above (which we leave as an exercise) establishes that $\mathbf{E}(\partial_*^\alpha Y_m) = O_h(m^{-1/h})$ for all non-zero α. The claim now follows from Theorem 1.38. □

The study of $B_h[g]$ sets is a popular topic in additive combinatorics. A detailed discussion of this topic is beyond the scope of our book. Let us, however, mention one new result of Cilleruelo, Ruzsa and Trujillo from [62]. Many other recent results can be found in [62, 191, 213, 61, 145, 272].

Let $A \subset [1, N]$ be a $B_h[g]$ set. A simple counting argument (related to (1.21)) gives $\binom{|A|+h-1}{h} \leq ghN$, which in turn yields the trivial bound $|A| \leq (ghh!N)^{1/h}$. Cilleruelo, Ruzsa and Trujillo gave the first non-trivial bounds for the case $g \geq 2$. They prove that $|A| \leq 1.864(gN)^{1/2} + 1$ when $h = 2$, and that

$$F_h(g, N) \leq (1 + \cos^h(\pi/h))^{-1/h}(hh!gN)^{1/h}$$

when $h > 2$. The proofs made use of harmonic analysis methods via the consideration of the trigonometric polynomials $f(t) = \sum_{a\in X} e^{iat}$. The authors also constructed sets to establish for any g, the existence of a $B_2[g]$ set $A \subset [1, N]$ with

$$|A| \geq \left(\frac{g + [g/2]}{\sqrt{g + 2[g/2]}} + o_g(1)\right) N^{1/2}.$$

Exercises

1.7.1 Consider the random graph $G(n, p)$ defined in Exercise 1.6.2, and set $p := n^{-1+\epsilon}$. Let Y be the number of triangles in $G(n, p)$. Give an upper bound and a lower bound for

$$\mathbf{P}\left(Y \geq \frac{3}{2}\mathbf{E}(Y)\right).$$

1.7.2 Verify the bound $\mathbf{E}(\partial_*^\alpha Y_m) = O_h(n^{-1/h})$ claimed in the Proof of Theorem 1.39.

1.8 Thin bases of higher order

We now return to the study of thin bases B and their associated counting functions $r_{k,B}(n)$, initiated in Section 1.3. However, in this section we can use Theorem 1.37 to present a proof of Theorem 1.15, which asserted for each $k \geq 1$ the existence of a base B of order k with $r_{k,B}(n) = O_k(\log n)$ for all large n. This was proven in the $k = 2$ case (see Theorem 1.13) using Chernoff's inequality, but that method does not directly apply for higher k because $r_{k,B}(n)$ cannot be easily expressed as the sum of independent random variables.

We begin with a simple lemma on boolean polynomials that shows that if $\mathbf{E}(X)$ is not too large, then at most points $(t_1, \ldots, t_n)$ of the sample space, the polynomial X does not contain too many independent terms (cf. Exercise 1.3.12).

Lemma 1.40 *Let $X = \sum_{A \in \mathcal{A}} \prod_{j \in A} t_j$ be a boolean polynomial of n independent boolean variables $t_1, \ldots, t_n$, let $B \subseteq [1, n]$ be the random set $B := \{j \in [1, n] : t_j = 1\}$, and let $D \in \mathbf{N}$ be the random variable, defined as the largest number of disjoint sets in $\mathcal{A}$ which are contained in B. Then for any integer $K \geq 1$ we have*

$$\mathbf{P}(D \geq K) \leq \frac{\mathbf{E}(X)^K}{K!}.$$

Proof Observe that for $A_1, \ldots, A_k$ disjoint,

$$\mathbf{I}(D \geq K) \leq \frac{1}{K!} \sum_{A_1, \ldots, A_K \in \mathcal{A}, \text{disjoint}} \prod_{j \in A_1} t_j \ldots \prod_{j \in A_k} t_j.$$

Taking expectations of both sides and using linearity of expectation (1.3) followed by independence, we conclude

$$\mathbf{P}(D \geq K) \leq \frac{1}{K!} \sum_{A_1, \ldots, A_K \in \mathcal{A}} \mathbf{E}\left(\prod_{j \in A_1} t_j\right) \ldots \mathbf{E}\left(\prod_{j \in A_k} t_j\right).$$

But by linearity of expectation again, the left-hand side is just $\mathbf{E}(X)^K / K!$, and the claim follows. □

This lemma is particularly useful when combined with the *sunflower lemma* of Erdős and Rado [95]. A collection of sets $A_1, \ldots, A_l$ forms a *sunflower* if the pairwise intersections $A_i \cap A_j$ for $i \neq j$ are all the same (the A_i are called the *petals* of the flower). We allow this common pairwise intersection to be empty.

Lemma 1.41 (Sunflower lemma) *If $\mathcal{A}$ is a collection of sets, each of size at most k, and $|\mathcal{A}| > (l-1)^k k!$, then $\mathcal{A}$ contains l sets forming a sunflower.*

This lemma can be proven by elementary combinatorics and is left as an exercise. It has the following consequence for the counting function $r_{k,B}(n)$.

Corollary 1.42 *Let $B \subset \mathbf{Z}^+$ and $k \geq 2$, and for each $n \in \mathbf{Z}^+$ let $D_{k,n}$ be the largest number of disjoint multisets[2] $\{x_1, \ldots, x_k\}$ of elements of B which sum to n. Then*

$$r_{k,B}(n) \leq k! k^k \max\left(D_{k,n}, \left(\sup_{m<n} r_{k-1,B}(m) - 1\right)^k\right).$$

Proof Fix n, and consider the collection $\mathcal{A}$ of sets which arise from taking the multisets $\{x_1, \ldots, x_k\}$ of elements of B which sum to n and then removing repeated

[2] A multiset is a set which is allowed to have repeated elements

elements. Clearly $r_{k,B}(n) \leq k^k|\mathcal{A}|$. Also observe that any sunflower in $\mathcal{A}$ has cardinality at most $D_{k,n}$ (if the petals are disjoint) or $\sup_{m<n} r_{k-1,B}(m)$ (if the petals are not disjoint); the latter follows by taking one of the elements in the common intersection of the sunflower and removing it once from each of the associated multisets. The claim then follows from the sunflower lemma. □

Using the above methods, we can now give a preliminary result towards proving Theorem 1.15.

Proposition 1.43 *Let $k \geq 2$, and let $B \subset \mathbf{Z}^+$ be a random subset of $\mathbf{Z}^+$, defined by letting $x \in B$ be independent with probability*

$$\mathbf{P}(x \in B) = \min\left(Cx^{1/k-1}\log^{1/k} x, 1\right)$$

for some positive constant $C > 1$. Then with probability 1, we have $\sup_n r_{k',B}(n) = O_{C,k,k',B}(1)$ for all $1 \leq k' < k$.

Proof We induce on k. The case $k = 1$ is obvious. Now suppose that $1 < k' < k$ and the claim has already been proven for $k' - 1$. Applying Corollary 1.42, we conclude that, with probability 1,

$$r_{k',B}(n) = O_{C,k,k',B}\left(\left(D_{k',n}+1\right)^{k'}\right). \tag{1.39}$$

On the other hand, if we apply Lemma 1.40 with $t_x := \mathbf{I}(x \in B)$ for $1 \leq x \leq n$, and $\mathcal{A} = \mathcal{A}_n$ equal to all the sets which arise from the multisets $\{x_1, \ldots, x_{k'}\}$ that sum to n, then we observe that

$$\mathbf{P}(D_{k',n} \geq K) \leq \frac{\mathbf{E}\left(\sum_{A\in\mathcal{A}_n}\prod_{j\in A} t_j\right)^K}{K!}$$

for any $K \in \mathbf{Z}^+$. However, from linearity of expectation (1.3) and independence we have

$$\begin{aligned}
\mathbf{E}\left(\sum_{A\in\mathcal{A}_n}\prod_{j\in A} t_j\right) &= \sum_{A\in\mathcal{A}_n}\prod_{j\in A}\min\left(Cj^{1/k-1}\log^{1/k} j, 1\right)\\
&\leq O_{C,k,k'}\left(\sum_{j_1\leq\cdots\leq j_{k'}: j_1+\ldots+j_{k'}=n} j_1^{1/k-1}\cdots j_{k'}^{1/k-1}\right)\log n\\
&\leq O_{C,k,k'}\left(\sum_{j_1,\ldots,j_{k'-1}\in[1,n]} j_1^{1/k-1}\cdots j_{k'-1}^{1/k-1}\right)n^{1/k-1}\log n\\
&= O_{C,k,k'}\left(\left(\sum_{j\in[1,n]} j^{1/k-1}\right)^{k'-1}\right)n^{1/k-1}\log n\\
&= O_{C,k,k'}\left(n^{k'/k-1}\log n\right).
\end{aligned}$$

Since $k' < k$, we thus see that, by choosing K depending on k sufficiently large (e.g. $K = 2k + 1$), we have

$$\mathbf{P}(D_{k',n} \geq K) = O_{C,k,k',K}\left(\frac{1}{n^2}\right).$$

Applying the Borel–Cantelli lemma (Lemma 1.2) we see that with probability 1, we have $D_{k',n} < K$ for all but finitely many n. Combining this with (1.39) we obtain the claim. □

Now we prove Theorem 1.15. It will suffice to show that

Proposition 1.44 *Let $k \geq 2$, and let $B \subset \mathbf{Z}^+$ be a random subset of $\mathbf{Z}^+$, defined by letting $x \in B$ be independent with probability*

$$\mathbf{P}(x \in B) = \min(Cx^{1/k-1}\log^{1/k} x, 1)$$

for some positive constant $C > 1$. If C is sufficiently large depending on k, then with probability 1, we have $r_{k,B}(n) = \Theta_{C,k}(\log n)$ for all but finitely many n. In particular, B is a thin basis of order k with probability 1.

Proof We shall estimate $r_{k,B}(n)$ in terms of two related expressions:

$$R(n) := \{(x_1, \ldots, x_k) \in B : x_1 + \cdots + x_k = n; n^{0.1} < x_1 < x_2 < \cdots < x_k\} \tag{1.40}$$

$$E(n) := \{(x_1, \ldots, x_k) \in B : x_1 + \cdots + x_k = n; x_1 = x_2 \text{ or } x_1 \leq n^{0.1}\}. \tag{1.41}$$

It is clear (using the symmetry of $x_1 + \cdots + x_k$ under permutations) that

$$k!R(n) \leq r_{k,B}(n) \leq k!R(n) + k^2E(n).$$

We view $R(n)$ as the main term and $E(n)$ as the error term; this reflects the intuitive fact that for most representations $n = x_1 + \cdots + x_k$, the x_i will be distinct and comparable in magnitude to n. It will suffice to show that with probability 1 we have

$$E(n) = O_{C,k,B}(1); \qquad R(n) = \Theta_{C,k,B}(\log n)$$

for all but finitely many n.

Let us deal first with the error term $E(n)$. We argue as in the proof of Proposition 1.43. Let $\mathcal{A}_n$ denote those sets which arise from the multisets $\{x_1, \cdots, x_k\}$ with $x_1 + \cdots + x_k = n$ and either $x_1 = x_2$ or $x_1 \leq n^{0.1}$. By arguing as in Corollary 1.42, we have

$$E(n) \leq k!k^k \max\left(D_n, \left(\sup_{m<n} r_{k-1,B}(m) - 1\right)^k\right)$$

where D_n is the largest number of disjoint sets that one can find in $\mathcal{A}_n$. Applying Proposition 1.43, we conclude that

$$E(n) = O_{C,k,B}(D_n + 1)$$

with probability 1. On the other hand, from Lemma 1.40, we have for any K that

$$\mathbf{P}(D_n \geq K) \leq \frac{\mathbf{E}(\sum_{A\in\mathcal{A}_n} \prod_{j\in A} t_j)^K}{K!}.$$

By arguing as in Proposition 1.43, one can establish

$$\mathbf{E}\left(\sum_{A\in\mathcal{A}_n} \prod_{j\in A} t_j\right) \leq O_k(n^{-1/k} n^{-0.9/k} \log n)$$

and thus, for a suitably large constant K depending only on k,

$$\mathbf{P}(D_n \geq K) = O_k(1/n^2).$$

From the Borel–Cantelli lemma we conclude that, with probability 1,

$$E(n) = O_{C,k,B}(1)$$

for all but finitely many n, and so the contribution of $E(n)$ is negligible.

Now we estimate the main term $R(n)$. Observe that we can write $R(n)$ as a homogeneous boolean polynomial $Y = Y(t_1, \ldots, t_n)$ of degree k; more explicitly, we have

$$Y(t_1, \ldots, t_n) = \sum_{A\in\mathcal{A}'_n} \prod_{j\in A} t_j$$

where $\mathcal{A}'_n$ is the collection of all sets $\{x_1, \ldots, x_k\}$ where $x_1 + \cdots + x_k = n$ and $n^{0.1} < x_1 < x_2 < \cdots < x_k$. Repeating the computations in Proposition 1.43 we see that

$$\mathbf{E}(Y) = \Theta_k(C \log n)$$

when n is sufficiently large depending on C, k. To conclude the proof it would thus suffice by the Borel–Cantelli lemma to establish the large deviation inequality

$$\mathbf{P}\left(|Y - \mathbf{E}(Y)| > \frac{1}{2}\mathbf{E}(Y)\right) = O_{C,k}\left(\frac{1}{n^2}\right)$$

for all large n. Applying Theorem 1.37 (and choosing C sufficiently large), we see that it suffices to show the derivative estimates

$$\mathbf{E}_1(Y), \ldots, \mathbf{E}_{k-1}(Y) \leq n^{-\gamma}$$

for all large n and some $\gamma > 0$. In other words, we need to establish

$$\mathbf{E}\left(\left(\frac{\partial}{\partial t_1}\right)^{\alpha_1} \ldots \left(\frac{\partial}{\partial t_n}\right)^{\alpha_n} Y(t_1, \ldots, t_n)\right) \leq n^{-\gamma}$$