{0,1\}^d$ is ϵ-close to independent, that is, for any vector $a \in \{0,1\}^d$, we have

$$\left| \mathbb{P}[s|_J = a] - \frac{1}{2^d} \right| \leq \epsilon.$$

In the main result of this section we show that allowing a bit of imperfectness in d-wise independence enables one to reduce the size of the sample space from the polynomial of Theorem 5.1 to a *polylogarithmic* function of N. More precisely, we will construct sample spaces that are ϵ-close to d-wise independent, and their size is only polylogarithmic in N and polynomial in their imperfectness measurements, i.e., in d and $\frac{1}{\epsilon}$.

Theorem 5.4 (Naor and Naor) *Let $N = 2^t$ with $t \in \mathbb{N}$, let $d \geq 1$ be an odd integer, and let $\epsilon > 0$. Then there is a sample space $R \subseteq \{0,1\}^N$ of size at most*

$$\frac{2\left(t\frac{d-1}{2} + 1\right)^2}{\epsilon^2} \sim \frac{d^2}{2\epsilon^2} \log^2 N,$$

which is ϵ-close to d-wise independent.

The main idea of the proof is to take the d-wise independent sample space $S(BG_m)$ we constructed in the last section and somehow reduce its size. The columns of the matrix BG_m are the 2^m linear combinations of the columns of B: each column of G_m is responsible for one. The plan is to take only an appropriately selected few of these, such that the d-wise independence is not ruined too much. It is tempting to select a few columns of G_m randomly, but we must remain sober and resist—we want an explicit construction. We will instead construct an m-dimensional sample space $S(Q)$ of quadratic size $p \sim \frac{m^2}{\epsilon^2}$, as opposed to 2^m, which is ϵ-close to independent. Then we will show that taking only this p linear combinations of the columns of B, as opposed to all 2^m, is enough to maintain the d-wise independence with an error ϵ. Namely, we will show that $S(BQ)$ is ϵ-close to d-wise independent.

In the next two subsections we work out the ingredients of this plan and then the proof of Theorem 5.4 will follow easily.

Linear tests

The property of being ϵ-close to d-wise independent is quite difficult to work with, let alone to show directly. Hence we develop a more effective way to establish it, a way which is much more apt to our plan to create our sample space via linear combinations.

If a sample space $S = S(M) \subseteq \{0,1\}^N$ is independent then we have seen in Exercise 5.6 that it is 1-independent, that is, the number of 0 and 1 in every row of M is the same. In the next exercise we generalize this to give yet another characterization of independent sample spaces.

Exercise 5.10 *A sample space $S \subseteq \{0,1\}^N$ is independent if and only if for every vector $a \in \{0,1\}^N \setminus \{0^N\}$,*

$$\mathbb{P}[s \cdot a = 0] = \mathbb{P}[s \cdot a = 1].$$

Here 0^N denotes the vector of length N having only 0 coordinates, while $s \cdot a = \sum_{i=1}^N s_i a_i$ represents the usual dot-product of vectors over $\mathbb{F}_2$.

The exercise involves $2^N - 1$ "linear test"s one performs on the sample space to verify its independence, each of which should produce a halving of the sample space. We will relax on the perfectness of these halvings to approach the concept of almost independence.

Definition: A sample space $S \subseteq \{0,1\}^N$ is called ϵ-*unbiased with respect to linear tests* if for any $a \in \{0,1\}^N \setminus \{0^N\}$,

$$|\mathbb{P}[s \cdot a = 0] - \mathbb{P}[s \cdot a = 1]| \leq \epsilon.$$

Note that S is ϵ-unbiased with respect to linear tests if and only if for any $a \in \{0,1\}^N \setminus \{0^N\}$, the 1-dimensional sample space $\{s \cdot a : s \in S\} \subseteq \{0,1\}$ is $\epsilon/2$-close to independent.

The equivalence of being ϵ-unbiased with respect to linear tests and being ϵ-close to independent, which was established in Exercise 5.10 for $\epsilon = 0$, does not hold for $\epsilon > 0$. This is shown in the next exercise.

Exercise 5.11 *Show that if a sample space $S \subseteq \{0,1\}^N$ is ϵ-close to independent then it is also $\epsilon 2^N$-unbiased with respect to linear tests. Construct a sample space that shows the statement being best possible (for all sensible values of the parameters N and ϵ).*

The following lemma states that one direction of Exercise 5.10 remains valid even if $\epsilon > 0$ and thus establishes linear tests as a method to prove ϵ-closeness to independence.

Lemma 5.4.1 (Vazirani) *Let $S \subseteq \{0,1\}^N$ be a sample space that is ϵ-unbiased with respect to linear tests. Then S is ϵ-close to independent.*

Proof. We introduce the probability distribution function p on $\mathbb{Z}_2^N$ by setting $p(x) := \mathbb{P}[s = x]$ for the probability of a vector $x \in \{0,1\}^N$ in the sample space S. We need to show that this function $p : \mathbb{Z}_2^N \to \mathbb{C}$ does not deviate more than ϵ from its average $\frac{1}{|\mathbb{Z}_2^N|} \sum_{x \in \mathbb{Z}_2^N} p(x) = \frac{1}{2^N}$. We make use of the basic properties of the discrete Fourier transform of p on the group $\langle H, + \rangle = \langle \mathbb{Z}_2^N, + \rangle$. In particular, applying Proposition A.35 we obtain that

$$\left| p(a) - \frac{1}{2^N} \right| \leq \Phi(p) |\mathbb{Z}_2^N| \tag{5.6}$$

for every $a \in \mathbb{Z}_2^N$, where

$$\Phi(p) = \max\{|\langle \chi, p \rangle| : \chi \in \widehat{\mathbb{Z}_2^N}, \chi \neq \chi_0\}$$

is the largest absolute value among the non-principal Fourier coeffiecients of p.

Recall that the characters of $\mathbb{Z}_2^N$ are defined by $\chi_b(a) = (-1)^{b \cdot a}$, for every $b \in \mathbb{Z}_2^N$ and $a \in \mathbb{Z}_2^N$. The key observation is that the probability difference between the occurrence of 0 and 1 upon making a linear test with some test vector $b \in \mathbb{Z}_2^N$ is precisely the (non-normalized) Fourier coefficient of p corresponding to character $\overline{\chi}_b$. The test vector $b = 0^N$ corresponds then to the principal character χ_0 and therefore our assumption on S implies that all, but the principal, non-normalized Fourier coefficients of p are at most ϵ. And that, via (5.6), implies that S is ϵ-close to independent.

Indeed, for any $b \in \mathbb{Z}_2^N \setminus \{0^N\}$ we have

$$\begin{aligned} \epsilon \geq \mathbb{P}[s \cdot b = 0] - \mathbb{P}[s \cdot b = 1] &= \sum_{\substack{a \in \mathbb{Z}_2^N \\ a \cdot b = 0}} \mathbb{P}[s = a] - \sum_{\substack{a \in \mathbb{Z}_2^N \\ a \cdot b = 1}} \mathbb{P}[s = a] \\ &= \sum_{a \in \mathbb{Z}_2^N} (-1)^{a \cdot b} p(a) = \sum_{a \in \mathbb{Z}_2^N} \chi_b(a) p(a) = |\mathbb{Z}_2^N| \langle \overline{\chi}_b, p \rangle, \end{aligned}$$

and the lemma is proved. □

Our eventual goal is the construction of a small sample space that is ϵ-close to d-wise independent. The next lemma describes an easy way to combine the generator matrix L of a d-wise independent linear sample space $S(LG_m)$ with a sample space $S(Q)$ which is ϵ-close to independent and obtain a sample space that is ϵ-close to d-wise independent.

We plan to use Lemma 5.4.1 to each d-dimensional restriction of the constructed sample space, in order to establish that they are all ϵ-close to independent, and then conclude that the sample space itself is ϵ-close to d-wise independent.

Lemma 5.4.2 (Naor and Naor) *Let B be an $(N \times m)$-matrix over $\mathbb{F}_2$ such that any d-rows are linearly independent and let Q be a $(m \times p)$-matrix over $\mathbb{F}_2$ such that the sample space $S(Q) \subseteq \{0,1\}^m$ of size p is ϵ-unbiased with respect to linear tests. Then the sample space $S(BQ) \subseteq \{0,1\}^N$ of size p is ϵ-close to d-wise independent.*

Proof. We have to check that for every subset $J \subseteq [N]$ of size d the rows, the restriction of the sample space $S(BQ)$ to these d rows is ϵ-close to independent. To this end we would like to use Lemma 5.4.1 and hence verify that the d-dimensional restriction $S(BQ)|_J = S(B|_J Q)$ is ϵ-unbiased with respect to linear tests. Let $a \in \{0,1\}^d \setminus \{0^d\}$ be a d-dimensional test vector. Since $a^T(B|_J Q) = (a^T B|_J)Q$, the linear test of $S(B|_J Q)$ with test vector a and the linear test of $S(Q)$ with test vector $a^T B_J$ are the same. Note that the test vector $a^T B|_J \in \{0,1\}^m$ is non-zero, since $a \neq 0^d$ and any d rows of B are linearly independent. By assumption the sample space $S(Q)$ is ϵ-unbiased with respect to linear tests, so the probability of 0 and the probability of 1 differ by at most ϵ in the 1-dimensional sample space $S((a^T B|_J)Q)$, and hence also in $S(a^T(B|_J Q))$. □

The first ingredient of Lemma 5.4.2, a matrix B with any d of its rows being linearly independent, was constructed in Theorem 5.1. In the next subsection we construct the second ingredient: a small sample space $S(Q) \subseteq \{0,1\}^m$ which is ϵ-unbiased with respect to linear tests.

Almost independent sample spaces via the quadratic character

A field has two operations: addition and multiplication. There are many examples of the vague phenomenon that being a regular structure in some additive sense and being a regular structure in some multiplicative sense are mutually exclusive, or at least very limited in size. As a simplest example one can think of are arithmetic and geometric progressions: the largest set that is both is of size two. Recall the Paley graph we discussed in the first section of this part: for a prime $p \cong 1 \pmod 4$, the Paley graph P_p was just the Cayley graph defined on the additive group of $\mathbb{F}_p$ by the generating set $S = QR_p$ of the quadratic residues. That is, the Paley graph is defined on the additive structure of a field by a generating set that is multiplicative in nature. While we know, modulo the Generalized Rieman Hypothesis, that the Paley graph is not a perfect source of randomness, we also know that it might be a pretty good imitation, in fact waay better than anything we are able to construct today.

We use this intuition, the quadratic residues being a pseudorandom random subset within the additive structure of the finite field $\mathbb{F}_p$. Recall that the value of the quadratic character $\varrho_p : \mathbb{F}_p^* \to \{-1,1\}$ is 1 for quadratic residues and -1 for quadratic non-residues. Convert these values to bits: let $r(x) = 0$ for quadratic residues and 1 for non-residues. Expressed with a formula, we have $\varrho_p(x) = (-1)^{r(x)}$. In other words, $r = \mathbb{1}_{NQR_p}$ is just the characteristic function of the quadratic non-residues modulo p. Imagine these values in the cyclic additive order of the field, that is $r(1), r(2), (3), \ldots, r(p-1), r(0)$. For 0 let us just extend r arbitrarily, say let us have $r(0) = 1$.

Our sample space will consits of the p bit-vectors that form an interval of length m in this cyclic ordering of length p. Since intervals are very regular additive structures, we hope that the multiplicatively defined values will be quite random. Naturally, we will have to assume that m is small enough compared to p. Formally, we define a $(m \times p)$-matrix $Q = Q_m^p$, whose colums $q^{(x)} \in \mathbb{F}_2^m$ are labeled by elements $x \in \mathbb{F}_p$ and $q_i^{(x)} := r(x+i)$ for every $i = 1, 2, \ldots, m$.

Proposition 5.5 (Alon, Goldreich, Hastad, and Peralta) *For every $m \leq \sqrt{p}$, the sample space $S(Q_m^p) = \{q^{(x)} : x \in \mathbb{F}_p\}$ is $\frac{m}{\sqrt{p}}$-unbiased with respect to linear tests.*

Note that for this proposition to have any power, we better have $m \leq \epsilon\sqrt{p}$ with some $\epsilon < 1$; the smaller the ϵ, the better.

Proof. Let us fix our "linear tester" $a \in \{0,1\}^m$. As we saw in the proof of Lemma 5.4.1, the probability difference in the definition of almost independence can be expressed as follows.

$$\mathbb{P}_{x\in\mathbb{F}_p}\left[q^{(x)} \cdot a = 0\right] - \mathbb{P}_{x\in\mathbb{F}_p}\left[q^{(x)} \cdot a = 1\right] = \sum_{\substack{b\in\mathbb{F}_p \\ q^{(b)}\cdot a=0}} \mathbb{P}_{x\in\mathbb{F}_p}[x=b] - \sum_{\substack{b\in\mathbb{F}_p \\ r^{(b)}\cdot a=1}} \mathbb{P}_{x\in\mathbb{F}_p}[x=b]$$
$$= \frac{1}{p}\sum_{b\in\mathbb{F}_p}(-1)^{q^{(b)}\cdot a} = \frac{1}{p}\sum_{b\in\mathbb{F}_p}\prod_{i=1}^{m}(-1)^{r(b+i)a_i}$$

We want to replace each product $\prod_{i=1}^m(-1)^{r(b+i)a_i}$ with $\prod_{i=1}^m(\varrho_p(b+i))^{a_i} = \varrho_p(\prod_{i=1}^m(b+i)^{a_i})$ and then use Weil's Theorem for the quadratic character ϱ_p and the polynomial $f(x) = \prod_{i=1}^m(x+i)^{a_i}$. We can certainly do this whenever $b \in \mathbb{F}_p$ is not in the interval $[p-b, p-1]$, because then $b+i \neq 0$ and hence $(-1)^{r(b+i)} = \varrho_p(b+i)$ for every $i = 1, 2, \ldots, m$, by the definition of r. These are most of the $b \in \mathbb{F}_p$; only those in the interval $[p-m, p-1]$ of length $m \leq \sqrt{p}$ are problematic. Whenever $b \in [p-m, p-1]$ the corresponding product contains a factor $(-1)^{r(b+i)a_i}$ with $b+i = 0$. Considering that for the sake of Weil's Theorem $\varrho_p(0)$ is defined to be 0, whenever $b+i=0$, we have that $\left|(-1)^{r(b+i)a_i} - \varrho_p(b+i)^{a_i}\right|$ is either 0 or 1 (depending on whether $a_i = 0$ or 1).

$$\left|\mathbb{P}_{x\in\mathbb{F}_p}\left[q^{(x)} \cdot a = 0\right] - \mathbb{P}_{x\in\mathbb{F}_p}\left[q^{(x)} \cdot a = 1\right]\right| = \left|\frac{1}{p}\sum_{b\in\mathbb{F}_p}\prod_{i=1}^{m}(-1)^{r(b+i)a_i}\right| \leq$$
$$\leq \frac{1}{p}\left|\sum_{b\in\mathbb{F}_p}\prod_{i=1}^{m}(\varrho_p(b+i))^{a_i}\right| + \frac{1}{p}\sum_{b\in[p-m,p-1]}\left|\prod_{i=1}^{m}(-1)^{r(b+i)a_i} - \prod_{i=1}^{m}(\varrho_p(b+i))^{a_i}\right|$$
$$\leq \frac{1}{p}\left|\sum_{b\in\mathbb{F}_p}\varrho_p\left(\prod_{i=1}^{m}(b+i)^{a_i}\right)\right| + \frac{m}{p}$$
$$\leq \frac{m-1}{\sqrt{p}} + \frac{m}{p} \leq \frac{m}{\sqrt{p}}.$$

In the last step we used $m \leq \sqrt{p}$ and in the next to last we applied Weil's theorem for the quadratic character ϱ_p which has order 2 and the polynomial $f(x) = \prod_{i=1}^{m}(x+i)^{a_i}$ which has at most m distinct roots and is certainly not a square. □

Proof. We can now put together the proof of Theorem 5.4 by using Lemma 5.4.2 with the almost independent independent sample space of Proposition 5.5 and the d-wise independent linear sample space of Theorem 5.1.

Let $m = t^{\frac{d-1}{2}} + 1$. First construct the $(N \times m)$-matrix B with the property that any d rows are linearly independent. Then, after choosing a prime p between $\frac{\left(t^{\frac{d-1}{2}}+1\right)^2}{\epsilon^2}$ and its double, construct the above sample space $S(Q^p_m) \subseteq \{0,1\}^p$ with $m = t^{\frac{d-1}{2}} + 1$. By Proposition 5.5 $S(Q^p_m)$ is $\frac{m}{\sqrt{p}}$-unbiased with respect to linear tests. Note that $\frac{m}{\sqrt{p}} \leq \epsilon$. According to Lemma 5.4.2 the sample space $S(BQ^p_m)$ of size p is ϵ-close to d-wise independent. This concludes the proof of the theorem. □

Better Ramsey-graphs

Let us now try to use our sample spaces from Theorem 5.4 which are ϵ-close to d-wise independent in our quest for explicit Ramsey graphs.

We could again take our constructive sample space, like we did earlier, interpret it as graphs on $N = \binom{n}{2}$ vertices and take the Abbott product of all of them. But in fact, since our sample space is now so small, we can do even better. We can return to the original idea of the Abbott construction: checking for the perfect "starter graph" with brute force in polynomial time, and then taking the Abbott-powers of this single graph with good Ramsey properties.

Our goal in this section is the construction of a graph G on n vertices in time polynomial in n with $\omega(G), \alpha(G) < 2^{\sqrt{\log n}\log\log n}$. In the solitude of your home you should check that it is equivalent to constructing a k-Ramsey graph with $k^{\frac{\log k}{(\log\log k)^2}}$ vertices. Recall that this will be a further improvement in the line of our constructive lower bounds: the exponent of the order of the construction in Subsection 5.1.2 was twice iterated logarithm and now we have essentially a single $\log k$ in the exponent (disregarding the lower order $(\log\log k)^2$ in the denominator.)

This construction was apparently folklore, here we follow the description of Baraz. Let us fix the number of vertices n and define the integer $k = 2^{\sqrt{\log n}}$.

We aim to find our "good starter" graph H on k vertices. What is special about the selection of k. We will see that on the one hand we can choose a sample space of size polynomial in n of graphs on k vertices, which γ-close to d-wise independent, where γ is small enough and d is large enough. On the other hand it is possible to check for small enough cliques on k vertices.

We take a sample space $S \subseteq \{0,1\}^{\binom{k}{2}}$ which is $2^{-5\log^2 k}$-close to being $4.5\log^2 k$-wise independent. By Theorem 5.4 there exists such a space of size

$$\approx 20.25 \log^4 k 2^{10\log^2 k} \log^2 \binom{k}{2} = k^{O(\log k)} = n^{O(1)},$$

i.e., the size of this space is polynomial in n.

Note that for any graph on k vertices we can check, just by brute force, whether the clique number and the independence number of it is at most $3\log k$, in time

$$\binom{k}{3\log k}\binom{3\log k}{2} = k^{O(\log k)} = n^{O(1)},$$

which is polynomial in n.

Hence in polynomial time we can check for each member of this sample space, whether its clique number and independence number is at most $3\log k$. What is left to prove is that in S, there exist such a graph. This follows from the almost d-wise independence of the space. Fix a subset L of the vertices, $|L| = 3\log k$. Then by the almost $4.5\log^2 k$-independence of the sample space,

$$\mathbb{P}[L \text{ is a clique or independent set}] = 2\cdot\left(\frac{1}{2^{\binom{|L|}{2}}} + \frac{1}{2^{5\log^2 k}}\right) \ll \frac{1}{\binom{k}{3\log k}}.$$

That is *there exists* a member of the sample space S for which no set of size $3\log k$ is a clique or an independent set. This will be our starter graph H and our brute force search will certainly find it in polynomial time in n.

Now take the $\sqrt{\log n}$th Abbott-power of H. This product graph has $k^{\sqrt{\log n}} = n$ vertices and can be constructed in time polynomial in n (Exercise 5.5). By (5.1), its clique number and independence numnber is certainly upper bounded by

$$(3\log k)^{\sqrt{\log n}} = (3\sqrt{\log n})^{\sqrt{\log n}} = 2^{\sqrt{\log n}\log\log n\left(\frac{1}{2}+\frac{\log 3}{\log\log n}\right)}.$$

The extra factor in the exponent is smaller than 1 for large enough n and hence we are done.

Note however a crucial difference in the construction of this last example and the rest of this section. When we took the Abbott-product of all graphs in Subsection 5.1.1 or when we took the Abbott-product of all graphs from the d-wise independent sample space in Subsection 5.1.2 we were not only constructing the adjacancy matrix of the graph in time polynomial in n, but were able to answer a query quickly requesting the adjacency relation of two particular vertices. The query containing the labels of the two vertices in question has only $2\log n$ bits, so one would possibly want to have the answer in time polynomial in $\log n$. This is possible in those constructions as the Abbott-product is efficient in this sense (see Exercise 5.5).

In our current construction one needs to construct the starter graph first before being able to answer adjacency queries about its Abbott-power and this alone already takes time polynomial in n, and not in $\log n$. This explains the following definition. A construction of a graph on n vertices is called *strongly explicit* if adjacency queries can be answered in time polynomial in $\log n$. A construction of a graph on n vertices is called *weakly explicit* if the adjacency matrix of the graph can be constructed in time polynomial in n.

One could suspect that the "definition" or rather "feeling" of explicit construction a'la Erdős would be closer to the definition of the strongly explicit one above. However there is an important "philosophical" distinction. At the time Erdős posed his question about a “constructive” lower bound for the Ramsey function, the computer scientific notion of “efficient” was just about to be created. Erdős refused to pay his award to Peter Frankl, who came up to him with the Abbott product construction. His refusal was not based on a mathematically founded argument, rather by a philosophically motivated one. “I don't know what a construction is, but I will know when I see one and this is not it” he might have said. The motivation behind his original question was rather the desire to see disorder in an understandable fashion. Erdős would not care about polynomial computability of the adjacency relation; a computer can calculate many things where the human mind is not able to see anything. On the other hand, he would also not worry about the adjacency relation in the Paley graph being really computable in polylogarithmic time, before proclaiming the Paley graph a "construction". The Paley graph is not an explicit construction because of efficient computability, it is explicit because one looks at it and sees mathematically explainable disorder (should number theorists finally be able to prove that so).

The best strongly explicit construction by the Abbott-product (from Subsection 5.1.2) has a twice iterated logarithm in the exponent. In the next section we discuss a surprisingly simple strongly explicit construction, which beats slightly even the weakly explicit Abbott-type construction above.

The following exercise is good preparation for that. It was the first real breakthrough over the quadratic constructive lower bound of the Turán graph and it came in the same year (1972) as the Abbott-product. Nagy defined an infinite sequence of k-Ramsey graphs on $\Theta(k^3)$ vertices. Let G be the graph with $V(G) = \binom{[k]}{3}$, and $A \sim B$ if $|A \cap B| = 1$. The proof of correctness of the construction, i.e. that they don't contain large clique and independent set, is a beautiful application of the Linear Algebra Method.

Exercise 5.12 *Prove that the graph of Nagy contains no clique and no independent set of order $k+1$. (Hint for a proof via linear algebra: Prove that set of characteristic vectors of an independent set (or a clique) is linearly independent over an appropriately chosen field. Hint for a combinatorial proof: there is one.)*

5.2 The construction of Frankl and Wilson

In 1977 Frankl extended the construction of Nagy using the theory of *sunflowers* to obtain a constructive superpolynomial lower bound $k^{f(k)}$, with $f(k) = \Omega\left(\frac{\log k}{\log\log k}\right) \to \infty$. Later Frankl and Wilson (1981) gave a simpler proof through the linear algebra method. This is what we will discuss here. Let p be a prime and define the graph G by

$$V(G) = \binom{[p^3]}{p^2-1}, \quad A \text{ and } B \text{ are adjacent if } |A \cap B| \equiv -1 \pmod{p}.$$

Observe that for $p = 2$ we get back Nagy's construction with $k = 8$.

Theorem 5.6 *Graph G contains no clique and no independent set of size*

$$\sum_{i=0}^{p-1} \binom{p^3}{i} + 1.$$

Provided that the theorem holds, we have a $\sim p^{2p}$-Ramsey graph on $\sim p^{p^2}$ vertices.

Exercise 5.13 *Check (precisely!) that for* every *k we have a k-Ramsey graph with $k^{\Omega\left(\frac{\log k}{\log\log k}\right)}$ vertices.*

The proof of Theorem 5.6 is again a wonderful application of the linear algebra method, which goes one step further than the proof of the theorem of Nagy. Now characteristic vectors do not suffice; we need a simple technical lemma about *function spaces.* Let F be a field and $\Omega \subseteq F^n$. Then the set $F^\Omega = \{f : \Omega \to F\}$ of functions is a *vector space over F.*

Lemma 5.6.1 *If $f_1, \dots, f_m \in F^\Omega$ and $v_1, \dots, v_m \in \Omega$ such that*

- *$f_i(v_i) \neq 0$, and*
- *$f_i(v_j) = 0$ for all $j < i$,*

then $f_1, \dots, f_m$ are linearly independent in F^Ω.

Proof. (of Lemma 5.6.1) Suppose $\lambda_1 f_1 + \cdots + \lambda_m f_m = 0$, and let j be the smallest index j with $\lambda_j \neq 0$. Substituting v_j into this function equation we have

$$\underbrace{\lambda_1 f_1(v_j) + \cdots + \lambda_{j-1} f_{j-1}(v_j)}_{=0,\text{ since } \lambda_i = 0,\ i<j} + \underbrace{\lambda_j f_j(v_j)}_{\neq 0} + \underbrace{\lambda_{j+1} f_{j+1}(v_j) + \cdots + \lambda_m f_m(v_j)}_{=0,\text{ since } f_i(v_j) = 0,\ j<i} = 0,$$

a contradiction. □

Proof. (of Theorem 5.6) For a set $A \in 2^{[p^3]}$ let $v_A \in \{0,1\}^{p^3}$ be the characteristic vector of A. The linear algebra method is based on a simple, but crucial identity connecting the size of the intersection of two sets to the inner product of their characteristic vectors, namely that $|A \cap B| = \langle v_A, v_B \rangle$.

Independent sets. Let $A_1, \dots, A_s$ be an independent set in G, so $|A_i \cap A_j| \not\equiv -1 \pmod p$ for every $i \neq j$. For each i let $v_i = v_{A_i}$ be the characteristic vector of A_i. Our plan is to define a function $f_i : \{0,1\}^{p^3} \to \mathbb{F}_p$ for every $i = 1, \dots s$, prove that they are linearly independent and bound the dimension of the vector space they span — giving us an upper bound on s. Let

$$\tilde{f}_i(x) = \prod_{l=0}^{p-2} (\langle x, v_i \rangle - l),$$

for all i. Obviously we have $\tilde{f}_i(v_i) \neq 0$, since $|A_i| \equiv -1 \pmod p$. On the other hand, we have $\tilde{f}_i(v_j) = 0$ for all $j \neq i$, since $\{A_1, \ldots, A_s\}$ is an independent set. Our technical lemma then implies that $\tilde{f}_1, \ldots, \tilde{f}_s$ are linearly independent. The dimension of the space these functions span could be quite large, since each variable x_j, $j = 1, \ldots, p^3$ could appear with powers ranging from 0 to $p-1$. To reduce the dimension of the space, we apply a "multilinearization trick" and define $f_i(x)$ from $\tilde{f}_i(x)$ by replacing each occurrence of a large power x_i^l $(l > 1)$ with x_i. Observe that $f_i \equiv \tilde{f}_i$ on $\{0,1\}^{p^3}$. Since all the f_i are multilinear polynomials, the dimension of the space spanned by them is the number of monomials of degree at most $p-1$,

$$1 + p^3 + \binom{p^3}{2} + \cdots + \binom{p^3}{p-1}.$$

Cliques. To bound the clique number of G we proceed similarly, but we will work over $\mathbb{R}$ instead of $\mathbb{F}_p$. Let $B_1, \ldots, B_t$ be a clique in G, so $|B_i \cap B_j| \equiv -1 \pmod p$ for every $i \neq j$. Let $L = \{p-1, 2p-1, \ldots, p^2-p-1\}$ be the set of possible intersection sizes. Note that $|L| = p-1$. For each i let $w_i = v_{B_i}$ be the characteristic vector of B_i and let

$$\tilde{f}_i(x) = \prod_{l \in L} (\langle x, w_i \rangle - l)$$

be functions $\{0,1\}^{p^3} \to \mathbb{R}$ for all i. Since $|B_i| = p^2 - 1 \notin L$, we have $\tilde{f}_i(w_i) \neq 0$. On the other hand, $\tilde{f}_i(w_j) = 0$ for all $j \neq i$. Lemma ?? then implies that $\tilde{f}_1, \ldots, \tilde{f}_t$ are linearly independent. Again, we multilinearize the functions and define $f_i(x)$ from $\tilde{f}_i(x)$ by replacing each occurrence of a large power x_i^l $(l > 1)$ with x_i. Since $|L| = p-1$, all the f_i are multilinear polynomials of degree at most $p-1$. Thus the dimension of the space spanned by them is at most

$$1 + p^3 + \binom{p^3}{2} + \cdots + \binom{p^3}{p-1}.$$

□

Exercise 5.14 *The proof of the following theorem is an immediate generalization of the claim we had about the clique number of the Frankl-Wilson graph. (Think this over!)*

Theorem *Let L be a set of integers with $|L| = s$. Let $B_1, \ldots, B_t \in 2^{[n]}$ be a uniform L-intersecting family, i.e. all $|B_i|$ have the same size and $|B_i \cap B_j| \in L$ for every $i \neq j$. Then $t \leq \sum_{i=0}^{s} \binom{n}{i}$.* □

Generalize this statement further to arbitrary *L-intersecting families, i.e. derive the same conclusion when the $|B_i|$ are not necessarily all equal. (Hint: Select the functions $\tilde{f}_i$ more carefully and use Lemma 5.6.1 in its full power.)*