Algebraic and Combinatorial Methods in the Theory of Set Addition

Károlyi Gyula

akadémiai doktori értekezés

Budapest, 2007

Contents

Foreword

This dissertation contains a good part of the results of my reserach in additive combinatorics I have been conducting during the last decade. It is based on the papers [20] and [51]–[58] all whose central theme is connected to the theory of set addition. The four main chapters contain results obtained by four different methods reflected in their respective titles. The results in the first three of those chapters nicely fit into a general framework that we explain in the introduction. The last chapter appears to be out of this context at a first glance. Most of the results therein, however, can be traced back to the Erdős–Heilbronn problem, which is in the center of these investigations. Therefore we feel that the present work contains quite a coherent section of our research curriculum.

Most of the above mentioned papers have already been refereed and published. Exceptions are [57] and [58], from which the whole Chapter 6 is extracted, and [56] that contains Section 4.1. The paper [20] I have written with coauthors; from that paper I only include here those results in which my contribution was more than essential.

During this work I benefitted a lot from the knowledge, support, encouragement and friendship of many colleagues, including Noga Alon, Imre Bárány, Marc Burger, Jean-Pierre Bourguignon, Shalom Eliahou, Komei Fukuda, Yahya Ould Hamidoune, Anna Lladó, Monique Laurent, Seva Lev, László Lovász, Hans-Jakob Lüthi, Péter Pálfy, Lajos Rónyai, Vera Rosta, Imre Ruzsa, Lex Schrijver, Oriol Serra, Balázs Szegedy, Tamás Szőnyi, Kati Vesztergombi, and Emo Welzl. I also greatfully acknowledge the support of the National Scientific Research Funds (OTKA) and the Bolyai Research Fellowship as well as the support and hospitality of the following institutions: the CRM in Montréal, the CWI in Amsterdam, the ETH in Zürich, the IAS in Princeton, the IHÉS in Bures-sur-Yvette, the RI in Budapest, and the UPC in Barcelona.

My greatest gratitude goes to Gabi and Béla Bollobás who helped me in every possible respect just when everything seemed to collapse.

I dedicate this dissertation to my father who could have been a great scientist.

Notation

If q is a power of a prime number, then the Galois field $\mathrm{GF}(q)$ of q elements will be denoted by $\mathbb{F}_q$. For a positive integer n, $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ denotes the cyclic group of order n, whereas $\mathbb{Q}_n = \mathbb{Q}(e^{2\pi i/n}) = \mathbb{Q}(x)/(\Phi_n(x))$ stands for the nth cyclotomic field, Φ_n denoting the nth cyclotomic polynomial. The symmetric group of degree k is denoted by S_k.

For a nontrivial group G we denote by $p(G)$ the order of the smallest nontrivial subgroup of G. If G is finite, then $p(G)$ equals the smallest prime divisor of the order of G. On the other hand, $p(G) = \infty$ if and only if G is torsion free. For an abelian group G and a natural number n we denote by G^n the direct sum of n copies of G.

A and B will always denote (usually nonempty) subsets of some group G. Unless declared otherwise, their cardinalities will be denoted by $|A| = k$ and $|B| = \ell$, respectively. In case of abelian groups we will use additive notation. In that case

$$A + B = \{a + b \mid a \in A,\ b \in B\}$$

stands for the usual Minkowski-sum of A and B, whereas

$$A \dot{+} B = \{a + b \mid a \in A,\ b \in B, a \neq b\}$$

denotes their so-called restricted sum. If G is not declared to be commutative, we will stick to the more accepted multiplicative notation. Thus, $AB = \{ab \mid a \in A,\ b \in B\}$ in such a case.

In the last chapter, for positive integers $a < b$ we will use the notation

$$[a, b] = \{a, a + 1, \ldots, b - 1, b\}.$$

The sum of the elements of a set B will be denoted by $\sigma(B)$, and $\Sigma(A) = \{\sigma(B) \mid B \subseteq A\}$ will represent the set of all possible subset sums of A, including $0 = \sigma(\emptyset)$. The notation

$$\Sigma_d(A) = \{\sigma(B) \mid B \subseteq A, |B| = d\}$$

is a deviation from the standard notation used in the context of restricted multiple set addition. Finally, if A is a set of integers and q is a positive integer, then $N_q(A)$ denotes the number of elements in A not divisible by q.

The rest of the notation we use throughout this dissertation is all standard.

Chapter 1

First Principles

Perhaps the most ancient result in combinatorial number theory is the following. Let p denote a prime number. If the nonempty sets A and B of integers intersect k and ℓ different residue classes modulo p, respectively, then in case $p \geq k + \ell - 1$, at least $k + \ell - 1$ different residue classes are represented by the numbers $a + b$ with $a \in A$, $b \in B$. In our terminology: If A, B are nonempty subsets of $\mathbb{Z}_p$, then $p \geq |A| + |B| - 1$ implies $|A + B| \geq |A| + |B| - 1$. This result is due to Cauchy [16] who invented it in relation to Lagrange's famous 'four squares theorem', and is referred to as the Cauchy–Davenport theorem. After Davenport [21] rediscovered the result in 1935, it was immediately generalized by Chowla [19] and Pillai [72]. The short but tricky combinatorial proof actually gives the following generalization (see e.g. [53]), which is a good starting point to the present dissertation.

Theorem 1.1. *If A and B are nonempty subsets of an abelian group G such that $p(G) \geq |A| + |B| - 1$, then $|A + B| \geq |A| + |B| - 1$.*

Proof. Assume that $|A| \leq |B|$. If $|A| = 1$, then clearly $|A+B| = |B| = |A|+|B|-1$. Otherwise assume for a moment that B intersects A properly, that is, $A \cap B \neq \emptyset$ and $A \setminus B \neq \emptyset$. In this case we may replace A with the set $A' = A \cap B$ and B with $B' = A \cup B$ such that $0 < |A'| < |A|$, $|A'| + |B'| - 1 = |A| + |B| - 1$ and $A' + B' \subseteq A + B$, implying $|A' + B'| \leq |A + B|$. If B does not intersect A properly, we still can do the following. Choose some $c \in G$ such that the set $B + c = B + \{c\}$ intersects A properly. Then replace A with the set $A' = A \cap (B + c)$ and B with $B' = A \cup (B + c)$. Note that $|B + c| = |B|$ and that $A + (B + c) = (A + B) + c$, implying $|A+(B+c)| = |A+B|$. Therefore again we have that $0 < |A'| < |A|$, $|A'|+|B'|-1 = |A|+|B|-1$ and $|A' + B'| \leq |A + B|$. Thus, it suffices to prove the estimate for the sets A' and B'. In a finite number of steps we can reduce the problem to the case when $|A| = 1$, and the result follows.

It only remains to prove that an appropriate $c \in G$ can be found. First, there is a $c_0 \in G$ such that $A \cap (B + c_0)$ is not empty. If A is not contained in $B + c_0$, then $c = c_0$ will do. Otherwise there are two different elements of A, say a and $b = a - c_1$, that both belong to

$B+c_0$. Since $|B+c_0| = |B| < p(G)$ and the numbers $a, a-c_1, a-2c_1, \ldots, a-(p(G)-1)c_1$ are all different, there is a smallest positive integer t such that $a-tc_1 \in B+c_0$ but $a-(t+1)c_1 \notin B+c_0$. Writing $c = c_0 + tc_1$ we can conclude that $a \in A \cap (B+c)$ and $b = a - c_1 \in A \setminus (B+c)$, which makes the proof complete. □

This idea has eventually led to Vosper's inverse theorem [87] and also to Kneser's theorem [61] that became a very powerful tool in combinatorial number theory.

Kneser's theorem states that if A, B are finite nonempty subsets of an abelian group G, then either $|A+B| \geq |A| + |B|$, or

$$|A+B| = |A+H| + |B+H| - |H|,$$

where $H = \{g \in G \mid (A+B)+g = A+B\}$ is the stabilizer, or the set of periods, of $A+B$. Note that H is clearly a subgroup of G and $A+B$ is a union of certain cosets of H. It implies Theorem 1.1 as follows. Assume that A, B are finite nonempty subsets of G such that $p(G) \geq |A| + |B| - 1$. If $|A+B| \geq |A| + |B|$, then we are ready. Otherwise, if 0 is the only period of $A+B$, then $|A+B| = |A+H| + |B+H| - |H| = |A| + |B| - 1$. Finally, if H is a nontrivial subgroup of G, then $|H| \geq p(G)$, and therefore $|A+H| \geq |H|$ and $|B+H| \geq |H|$ imply

$$|A+B| = |A+H| + |B+H| - |H| \geq |H| \geq p(G) \geq |A| + |B| - 1.$$

Instead of going deeper into the history at this point, we present in the next section a list of statements that are relevant to our work and can be easily proved in any linearly ordered abelian group. A standard compactness argument implies that the statements are valid in any abelian group G with $p(G)$ large enough. A more effective principle is discussed in the section that follows. After that we return to the history of the subject and describe our main new results in this context. This is followed by a brief description of the algebraic background and the new methods we employ in the dissertation.

1.1 A General Framework

Let G be an abelian group and let A, B be nonempty subsets of G. Assume that, like in the case of $\mathbb{Z}$ and $\mathbb{Q}$, there is a linear order $<$ on G, which is compatible with the addition on G, that is, for arbitrary elements $a, b, c \in G$, $a < b$ implies $a + c < b + c$. It is immediate that such a linearly orderable group cannot have any nonzero element of finite order. It is also easy to see, that if the abelian groups G and H are linearly orderable, then so is their direct sum $G \oplus H$. Thus, every finitely generated torsion free abelian group can be equipped with such a linear order. In fact, it can be proved using transfinite induction, that even the direct sum

of infinitely many linearly orderable abelian groups can be ordered. Since every torsion free abelian group is a subgroup of the direct sum of some isomorphic copies of $\mathbb{Q}$ (see e.g. [76]), we arrive at the (well known) conclusion that an abelian group can be ordered if and only if it is torsion free.

Thus, if G is torsion free, then the elements of A and B can be enumerated as $a_1 < a_2 < \ldots < a_k$ and $b_1 < b_2 < \ldots < b_\ell$ such that

$$a_1 + b_1 < a_2 + b_1 < \ldots < a_k + b_1 < a_k + b_2 < \ldots < a_k + b_\ell.$$

Moreover, at most one element of A can be equal to b_1, and no more than one member of B can equal a_k. It follows that the following statements are valid in any torsion free abelian group G.

Statement 1.2. *If A and B are nonempty finite subsets of the abelian group G,then $|A+B| \geq k + \ell - 1$.*

Statement 1.3. *If A and B are nonempty finite subsets of the abelian group G,then $|A\dot{+}B| \geq k + \ell - 3$.*

In particular,

Statement 1.4. *If A is a finite subset of the abelian group G, then $|A + A| \geq 2k - 1$.*

Statement 1.5. *If A is a finite subset of the abelian group G, then $|A\dot{+}A| \geq 2k - 3$.*

If A is different from B, then we can say something stronger:

Statement 1.6. *If A and B are nonempty finite subsets of the abelian group G such that $A \neq B$, then $|A\dot{+}B| \geq k + \ell - 2$.*

Indeed, if $k = 1$, then $|A\dot{+}B| \geq |B| - 1 = k + \ell - 2$, and we can argue in a similar way if $\ell = 1$. Thus, we may assume that $k, \ell \geq 2$ and we have already proved that $|A'| + |B'| < k + \ell$ and $|A'\dot{+}B'| = |A'| + |B'| - 3$ implies $A' = B'$. If $a_1 \neq b_1$, then we may assume without any loss of generality that $b_1 < a_1$. In this case no element of A can be equal to b_1, so at least $k + \ell - 2$ out of the $k + \ell - 1$ different numbers

$$a_1 + b_1 < a_2 + b_1 < \ldots < a_k + b_1 < a_k + b_2 < \ldots < a_k + b_\ell$$

belong to $A\dot{+}B$. Thus, we may assume that $a_1 = b_1$, and also that $k \leq \ell$, say. Since $A \neq B$, there is a smallest integer t with the property that $a_t = b_t$ but $a_{t+1} \neq b_{t+1}$. If $t = k$, that is, a_{t+1} does not even exist, we find that $\ell > k \geq 2$ and then $A\dot{+}B$ contains the following $k + \ell - 2$ different numbers:

$$a_1 + b_2 < \ldots < a_1 + b_k < \ldots < a_{k-1} + b_k <$$

$$a_{k-1} + b_{k+1} < a_k + b_{k+1} < \ldots < a_k + b_\ell.$$

Otherwise we may assume that $a_{t+1} < b_{t+1}$, and even if $t = 1$, we can consider the following $2t-2$ elements of $A\dot{+}B$:

$$a_1 + b_2 < \ldots < a_1 + b_t < \ldots < a_{t-1} + b_t < a_{t+1} + b_{t-1}.$$

Defining $A' = A \setminus \{a_1, \ldots, a_t\}$ and $B' = B \setminus \{b_1, \ldots, b_t\}$ we find that $A' \neq B'$, so by our induction hypothesis, $|A'\dot{+}B'| \geq (k-t)+(\ell-t)-2$. This way we found $k+\ell-2t-2$ elements of $A\dot{+}B$, each larger than the previously found $2t-2$ numbers. Finally, the elements $a_{t+1}+b_t$ and $a_t + b_{t+1}$ also belong to $A\dot{+}B$ and they are both larger than the first $2t-2$ numbers and at the same time smaller than the elements of $A'\dot{+}B'$. That is,

$$|A\dot{+}B| \geq (2t-2) + (k+\ell-2t-2) + 2 = k+\ell-2,$$

as we wanted to prove. □

It is not difficult to characterize the sets A and B for which equality holds in Statement 1.2, a proof can be found in [69].

Statement 1.7. *If A and B are nonempty finite subsets of the abelian group G such that $|A+B| = k+\ell-1$, then A and B are both arithmetic progressions of the same difference.*

In particular, the following statement is also valid in every torsion free abelian group:

Statement 1.8. *If A is a nonempty finite subset of the abelian group G such that $|A+A| = 2k-1$, then A is an arithmetic progressions.*

In view of Statement 1.6, $|A\dot{+}B| = k+\ell-3$ is only possible if $A = B$. If k is 2 or 3, then clearly $|A\dot{+}A| = 2k-3$. If k is 4, then $|A\dot{+}A|$ is either 5 or 6, where the first case happens if and only if $a_1 + a_4 = a_2 + a_3$. Otherwise the analogue of the previous statement is true, see [69].

Statement 1.9. *If A is a finite subset of the abelian group G such that $k = |A| \geq 5$ and $|A\dot{+}A| = 2k-3$, then A is an arithmetic progression.*

Assume now that $a_1 \leq a_2 \leq \ldots \leq a_k$ and $b_1 < b_2 < \ldots < b_k$, then clearly

$$a_1 + b_1 < a_2 + b_2 < \ldots < a_k + b_k.$$

Consequently, the following statements are also valid in every torsion free abelian group G.

Statement 1.10. *If A and B are subsets of the abelian group G, each of cardinality k, then there are numberings $a_1, a_2, \ldots, a_k$ and $b_1, \ldots, b_k$ of the elements of A and B, respectively, such that the sums $a_1+b_1, a_2+b_2, \ldots, a_k+b_k$ are pairwise different.*

Statement 1.11. *Let $A = (a_1, \ldots, a_k)$ be a sequence of k elements in the abelian group G. Then for any subset $B \subset G$ of cardinality k there is a numbering $b_1, \ldots, b_k$ of the elements of B such that the sums $a_1+b_1, a_2+b_2, \ldots, a_k+b_k$ are pairwise different.*

That is, Statement 1.10 is also true if A is a multiset. Finally, if A is a finite multiset of at least two nonzero elements in a linearly ordered abelian group, then it can be partitioned into two nonempty multisets containing the negative and the positive elements of A, respectively, such that no elements in the same part can add up to zero (take any partition if all the elements of A have the same sign). Consequently, the following is true in torsion free abelian groups G.

Statement 1.12. *Any multiset of $k \geq 2$ nonzero elements of G can be partitioned into two nonempty parts such that in none of the parts does a zero subsum occur.*

Common features of all the above statements are that for fixed values of k and ℓ they can be written as a closed formula in the first order language of abelian groups, and that they are valid in every linearly ordered, and thus also in every torsion free abelian group. Based on a standard compactness argument it follows that the same statements hold in any abelian group G for which $p(G)$ is large enough compared to k and ℓ.

Theorem 1.13. *Let Φ be any statement that can be formulated as a sentence in the first order language of abelian groups. Assume that Φ is true in every linearly orderable abelian group. Then there is an integer $p_0 = p_0(\Phi)$ such that Φ is valid in every abelian group G with $p(G) \geq p_0$.*

Proof. Assume that, on the contrary, there is an infinite sequence of prime numbers $p_1 < p_2 < p_3 < \ldots$ such that, for every positive integer i, there is an abelian group G_i with the property that $p(G_i) = p_i$ and Φ is not valid in G_i. Let U denote any non-principal ultrafilter on the set of positive integers $\mathbb{Z}_+$, it contains all co-finite subsets of $\mathbb{Z}_+$. Let $G = \prod G_i/U$ be the ultraproduct of the groups G_i with respect to U.

According to the fundamental theorem of ultraproducts, also known as Łoś's theorem (cf. [17, 40]), a sentence Ψ in the first order language of abelian groups is true in G if and only if the set

$$\{i \in \mathbb{Z}_+ \mid \Psi \text{ is valid in } G_i\}$$

belongs to U. Since $\neg\Phi$ is valid in every G_i and, by definition, $\mathbb{Z}_+ \in U$, it follows that Φ is not valid in G.

Notice that, for any fixed k, the statement Ψ_k 'there is no nonzero element whose order is less than k' is in fact a first order sentence for abelian groups. Since for any fixed k there is only a finite number of indices i with $p_i < k$, the set of indices for which Ψ_k is valid in G_i belongs to U. It follows that for every k, no element of G other than 0 can have an order less than k, implying that G is torsion free. Consequently, G can be ordered, and thus Φ is valid in G. This contradiction completes the proof. □

We note that a similar argument has also been suggested by Ambrus Pál [71], see also [49].

As we have already mentioned, all the above statements of the previous section can be expressed as a first order sentence, and thus must be valid, in the view of the above theorem, whenever $p(G)$ is large enough compared to k and ℓ.

Now we turn our attention to more efficient methods. The drawback of above argument on one hand is that it depends on the axiom of choice, and on the other hand is that it does not say how large $p(G)$ should indeed be. An effective, though in general exponentional admissible bound can be obtained by the rectification principle of Freiman [37], worked out by Bilu, Ruzsa and Lev for cyclic groups of prime order in [13]. We elaborate on this idea in the next section.

1.2 The Rectification Principle

Let Φ be any closed formula in the first order language of abelian groups, written inductively in the usual way. Every atomic formula that occurs in Φ is of the form $\tau = \sigma$ where

$$\tau = x_1 + x_2 + \ldots + x_{v(\tau)} \text{ and } \sigma = y_1 + y_2 + \ldots + y_{v(\sigma)},$$

such that $x_1, x_2, \ldots, x_{v(\tau)}$ and $y_1, y_2, \ldots, y_{v(\sigma)}$ are not necessarily different variables of Φ. We say that Φ is an (s,t)-*sentence* if $\Phi = \forall x_1 \ldots \forall x_t \Psi$, where Ψ only contains the open variables $x_1, \ldots, x_t$ and, for every atomic formula $\tau = \sigma$ that occurs in Φ, we have $v(\tau) + v(\sigma) \leq s$. We will assume that $s \geq 2$. For example, Statement 1.12 in the case $k = 3$ can be written as a $(2,3)$-sentence as follows:

$$\forall x \forall y \forall z((\neg(x = 0) \wedge \neg(y = 0) \wedge \neg(z = 0)) \rightarrow$$

$$(\neg(x + y = 0) \vee \neg(x + z = 0) \vee \neg(y + z = 0))),$$

a formula that is clearly valid in every abelian group G with $p(G) > 2$. Here, in the atomic sub-formula $x + y = 0$, we have $v(x + y) = 2$ and $v(0) = 0$.

An effective version of Theorem 1.13 is the following

Theorem 1.14. *Let Φ be an (s,t)-sentence in the first order language of abelian groups. If Φ is true in $\mathbb{Z}$, then it is valid in every abelian group G with $p(G) > s^t$.*

Thus we have a tool even for such problems, where we cannot argue using the appropriate ordering of torsion free abelian groups, but instead of that we somehow can exploit the arithmetic and/or some other properties of $\mathbb{Z}$, like in the following well-known exercise: *If $n_1, n_2, \ldots, n_{2k+1}$ are integers with the property that, whichever number we omit, the rest can be partitioned into two k-element groups with equal sums, then all the numbers are equal.*

To prove Theorem 1.14 we follow [13]. Note that we may readily assume that G is finitely generated. We use the following notion of Freiman-isomorphism. For subsets K and L of

the abelian groups G and H, respectively, we say that the bijection $\varphi : K \to L$ is an $\tilde{F}_s$-isomorphism, if for any $a_1, \ldots, a_u \in K$ and $b_1, \ldots, b_v \in K$ with $u + v \leq s$, we have

$$a_1 + \ldots + a_u = b_1 + \ldots + b_v$$

if and only if

$$\varphi(a_1) + \ldots + \varphi(a_u) = \varphi(b_1) + \ldots + \varphi(b_v).$$

Denote by $z_1, z_2, \ldots, z_t$ the variables that occur in Φ. Let $g_1, g_2, \ldots, g_t$ be arbitrary elements of G and let $K = \{g_1, g_2, \ldots, g_t\}$, then $|K| \leq t$. Assume that K is $\tilde{F}_s$-isomorphic to some subset K' of $\mathbb{Z}$, and denote by φ the corresponding bijection. In G, substitute $z_i = g_i$ in Φ; in $\mathbb{Z}$, do the same with $z_i = \varphi(g_i)$. Then we get the same truth assignment in the case of each atomic sub-formula of Φ. Since Φ is valid in $\mathbb{Z}$, it follows that the above substitution makes Φ valid in G. Thus, it is enough to prove the following

Theorem 1.15. *Let K be a t-element subset of the finitely generated abelian group G. If $p(G) > s^t$ then there exists an $\tilde{F}_s$-isomorphism $\varphi : K \to K'$ for some set $K' \subseteq \mathbb{Z}$.*

The starting point is the following direct generalization of [13, Theorem 3.1] whose proof we include for the sake of completeness.

Lemma 1.16. *Let K be a t-element subset of $\mathbb{Z}_q$ where q is a power of a prime $p > s^t$. Then there exists a set of integers K' such that the canonical homomorphism $\mathbb{Z} \to \mathbb{Z}_q = \mathbb{Z}/q\mathbb{Z}$ induces an $\tilde{F}_s$-isomorphism of K' onto K.*

Proof. Identify the elements of K with the unique integers $0 \leq a_1, \ldots, a_t < q$ they represent. Let $\mathbf{e}_i$ $(0 \leq i \leq t)$ be the standard basis for $\mathbb{Z}^{t+1}$ and consider the lattice Λ generated by the vectors

$$\mathbf{e}_0 + \sum_{i=1}^{t} \frac{a_i}{q}\mathbf{e}_i, \ -\mathbf{e}_1, \ -\mathbf{e}_2 \ , \ldots, \ -\mathbf{e}_t.$$

The volume of the fundamental domain of Λ is 1. Since $p(1/s)^t > 1$, it follows from Minkowski's convex body theorem that Λ has a nonzero vector in the rectangular box

$$(-p, p) \times (-1/s, 1/s) \times \ldots \times (-1/s, 1/s),$$

that is, there are integers n_i, not all of them zero, such that $|n_0| < p$ and

$$\left| \frac{n_0 a_i}{q} - n_i \right| < \frac{1}{s}$$

for $1 \leq i \leq t$. Were $n_0 = 0$ it would imply $n_i = 0$ for $1 \leq i \leq t$. Thus we can conclude that n_0 is not divisible by p and that there are integers m_i such that $|m_i| < q/s$ and $n_0 a_i \equiv m_i \pmod{q}$. If r is any multiplicative inverse of n_0 modulo q, then $r m_i \equiv a_i \pmod{q}$, and thus the canonical homomorphism $\varphi : \mathbb{Z} \to \mathbb{Z}_q$ maps $K' = \{r m_1, \ldots, r m_t\}$ onto K. Moreover,

$$a_{i_1} + \ldots + a_{i_u} = a_{j_1} + \ldots + a_{j_v}$$

in $\mathbb{Z}_q$ if and only if

$$\sum_{k=1}^{u} a_{i_k} - \sum_{k=1}^{v} a_{j_k}$$

is divisible by q, which by $(n_0, q) = 1$ exactly happens if

$$\sum_{k=1}^{u} m_{i_k} - \sum_{k=1}^{v} m_{j_k}$$

is divisible by q. Since $|m_i| < q/s$, under the assumption that $u+v \le s$ this is equivalent with saying that the above expression is zero, or what is the same,

$$rm_{i_1} + \ldots + rm_{i_u} = rm_{j_1} + \ldots + rm_{j_v}.$$

This indicates that φ indeed induces an $\tilde{F}_s$-isomorphism. □

Since the identical map $\imath : \mathbb{Z} \to \mathbb{Z}$ obviously induces an $\tilde{F}_s$-isomorphism of any subset of $\mathbb{Z}$ onto itself, in view of the fundamental theorem of finitely generated abelian groups, to verify Theorem 1.15, it is enough to prove that whenever the theorem is true for the abelian groups G_1 and G_2, it is true for their direct sum $G = G_1 \oplus G_2$ as well. This we can do as follows. Assume that $p(G) > s^t$, then $p(G_i) > s^t$ for $i = 1, 2$. Let

$$K_1 = \{g \in G_1 \mid \exists h \in G_2 \text{ with } (g, h) \in K\},$$

and define K_2 in a similar way as the projection of K to G_2. Then $t_i = |K_i| \le |K| \le t$, so $s^{t_i} < p(G_i)$ and by our hypothesis there exist $\tilde{F}_s$-isomorphisms $\varphi_i : K_i \to K_i'$ for some appropriate t_i-element sets $K_i' \subset \mathbb{Z}$. With $m = \max\{|n| \ : \ n \in K_2'\}$ and with any integer $\alpha > sm$, define the map

$$\varphi : K_1 \times K_2 \to \{\alpha n_1 + n_2 \mid n_1 \in K_1', n_2 \in K_2'\}$$

by $\varphi((g, h)) = \alpha\varphi_1(g) + \varphi_2(h)$. Since $\alpha n_1 + n_2 = \alpha n_1' + n_2'$ implies that α divides the number $n_2 - n_2'$ whose modulus is not larger than $2m < \alpha$, that is, it implies $n_2 = n_2'$, and in turn also $n_1 = n_1'$, we find that φ is a bijection. A similar argument shows that φ is in fact an $\tilde{F}_s$-isomorphism, and thus its restriction to K is also an $\tilde{F}_s$-isomorphism. This completes the proof of Theorem 1.15 and in turn also that of Theorem 1.14. □

The above proof appeared in our expository paper [53]. Theorem 1.14 can be applied to all statements of Section 1.1, with $t = k$ or $t = k + \ell$ and $s = 4$ or, in the case of Statement 1.12, $s = k - 1$. It yields a bound that is exponentially large in k (and ℓ). Such a strong restriction on $p(G)$ is sometimes necessary, as it happens in the case of Statement 1.12, see [49]. In many cases, however, more effective results can be obtained. In Chapter 3–5 we study problems related to Statements 1.2–1.11, giving the ultimate answer in many cases.

Chapter 2

An Overview

In this chapter first we give an overview of our main results in the perspective of the relevant developments in the field. This is done, whenever appropriate, in the framework presented in the previous chapter. This is followed by a section in which we briefly explain the tools and methods we use, and how the dissertation is structured.

2.1 History and Results

In the context of Section 1.1, the Cauchy–Davenpont theorem claims that Statement 1.2 is valid in any cyclic group $\mathbb{Z}_p$ with a prime $p \geq k + \ell - 1$. Moreover, it is also valid in any abelian group G with $p(G) \geq k + \ell - 1$, according to Theorem 1.1. Most of the results that follow can be appreciated in a similar sense.

Unrestricted Set Addition

In addition to the already mentioned papers [19, 72], there are various further generalizations of the Cauchy–Davenpont theorem, see for example Shatrowsky [82], Pollard [73] and Yuzvinsky [88]. Kemperman [59] proved the analogue of Statement 1.2 in arbitrary (that is, not necessarily commutative) torsion free groups. In Chapter 5 we will prove that it is also valid in an arbitrary finite group G with $p(G) \geq k + \ell - 1$. Using multiplicative notation:

Theorem 2.1. *If A and B are nonempty subsets of a finite group G such that $p(G) \geq |A| + |B| - 1$, then $|AB| \geq |A| + |B| - 1$.*

It is easy to see that both the condition and the bound are sharp. Denote by $\mu_G(k, \ell)$ the minimum size of the product set AB where A and B range over all subsets of G of cardinality k and ℓ, respectively. For finite abelian groups G, the function μ_G has been exactly determined by Eliahou, Kervaire and Plagne [29]. Some partial results in the noncommutative case were

found recently by Eliahou and Kervaire [27, 28]. In particular, they proved the inequality $\mu_G(k,\ell) \le k+\ell-1$ for all possible values of k and ℓ when G is a finite solvable group. That equality holds here for $k+\ell-1 \le p(G)$, a case in which the upper bound is folklore, is the essence of the above theorem that we proved in [55].

The case of equality in the Cauchy–Davenport theorem was characterized by Vosper [87]. This first inverse theorem in the theory of set addition is the following.

Theorem 2.2. *If A, B are nonempty subsets of $\mathbb{Z}_p$ such that $|A+B| = |A|+|B|-1$, then either $|A|+|B|-1 = p$ (that is, $A+B = \mathbb{Z}_p$), or one of the sets A and B contains only one element, or $|A+B| = p-1$ and with the notation $\{c\} = \mathbb{Z}_p \setminus (A+B)$, B is the complement of the set $c-A$ in $\mathbb{Z}_p$, or both A and B are arithmetic progressions of the same difference.*

Hamidoune and Rødseth [48] go one step further; they characterize all pairs A, B with $|A+B| = |A|+|B|$.

In the special case when $A = B$, Vosper's theorem can be stated as

Theorem 2.3. *Let A be a set of k residue classes modulo a prime $p > 2k-1$. Then $|A+A| = 2k-1$ if and only if A is an arithmetic progression.*

An extension of Vosper's theorem to arbitrary abelian groups is due to Kemperman [60], who employed Kneser's theorem to obtain a recursive characterization of all critical pairs, that is, all pairs (A, B) with $|A+B| \le |A|+|B|-1$. For a related result, see Lev [64]. In particular, Theorem 2.3 can be extended as

Theorem 2.4. *Let A be a set of k elements of an abelian group G with $p(G) > 2k-1$. Then $|A+A| = 2k-1$ if and only if A is an arithmetic progression.*

That is, Statement 1.8 is valid whenever $p(G) \ge 2k$. In fact, Kemperman's result also implies that Statement 1.7 is true for $p(G) \ge k+\ell+1$.

Kneser's theorem cannot be extended to noncommutative groups in a natural way ([70, 89]), and the simple combinatorial proof does not work either. However, Vosper's theorem has been extended to torsion free groups by Brailovsky and Freiman [14]. A generalization to arbitrary noncommutative groups has been obtained by Hamidoune [45]. To state it, we first have to recall the following notion. Let B be a finite subset of a group G such that $1 \in B$. B is called a *Cauchy-subset* of G if, for every finite nonempty subset A of G,

$$|AB| \ge \min\{|G|, |A|+|B|-1\}.$$

If the group G is finite, then a subset S that contains the unit element is known to be a Cauchy subset if and only if for every subgroup H of G,

$$\min\{|SH|, |HS|\} \ge \min\{|G|, |H|+|S|-1\},$$

see Corollary 3.4 in [45]. Now Theorem 6.6 in the same paper can be stated as follows. (Here $\langle q \rangle$ denotes the subgroup generated by the element q.)

Theorem 2.5. *Let G be a finite group and let B be a Cauchy subset of G such that $|G|$ is coprime to $|B|-1$. Assume that $|AB| = |A| + |B| - 1 \leq |G| - 1$ holds for some subset A of G. Then either $|A| = 1$, or $A = G \setminus aB^{-1}$ for some $a \in G$, or there are elements $a, b, q \in G$ and natural numbers k, l such that*

$$A = \{a, aq, aq^2, \ldots, aq^{k-1}\} \quad \text{and} \quad B = (G \setminus \langle q \rangle b) \cup \{b, qb, q^2b, \ldots q^{l-1}b\}.$$

Since without any loss of generality we may assume in Vosper's theorem that $1 \in B$, and any such B with $|B| \geq 2$ is a Cauchy subset of $\mathbb{Z}_p$, Vosper's theorem follows immediately from the above result of Hamidoune. Note that if G is not a cyclic group of prime order, then a subset B of G with $2 \leq |B| \leq p(G)$ is not a Cauchy subset in general. Thus the following result of ours [55] gives a different kind of generalization of Vosper's inverse theorem, more in the spirit of Theorem 2.1.

Theorem 2.6. *Let A, B be subsets of a finite group G such that $|A| = k$, $|B| = \ell$ and $k + \ell - 1 \leq p(G) - 1$. Then $|AB| = k + \ell - 1$ if and only if one of the following conditions holds:*

(i) $k = 1$ *or* $\ell = 1$*;*

(ii) there exists $a, b, q \in G$ such that

$$A = \{a, aq, aq^2, \ldots, aq^{k-1}\} \quad \text{and} \quad B = \{b, qb, q^2b, \ldots q^{l-1}b\};$$

(iii) $k + \ell - 1 = p(G) - 1$ *and there exists a subgroup F of G of order $p(G)$ and elements $u, v \in G$, $z \in F$ such that*

$$A \subset uF, \ B \subset Fv \quad \text{and} \quad A = u(F \setminus zvB^{-1}).$$

Our proof of Theorems 2.1 and 2.6 depend heavily on the solvability of groups of odd order and the structure of group extensions. Very recently Ruzsa [80] found in an ingenious way alternative proofs of these results that do not rely on the Feit–Thompson theorem.

Another far reaching generalization of Vosper's inverse theorem is due to Freiman. The starting point is Freiman's so-called '$3k-4$' theorem [34, 37]:

Theorem 2.7. *Let A be a set of $k \geq 3$ integers. If $|A + A| = 2k - 1 + b \leq 3k - 4$, then A is contained in an arithmetic progression of length $k + b$.*

This again must be true in any abelian group G with $p(G)$ large enough compared to k. Freiman [35, 37] derived the following analogue for cyclic groups of prime order.

Theorem 2.8. *If A is a large enough k-element subset of $\mathbb{Z}_p$, p a prime, such that $k \leq p/35$ and $|A+A| = 2k-1+b \leq 2.4k-3$, then A is contained in an arithmetic progression of length $k+b$.*

Finally we mention Freiman's theorem [36, 37] asserting that if a finite set A of integers satisfies $|A + A| \ll |A|$, then A is contained in a 'generalized arithmetic progression' whose size and dimension is bounded in terms of the implied constant, see also Ruzsa [78, 79], Bilu [12] and Chang [18]. Very recently Green and Ruzsa [42] generalized the result to arbitrary abelian groups.

The Erdős–Heilbronn Problem

The case of restricted addition is apparently more difficult. In 1994 Dias da Silva and Hamidoune [22] proved the following analogue of the Cauchy–Davenport theorem, thus settling a problem of Erdős and Heilbronn (see [30, 32]).

Theorem 2.9. *If A is a k-element subset of the p-element group $\mathbb{Z}_p$, p a prime, then*

$$|A\dot{+}A| \geq \min\{p, 2k-3\}.$$

More generally, they proved

Theorem 2.10. *If A is any subset of the cyclic group $\mathbb{Z}_p$, then*

$$|\Sigma_d(A)| \geq \min\{p, d(|A|-d)+1\}.$$

These results were obtained via exterior algebra methods and the representation theory of the symmetric groups. Shortly afterwards Alon, Nathanson and Ruzsa [7, 8] applying the so-called 'polynomial method' gave a simpler proof that also yields

$$|A\dot{+}B| \geq \min\{p, |A|+|B|-2\}$$

if $|A| \neq |B|$. Some lower estimates on the cardinality of $A\dot{+}B$ in arbitrary abelian groups were obtained recently by Lev [65, 66], and also by Hamidoune, Lladó and Serra [47] in the case $A = B$. Some ramifications in elementary abelian p-groups have been explored in a series of papers by Eliahou and Kervaire [24, 25, 26].

In [52] we established the following extension of the Dias da Silva–Hamidoune theorem.

Theorem 2.11. *If A is a k-element subset of an abelian group G, then*

$$|A\dot{+}A| \geq \min\{p(G), 2k-3\}.$$

Thus, Statement 1.5 holds in every abelian group G with $p(G) \geq 2k-3$, and this result is sharp. The result of Alon, Nathanson and Ruzsa implies Statement 1.3 for $G = \mathbb{Z}_p$ if $p \geq k + \ell - 3$. For more than ten years it has been open, whether Statement 1.6 can be generalized the same way. We prove [56] that this is indeed the case:

Theorem 2.12. *Let $A \neq B$ be nonempty subsets of the additive group of a field of characteristic p. Then $|A\dot{+}B| \geq \min\{p, |A| + |B| - 2\}$.*

Thus, if A, B are nonempty subsets of an elementary abelian p-group, with $p \geq |A| + |B| - 2$, then $|A\dot{+}B| \geq |A|+|B|-3$, and equality can only be attained if $A = B$. As opposed to the case of unrestricted set addition, only partial results have been known about the case of equality here. First, if $p(G) \leq 2k-3$ and A is contained in a subgroup H of G with $|H| = p(G)$, then $|A\dot{+}A| = H$ in view of Theorem 2.9. Next, if $k \geq 2$, $p(G) \geq 2k-3$, and the elements of A form an arithmetic progression, then $A\dot{+}A$ is an arithmetic progression of length $2k-3$. Finally, assume that $p(G) > 2k-3$. If k is 2 or 3, then clearly $|A\dot{+}A| = 2k-3$. If k is 4, then $|A\dot{+}A|$ is either 5 or 6, where the first case happens if and only if $a+b = c+d$ for some order a, b, c, d of the elements of A. If $k \geq 5$ and G is torsion free, then $|A\dot{+}A| = 2k-3$ happens if and only if A is an arithmetic progression. As we have seen, Statement 1.9 must be true under the assumption that $p(G)$ is large enough. This has been first proved in $\mathbb{Z}/p\mathbb{Z}$, where p is a large enough prime, by Pyber [74]. The same is proved in [13] under the assumption that $p > ck$, where c is an effective constant. Further improvements can be derived from the works of Freiman, Low and Pitman [39] and Lev [65] in the case when k is large enough. Roughly speaking, under some assumptions on k and p they prove that if $|A\dot{+}A|$ is close to $2k-3$ then A is contained in a short arithmetic progression. In particular, Theorem 2 of Lev [65] can be stated as follows.

Theorem 2.13. *Let A be a k-element subset of $\mathbb{Z}_p$, p a prime, such that $200 \leq k \leq p/50$. If $k' = |A\dot{+}A| \leq 2.18k-6$, then A is contained in an arithmetic progression of length $k'-k+3$. In particular, if $|A\dot{+}A| = 2k-3$, then the elements of A form an arithmetic progression.*

That is, there is a general inverse theorem that parallels the Freiman–Vosper theorem (Theorem 2.8). Part of the proof depends on estimates with exponential sums, which explains why the (somewhat flexible) conditions on p and k enter the theorem.

Here we exploit an algebraic method to get rid of these unnecessary restrictions when $|A\dot{+}A| = 2k-3$. Probably the most important result in this dissertation is the following inverse counterpart of Theorem 2.9 that we obtained in [54].

Theorem 2.14. *Let A be a set of $k \geq 5$ residue classes modulo a prime $p > 2k-3$. Then $|A\dot{+}A| = 2k-3$ if and only if A is an arithmetic progression.*

In fact, with the help of ideas from [52, 53] we can transfer this result, first to cyclic groups of prime power order then to direct sums, in order to prove the following extension [54].

Theorem 2.15. *Let A be a set of $k \geq 5$ elements of an abelian group G with $p(G) > 2k-3$. Then $|A\dot{+}A| = 2k-3$ if and only if A is an arithmetic progression.*

It is clear from what we have said before, that the bounds on k and p, resp. $p(G)$ cannot be improved upon in the above theorems. In view of our remarks, Theorems 2.12 and 2.15 imply the following:

Corollary 2.16. *Let A, B be nonempty subsets of the additive group of a field of characteristic $p \geq |A| + |B| - 2$. Then $|A\dot{+}B| \geq |A| + |B| - 2$, unless $A = B$ and one of the following holds:*

(i) $|A| = 2$ *or* $|A| = 3$*;*

(ii) $|A| = 4$*, and* $A = \{a, a+d, c, c+d\}$*;*

(iii) $|A| \geq 5$*, and A is an arithmetic progression.*

Further developing some ideas from our papers [51, 52, 55], very recently Balister and Wheeler [11] established

$$|\{a\vartheta(b) \mid a \in, b \in B, a \neq b\}| \geq \min\{p(G), |A| + |B| - 3\}$$

for every finite group G and automorphism $\vartheta \in \mathrm{Aut}(G)$. It is quite plausible, that the above corollary can also be generalized in the very same spirit.

Snevily's Problem

A *transversal* of an $n \times n$ matrix is a collection of n cells, no two of which are in the same row or column. A transversal of a matrix is a *Latin transversal* if no two of its cells contain the same element. A conjecture of Snevily [83, Conjecture 1] asserts that, for any odd n, every $k \times k$ sub-matrix of the Cayley addition table of $\mathbb{Z}_n$ contains a Latin transversal. Putting it differently, for any two subsets A and B with $|A| = |B| = k$ of a cyclic group G of odd order $n \geq k$, there exist numberings $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ of the elements of A and B respectively such that the k sums $a_i + b_i$, $1 \leq i \leq k$, are pairwise different. In fact, this is also conjectured for arbitrary abelian groups G of odd order [83, Conjecture 3]. That is, Statement 1.10 must be valid in any finite abelian group G with $p(G) \geq 3$. The statement does not hold for cyclic groups of even order as shown, for example, by taking $A = B = G$, whereas for this choice it clearly holds when $|G|$ is odd (just take $a_i = b_i, i = 1, \ldots, n$). For arbitrary groups of even order take $A = B = \{0, g\}$, with g an involution, to get a counterexample. Here we first verify Snevily's conjecture for arbitrary cyclic groups of odd order.

Theorem 2.17. *Let G be a cyclic group of odd order. Let $A = \{a_1, a_2, \ldots, a_k\}$ and B be subsets of G, each of cardinality k. Then there is a numbering $b_1, \ldots, b_k$ of the elements of B such that the sums $a_1 + b_1, \ldots, a_k + b_k$ are pairwise different.*

Alon [3] proved the conjecture in the particular case when $n = p$ is a prime number. Actually he proved a stronger result which can be considered as a special case of the following result when $\alpha = 1$.

Theorem 2.18. *Let p be a prime number, α a positive integer and $G = \mathbb{Z}_{p^\alpha}$ or $G = (\mathbb{Z}_p)^\alpha$. Let $(a_1, \ldots, a_k)$, $k < p$, be a sequence of not necessarily distinct elements in G. Then, for any subset $B \subset G$ of cardinality k, there is a numbering $b_1, \ldots, b_k$ of the elements of B such that the sums $a_1 + b_1, \ldots, a_k + b_k$ are pairwise different.*

Note that the above theorem is not true with $k = p$ (see [3]). Putting it otherwise: if G is a finite elementary abelian group, or a cyclic group of prime power order, then Statement 1.11 is true, assuming $p(G) > k$.

The above results appeared in [20]. They are discussed, along with proofs, in the recent monograph of Tao and Vu [86], and were briefly indicated in the 2002 ICM talk of Alon [4]. Based on our methods, various generalizations were obtained by Sun and Yeh [84, 85]. Employing one of the results in our paper for group rings, Gao and Wang [41] proved that Statement 1.10 is valid in every finite abelian group G with $p(G) > k^2$. They also verified Statement 1.11 for finite abelian p-groups with $p > k^2/4$.

The Subset Sum Problem

Representing integers as the sum of some elements of a given set A of integers is a very old problem, which has many ramifications. Several interesting questions are discussed by Erdős and Graham in [32]. If A is sufficiently dense, then $\Sigma(A)$ contains long arithmetic progressions. This phenomenon has received a lot of attention lately, see for example the last chapter of the recent monograph by Tao and Vu [86]. The following result is due to Lev [63].

Theorem 2.19. *If $A \subset [1, \ell]$ is a set of n integers and $\ell \leq 3n/2 - 2$, then*

$$[2\ell - 2n + 1, \sigma(A) - (2\ell - 2n + 1)] \subseteq \Sigma(A).$$

Motivated by a possible extension, at the Workshop on Combinatorial Number Theory held at DIMACS, 1996, V.F. Lev proposed the following problem. Suppose that $1 \leq a_1 < a_2 < \ldots < a_n \leq 2n - 1$ are integers such that their sum $\sigma = \sum_{i=1}^{n} a_i$ is even. Does there always exist $I \subset \{1, 2, \ldots, n\}$ such that $\sum_{i \in I} a_i = \sigma/2$? The answer is in the affirmative if n is large enough. Note that such a restriction has to be imposed on n, since the sequences $(1, 4, 5, 6)$ and $(1, 2, 3, 9, 10, 11)$ provide counterexamples otherwise. The answer can be easily derived from the following theorem [57].

Theorem 2.20. *Let $1 \leq a_1 < a_2 < \ldots < a_n \leq 2n-1$ denote integers such that $a_{\nu+1} - a_\nu = 1$ holds for at least one index $1 \leq \nu \leq n - 1$. If $n \geq n_0 = 89$, then there exist $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, +1\}$ such that $|\varepsilon_1 + \ldots + \varepsilon_n| \leq 1$ and $|\varepsilon_1 a_1 + \ldots + \varepsilon_n a_n| \leq 1$.*

More generally, every integer in a long interval can be expressed as a 'balanced' subset sum:

Theorem 2.21. *If n is large enough and $1 \le a_1 < a_2 < \ldots < a_n \le 2n-2$ are integers, then for every integer*

$$k \in [\sigma/2 - n^2/24, \sigma/2 + n^2/24]$$

there exists a set of indices $I \subset \{1, 2, \ldots, n\}$ such that $|I| \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$ and $\sum_{i \in I} a_i = k$.

Lev conjectured that if n is sufficiently large, then Theorem 2.19 must remain true under the weaker condition $\ell \le 2n-c$ with a suitable constant c. Based on the Dias da Silva–Hamidoune theorem (Theorem 2.10) we verify this conjecture in the ultimate way [58].

Theorem 2.22. *If $A \subset [1, \ell]$ is a set of $n \ge n_0$ integers and $\ell \le 2n-6$, then*

$$[2\ell - 2n + 1, \sigma(A) - (2\ell - 2n + 1)] \subseteq \Sigma(A).$$

The example $A = [\ell - n, 2\ell - 2n - 1] \cup [2\ell - 2n + 1, \ell]$ demonstrates that the interval in the theorem cannot be extended, whereas $A = [n, 2n-1]$ certifies that the result is no longer valid with $\ell \ge 2n - 1$.

A different but closely related problem is the following. For positive integers ℓ and m, let $f(\ell, m)$ denote the maximum cardinality of a set $A \subset [1, \ell]$ such that $m \notin \Sigma(A)$. The study of this function was initiated by Erdős and Graham, see [31]. Clearly $f(\ell, m) \ge \ell/\mathrm{snd}(m)$, where $\mathrm{snd}(m)$ denotes the smallest positive integer that does not divide m. In [1], Alon proved that $f(\ell, m) \le c(\varepsilon) \cdot \ell/\mathrm{snd}(m)$ for every $\ell^{1+\varepsilon} < m < \ell^2/\ln^2 \ell$, and conjectured that in fact $f(\ell, m) = (1 + o(1)) \cdot \ell/\mathrm{snd}(m)$ holds for $\ell^{1.1} < m < \ell^{1.9}$ as $\ell \to \infty$. This was verified by Lipkin [68] in the range $\ell \ln \ell < m < \ell^{3/2}$. Finally Alon and Freiman [5] determined the exact value of $f(\ell, m)$ as

$$f(\ell, m) = \Big\lfloor \frac{\ell}{\mathrm{snd}(m)} \Big\rfloor + \mathrm{snd}(m) - 2$$

for every $\varepsilon > 0$, $\ell > \ell_0(\varepsilon)$ and m satisfying $3\ell^{5/3+\varepsilon} < m < \ell^2/20 \ln^2 \ell$. The proof of these results employed the Hardy–Littlewood circle method. It turns out that one can replace the circle method by subtle combinatorics to solve this problem completely. Our first solution was based on the ideas we employed to prove Theorem 2.22. A slightly better result can be obtained, however, by the following theorem of Lev [67]. For any positive integer q we denote by $N_q(A)$ the number of elements in A that are not divisible by q.

Theorem 2.23. *Let A be a set of $n \ge n_0$ integers in the interval $[1, \ell]$, where $n \ge 20(\ell \ln n)^{1/2}$, and let $\lambda = 280\ell/n^2$. Then there exists a positive integer $d < 2\ell/n$ such that $\Sigma(A)$ contains all multiples of d that belong to the interval*

$$[\lambda\sigma(A), (1 - \lambda)\sigma(A)].$$

Moreover, if $N_q(A) \geq q-1$ *holds for every positive integer* $q < 2\ell/n$*, then the statement is valid with* $d = 1$*.*

See Freiman [38] and Sárközy [81] for the forerunners of this result. Lev [67] notes that the above theorem is essentially best possible in many respects. In [58] we give the following refinement.

Theorem 2.24. *Let* A *be a set of* $n \geq n_1$ *integers in the interval* $[1, \ell]$*, where*

$$n > \frac{\ell}{d} + d - 2 \quad \textit{for some integer} \quad 2 \leq d \leq \frac{n}{400 \ln n}.$$

Then there exists an integer $t \in [1, d-1]$ *such that* $\Sigma(A)$ *contains all multiples of* t *that belong to the interval* $[280d\ell, \sigma(A) - 280d\ell]$*.*

It is clear, that the theorem is now best possible also in regard to the common difference t of the long homogeneous arithmetic progression contained in $\Sigma(A)$. An almost immediate consequence is the following ultimate solution to the conjecture of Alon.

Theorem 2.25. *For every* $\varepsilon > 0$*, there is an* $\ell_0 = \ell_0(\varepsilon)$ *such that if* $\ell \geq \ell_0$*, then*

$$f(\ell, m) = \left\lfloor \frac{\ell}{\mathrm{snd}(m)} \right\rfloor + \mathrm{snd}(m) - 2$$

holds for any $(280 + \varepsilon)\ell \ln \ell < m < \ell^2/(8 + \varepsilon) \ln^2 \ell$*.*

2.2 Methods and Tools

The most frequently applied and highly developed methods in the structural theory of set addition are Kneser's theorem, the method of exponential sums, the isoperimetric method, and most recently also the polynomial method. A broad perspective of these methods can be gained from the book of Nathanson [69]. Our work during the last decade was highly influenced by the latter, which we briefly discuss below.

The Polynomial Method

The roots of this method go back as much as to Rédei, who used polynomials to study extremal problems in finite geometries. The idea has also occurred several times later, see e.g. Brouwer and Schrijver [15], Alon and Tarsi [9, 10], Alon and Füredi [6] and of course the already cited papers of Alon, Nathanson and Ruzsa. A major breakthrough is due to Alon [2], who formulated the following two theorems that can be applied directly in various situations.

Crucial to our work is the so-called Combinatorial Nullstellensatz. It is a simple consequence of a division algorithm for multivariate polynomials; it can be also viewed as a special case of Lasker's unmixedness theorem, see e.g. [23].

Theorem 2.26. *Let F be an arbitrary field and let $f = f(x_1, \ldots, x_k)$ be a polynomial in $F[x_1, \ldots, x_k]$. Let $S_1, \ldots, S_k$ be nonempty finite subsets of F and $g_i(x_i) = \prod_{s \in S_i}(x_i - s)$. If $f(s_1, s_2, \ldots, s_k) = 0$ for all $s_i \in S_i$, then there exist polynomials $h_1, h_2, \ldots, h_k \in F[x_1, \ldots, x_k]$ satisfying $\deg(h_i) \leq \deg(f) - \deg(g_i)$ such that $f = \sum_{i=1}^{k} h_i g_i$.*

In comparison with Hilbert's Nullstellensatz, three observations are in due order. First, the field F need not be algebraically closed, which is a convenience but not crucial in our proofs. Next, it is inherent in the above theorem that the ideal generated by the polynomials g_i is a radical ideal, this is the truly algebraic explanation why we can express f, instead of some unknown power of it, in the desired form. It is also very important to us that we have an explicit bound on the degree of coefficient polynomials h_i coming form the division algorithm.

An immediate consequence of the above theorem is what is often referred to as the polynomial lemma:

Theorem 2.27. *Let F be an arbitrary field and let $f = f(x_1, \ldots, x_k)$ be a polynomial in $F[x_1, \ldots, x_k]$. Suppose that $\prod_{i=1}^{k} x_i^{t_i}$ is a monomial such that $\sum_{i=1}^{k} t_i$ equals the degree of f and whose coefficient in f is nonzero. Then, if $S_1, \ldots, S_k$ are subsets of F with $|S_i| > t_i$ then there are $s_1 \in S_1, s_2 \in S_2, \ldots, s_k \in S_k$ such that $f(s_1, \ldots, s_k) \neq 0$.*

It can be also applied to derive the Chevalley–Warning theorem, which is a frequently used tool in zero-sum combinatorics. See [2] for a survey of applications. The most beautiful example in this direction is due to Rónyai [77] that led to the recent solution of Kemnitz's conjecture by Reiher [75].

Although the polynomial method has already demonstrated its power in the additive theory, to our best knowledge our paper [54] is the first instance when a structure theorem is obtained via this method. The polynomial lemma is a very convenient tool and has been widely applied for various problems in extremal combinatorics during the last decade. Direct applications of the Combinatorial Nullstellensatz appear to be a lot more complicated. Its strength over the polynomial lemma, informally speaking, lies in the fact that applying the latter we extract information encoded in one particular coefficient of a suitable polynomial, whereas applying Theorem 2.26 we have access to much more information encoded in a maze of coefficients.

A Brief Overview of the Contents

In Chapter 1 we generalized the rectification principle of Freiman, a minor contribution. The main novelties of our work are

- the application of the polynomial method in a multiplicative setting that led to the solution of a problem of Snevily, the extension of a result of Alon, and a generalization of the Erdős–Heilbronn conjecture to cyclic groups of prime power order;

- the application of the Combinatorial Nullstellensatz to obtain structural theorems related to the Erdős–Heilbronn problem;
- the application of group extensions to obtain results in the theory of set addition for more general, even noncommutative groups;
- and the application of elementary combinatorial arguments in conjunction with the Dias da Silva–Hamidoune theorem to prove some conjectures of Alon and Lev related to the subset sum problem.

These methods are respectively the main themes of the four main chapters that follow. Accordingly, in Chapter 3 we prove Theorems 2.17, 2.18 and a generalization of Theorem 2.9 to cyclic groups of prime power order. To study a few more examples we apply elementary algebraic number theory. We exploit some basic properties of cyclotomic fields, and the fact that the multiplicative group of any finite field is cyclic.

Chapter 4 is devoted to the proof of Theorems 2.12 and 2.14. As a by-product we get an independent proof of Theorem 2.3. Besides the Combinatorial Nullstellensatz only the notion of algebraic closure and the Vièta formulae are needed.

Theorems 2.1, 2.6, 2.11 and 2.15 are proved in Chapter 5, which also includes a self-contained proof of Theorem 2.4. We depend on the structure theory of finitely generated abelian groups, the Jordan–Hölder theorem, the structure of group extension in general, and in particular that of cyclic extensions, the Feit–Thompson theorem, Vosper's inverse theorem, and a result of Hamidoune (Theorem 2.5).

In Chapter 6 we derive among others, Theorems 2.20, 2.21, 2.22, 2.24 and 2.25. In addition to the Dias da Silva–Hamidoune theorem (Theorem 2.10) we use Theorems 2.19 and 2.23 of Lev, and we rely on the prime number theorem.

Chapter 3

The Polynomial Method

The main objective of this chapter is to prove the results related to the problems of Snevily. This is done by the application of the polynomial lemma in a multiplicative setting. A similar idea can be used in relation to restricted set addition to extend the Dias da Silva–Hamidoune theorem to cyclic groups whose order is a power of a prime.

3.1 Snevily's Problem

Following Alon's approach, our starting point will be the polynomial lemma (Theorem 2.27). For the case $G = (\mathbb{Z}_p)^\alpha$ the proof of Theorem 2.18 is almost the same as the one given by Alon in [3] which we sketch here to demonstrate the method. Identify the group $G = (\mathbb{Z}_p)^\alpha$ with the additive group of finite field $\mathbb{F}_q$ of order $q = p^\alpha$. Consider the polynomial

$$\begin{aligned} f(x_1, \ldots, x_k) &= \prod_{1 \le j < i \le k} ((x_i - x_j)(a_i + x_i - a_j - x_j)) \\ &= \prod_{1 \le j < i \le k} ((x_i - x_j)(x_i - x_j)) + \text{terms of lower degree.} \end{aligned}$$

The degree of f is $k(k-1)$ and the coefficient of $\prod_{i=1}^k x_i^{k-1}$ in f is $c = (-1)^{\binom{k}{2}} k!$ as we will see it in the following subsection. Since the characteristic of the field is $p > k$, it follows that c is a nonzero element. By applying Theorem 2.27 with $t_i = k - 1$ and $S_i = B$ for $i = 1, \ldots, k$, we obtain elements $b_1, \ldots b_k \in B$ such that

$$\prod_{1 \le j < i \le k} ((b_i - b_j)(a_i + b_i - a_j - b_j)) \neq 0.$$

Therefore, the elements $b_1, \ldots, b_k$ are pairwise distinct and so are the k sums $b_1 + a_1, \ldots, b_k + a_k$. This completes the proof for $G = (\mathbb{Z}_p)^\alpha$. □

So far we only have exploited the additive structures of finite fields; and it is clear that $(\mathbb{Z}_p)^\alpha$ are the only groups that can be treated this way. On the other hand, every cyclic group is

the subgroup of the multiplicative group of certain fields, and there exists a multiplicative analogue of the above described method, which is worked out in the first subsection. We apply this method to obtain Theorems 2.17 and 2.18 in the subsection that follows. In the remaining part of this section we study the possibility of further extending these results. In particular, we attempt to attack another conjecture of Snevily [83, Conjecture 2], namely that, if n is even, a $k \times k$ sub-matrix of the Cayley addition table of $\mathbb{Z}_n$ contains a Latin transversal unless k is an even divisor of n and the rows and columns of the sub-matrix are each cosets of the unique subgroup of order k in $\mathbb{Z}_n$.

The multiplicative analogue

In this subsection we study how to modify Alon's method if we wish to identify G with a subgroup of the multiplicative group of a suitable field. This will reduce the original problems to the study of permanents of certain Vandermonde matrices. Denote by $V(y_1, \ldots, y_k)$ the Vandermonde matrix

$$V(y_1, \ldots, y_k) = \begin{pmatrix} 1 & y_1 & \ldots & y_1^{k-1} \\ 1 & y_2 & \ldots & y_2^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & y_k & \ldots & y_k^{k-1} \end{pmatrix}.$$

For a matrix $M = (m_{ij})_{1 \le i,j \le k}$, the permanent of M is

$$\mathrm{Per}M = \sum_{\pi \in S_k} m_{1\pi(1)} m_{2\pi(2)} \ldots m_{k\pi(k)}.$$

Lemma 3.1. *Let F be an arbitrary field and suppose that $\mathrm{Per}V(a_1, \ldots, a_k) \neq 0$ for some elements $a_1, a_2, \ldots, a_k \in F$. Then, for any subset $B \subset F$ of cardinality k there is a numbering $b_1, \ldots, b_k$ of the elements of B such that the products $a_1 b_1, \ldots, a_k b_k$ are pairwise different.*

Proof. Consider the following polynomial in $F[x_1, \ldots, x_k]$

$$f(x_1, \ldots, x_k) = \prod_{1 \le j < i \le k} ((x_i - x_j)(a_i x_i - a_j x_j)).$$

The degree of f is clearly not greater than $k(k-1)$. In addition,

$$\begin{aligned} f(x_1, \ldots, x_k) &= \mathrm{Det}V(x_1, \ldots, x_k) \cdot \mathrm{Det}V(a_1 x_1, a_2 x_2, \ldots, a_k x_k) \\ &= \left(\sum_{\pi \in S_k} (-1)^{I(\pi)} \prod_{i=1}^{k} x_{\pi(i)}^{(i-1)} \right) \left(\sum_{\tau \in S_k} (-1)^{I(\tau)} \prod_{i=1}^{k} (a_{\tau(i)} x_{\tau(i)})^{(i-1)} \right) \\ &= \left(\sum_{\pi \in S_k} (-1)^{I(\pi)} \prod_{i=1}^{k} x_{\pi(i)}^{(i-1)} \right) \left(\sum_{\tau \in S_k} (-1)^{I(\tau)} \prod_{i=1}^{k} (a_{\tau(k+1-i)} x_{\tau(k+1-i)})^{(k-i)} \right) \\ &= \left(\sum_{\pi \in S_k} (-1)^{I(\pi)} \prod_{i=1}^{k} x_{\pi(i)}^{(i-1)} \right) \left(\sum_{\tau \in S_k} (-1)^{\binom{k}{2} - I(\pi)} \prod_{i=1}^{k} (a_{\pi(i)} x_{\pi(i)})^{(k-i)} \right). \end{aligned}$$

Therefore, the coefficient $c(a_1,\ldots,a_k)$ of the monomial $\prod_{i=1}^k x_i^{k-1}$ in f,

$$\begin{aligned} c(a_1,\ldots,a_k) &= \sum_{\pi\in S_k}(-1)^{\binom{k}{2}}\prod_{i=1}^k a_{\pi(i)}^{k-i} \\ &= (-1)^{\binom{k}{2}}\sum_{\pi\in S_k}\prod_{i=1}^k a_{\pi(k+1-i)}^{i-1} \\ &= (-1)^{\binom{k}{2}}\sum_{\tau\in S_k}\prod_{i=1}^k a_{\tau(i)}^{i-1} \\ &= (-1)^{\binom{k}{2}}\mathrm{Per}V(a_1,\ldots,a_k) \end{aligned}$$

is different from 0 (in particular, $c(1,\ldots,1)=(-1)^{\binom{k}{2}}k!$). Consequently, f is of degree $k(k-1)$, and we can apply Theorem 2.27 with $t_i=k-1$ and $S_i=B$ for $i=1,\ldots,k$ to obtain k distinct elements $b_1,\ldots,b_k$ in B such that the products $a_1b_1,\ldots,a_kb_k$ are pairwise distinct. This completes the proof of the lemma. □

Proof of the Theorems

Proof of Theorem 2.17. Write $|G|=m$ and let $\alpha=\phi(m)$, where ϕ is Euler's totient function; then $2^\alpha\equiv 1 \pmod m$. Consider $F=\mathbb{F}_{2^\alpha}$, its multiplicative group $F^\times$ is a cyclic group of order $2^\alpha-1$. Thus, G can be identified with a subgroup of $F^\times$, the operation on G being the restriction of the multiplication in F. Since F is of characteristic 2, we have

$$\mathrm{Per}V(a_1,\ldots,a_k)=\mathrm{Det}V(a_1,\ldots,a_k)=\prod_{1\le j<i\le k}(a_i-a_j)\neq 0\ .$$

The result follows immediately from Lemma 3.1. □

Proof of Theorem 2.18 for $G=\mathbb{Z}_{p^\alpha}$. Consider the cyclotomic field $F=\mathbb{Q}(\xi)$, where ξ is a primitive q^{th} root of unity and $q=p^\alpha$. The degree of this extension is $[\mathbb{Q}(\xi):\mathbb{Q}]=p^\alpha-p^{\alpha-1}$. Identify G with the multiplicative subgroup $\{1,\xi,\xi^2,\ldots,\xi^{q-1}\}$ of $\mathbb{Q}(\xi)$. As before, the result would be an immediate consequence of the fact $\mathrm{Per}V(a_1,\ldots,a_k)\neq 0$. To verify this fact, note that each term $\prod_{i=1}^k a_{\tau(i)}^{i-1}$ of this permanent is a q^{th} root of unity. Thus, $\mathrm{Per}V(a_1,\ldots,a_k)$ is the sum of q^{th} roots of unity, where the number of summands, $k!$, is not divisible by p. Therefore, it is enough to prove the following lemma.

Lemma 3.2. *If $\epsilon_1,\ldots,\epsilon_t$ are q^{th} roots of unity such that $\sum_{i=1}^t\epsilon_i=0$, then t is divisible by p.*

Lemma 3.2 follows from the more precise statement in Lemma 3.3 below. Let $\omega_p=e^{2\pi i/p}$. For each $\eta\in F$ such that $\eta^q=1$ we have $\sum_{i=1}^p\eta\omega_p^i=\eta\sum_{i=1}^p\omega_p^i=0$. We say that a set $X=\{\epsilon_1,\ldots,\epsilon_p\}$ of q^{th} roots of unity is *simple* if there is $\eta\in F$ with $\eta^q=1$ such that $X=\{\eta\omega_p,\eta\omega_p^2,\ldots,\eta\omega_p^p\}$.

Lemma 3.3. *Let $\epsilon_i, i \in I$ be q^{th} roots of unity such that $\sum_{i\in I} \epsilon_i = 0$. Then there is a partition $I = \cup J_r$ such that $\{\epsilon_j \mid j \in J_r\}$ is a simple set for each r.*

Proof. Consider $V = \mathbb{Q}(\xi)$ as a vector space over $\mathbb{Q}$. The dimension of V is $\phi(q) = p^\alpha - p^{\alpha-1}$. Let, for $0 \le s \le q-1$, $K_s = \{i \mid \epsilon_i = \xi^s\}$, and write $c_s = |K_s|$. Let $s \equiv \bar{s} \bmod p^{\alpha-1}$, $0 \le \bar{s} < p^{\alpha-1}$. Note that $\{\xi^s, \xi^{s+p^{\alpha-1}}, \ldots, \xi^{s+(p-1)p^{\alpha-1}}\}$ is a simple set for every $0 \le s < p^{\alpha-1}$. Thus,

$$0 = \sum_{i\in I} \epsilon_i = \sum_{s=0}^{q-1} c_s\xi^s = \sum_{s=0}^{q-1} c_s\xi^s - \sum_{s=0}^{p^{\alpha-1}-1} c_s(\xi^s + \xi^{s+p^{\alpha-1}} + \ldots + \xi^{s+(p-1)p^{\alpha-1}})$$

$$= \sum_{s=0}^{q-1}(c_s - c_{\bar{s}})\xi^s = \sum_{s=p^{\alpha-1}}^{q-1} (c_s - c_{\bar{s}})\xi^s \ .$$

Since $\{1, \xi, \xi^2, \ldots, \xi^{\phi(q)-1}\}$ is a basis of V, $\{\xi^s \mid p^{\alpha-1} \le s \le p^\alpha - 1\}$ is also an independent set. Thus, $c_s = c_{\bar{s}}$ for every $0 \le s \le q-1$. Each set J_r of the desired partition of I can then be obtained by choosing one element in each one of the sets $K_s, K_{s+p^{\alpha-1}}, \ldots, K_{s+(p-1)p^{\alpha-1}}$, for every choice of s, $0 \le s < p^{\alpha-1}$ such that $K_s \neq \emptyset$. □

Since every simple set has exactly p elements, Lemma 3.2 follows and the proof is complete. □

The following short proof of Lemma 3.2 was suggested by Imre Ruzsa. There exist positive integers α_i with $\varepsilon_i = \varepsilon^{\alpha_i}$. Consider the polynomial $R(x) = \sum_{i=1}^t x^{\alpha_i}$, then $R(\varepsilon) = 0$. It follows that the q^{th} cyclotomic polynomial Φ_q, which is irreducible in $\mathbb{Z}[x]$, is a divisor of R in the ring $\mathbb{Z}[x]$. Consequently, $p = \Phi_q(1)$ divides $R(1) = t$.

Bad sequences

One of Snevily's yet unsolved conjectures asserts that the statement of Theorem 2.17 holds whenever $|G|$ is not divisible by 2. We believe that the statement of Theorem 2.18 is always true if the smallest prime divisor of $|G|$ exceeds k. We also believe that the structure of the counterexamples in other cases cannot be arbitrary, see Problem 3.7 below.

Let G be any abelian group and $A = (a_1, a_2, \ldots, a_k)$, $k \le |G|$, be any sequence of group elements. A is said to be a *bad* sequence if there is a subset $B \subset G$ of cardinality k such that, for any numbering $b_1, \ldots, b_k$ of the elements of B, there are $1 \le i < j \le k$ such that $a_i + b_i = a_j + b_j$. Assume that G is a subgroup of the multiplicative group of some field F. It follows from Lemma 3.1 that A cannot be bad if $\mathrm{Per}V(a_1, \ldots, a_k) \neq 0$ in F. It is possible that a better understanding of permanents of Vandermonde matrices may even help in the characterization of bad sets. We will illustrate this point with the study of the cases $k = 2, 3$. There must be, however, certain limitations to this approach, as shown by the following example.

Example 3.4. *Suppose that $G \cong \mathbb{Z}_8$ is the subgroup of the multiplicative group of some field, and $A = \{a_1 = 1, a_2 = g^2, a_3 = g^3\}$ where g is a generator for G. Then* $\mathrm{Per}V(a_1, a_2, a_3) = 0$ *although A is not a bad sequence.*

Proof. Writing additively $A = \{0, 2, 3\}$, a short case analysis based on the number of even/odd elements of $B \subset G$, $|B| = 3$ shows that a required numbering b_1, b_2, b_3 of the elements of B always exists. On the other hand,

$$\mathrm{Per}V(a_1, a_2, a_3) = \mathrm{Per}\begin{pmatrix} 1 & 1 & 1 \\ 1 & g^2 & g^4 \\ 1 & g^3 & g^6 \end{pmatrix} = g^2(1 + g + g^2)(1 + g^4) = 0 \ ,$$

given that $g^4 = -1$. □

Next we give a complete description of the bad sequences of length ≤ 3 in cyclic groups.

Example 3.5. *Characterization of the bad sequences in the case $k = 2$.*

Identify $G \cong \mathbb{Z}_n$ with a subgroup of $\mathbb{C}^\times$, as in the proof of Theorem 2.18. Let ϵ, η be n^{th} roots of unity. Then $\mathrm{Per}V(\epsilon, \eta) = \epsilon + \eta = 0$ if and only if $\eta = -\epsilon = \omega_n^{n/2}\epsilon$. Consequently, $A = (a_1, a_2)$ can be a bad sequence in $\mathbb{Z}_n$ only if n is even and $a_2 = a_1 + n/2$, in which case it is indeed a bad sequence.

Example 3.6. *Characterization of the bad sequences in the case $k = 3$.*

Again we identify $G \cong \mathbb{Z}_n$ with a subgroup of $\mathbb{C}^\times$. Let ϵ, η, ζ be n^{th} roots of unity, $n \geq 3$. In this case $\mathrm{Per}V(\epsilon, \eta, \zeta) = 0$ if and only if

$$(\epsilon + \eta)(\eta + \zeta)(\zeta + \epsilon) = 2\epsilon\eta\zeta \ ,$$

that is,

$$(1 + x)(1 + y)(1 + z) = 2 \tag{3.1}$$

where $x = \eta/\epsilon, y = \zeta/\eta, z = \epsilon/\zeta$ are all n^{th} roots of unity and $xyz = 1$.

Recall (see e.g. [62]) that for ω a primitive n^{th} root of unity ($n > 1$), the norm of $1 - \omega$ in the n^{th} cyclotomic field $\mathbb{Q}_n = \mathbb{Q}(\omega)$ is

$$N_{\mathbb{Q}_n/\mathbb{Q}}(1 - \omega) = \prod_{\substack{1 \leq j < n \\ (j,n)=1}} (1 - \omega^j) = \begin{cases} 1 & \text{if } n \text{ is not a prime power,} \\ p & \text{if } n \text{ is a power of the prime } p. \end{cases}$$

Moreover, $-\omega$ is also a primitive n^{th} root of unity if n is even and a primitive $(2n)^{th}$ root of unity otherwise. Consequently,

$$N_{\mathbb{Q}_{2n}/\mathbb{Q}}(1 + \omega) = \begin{cases} 2^{\phi(2n)} & \text{if } \omega = 1, \\ 0 & \text{if } \omega = -1, \\ 2^{\phi(2n)/2^{\alpha-1}} & \text{if } \omega \text{ is a primitive } (2^\alpha)^{th} \text{ root of unity, } \alpha \geq 2, \\ 1 & \text{otherwise.} \end{cases}$$

By the multiplicative property of the norm, equality (3.1) can hold only if

- one of x, y, z (say x) is 1, or
- one of x, y, z (say x) is a primitive 4^{th} root of unity, while y and z are primitive 8^{th} roots of unity.

In the first case we have $\epsilon = \eta$, and with $u = \zeta/\epsilon$,

$$\mathrm{Per}V(\epsilon, \eta, \zeta) = \epsilon^3 \mathrm{Per}V(1, 1, u) = 2\epsilon^3(1 + u + u^2)$$

is 0 if and only if u is a primitive 3^{rd} root of unity, in which case (ϵ, η, ζ) is indeed a bad sequence. In the second case

$$\mathrm{Per}V(\epsilon, \eta, \zeta) = \epsilon^3 \mathrm{Per}V(1, x, xy)) = \epsilon^3((x-1) - y^2(1+x))$$

is 0 if and only if $y^2 = x = \pm i$. This, however, yields no bad sequences, see Example 3.4.

Consequently, $A = (a_1, a_2, a_3)$ is a bad sequence in $\mathbb{Z}_n$ if and only if n is divisible by 3, and for some permutation (i, j, k) of the indices $(1, 2, 3)$, $a_i = a_j = a_k \pm n/3$.

These results could have certainly been obtained without any algebraic consideration. We only worked them out to indicate that there may be further applications of our method. The above calculations also yield to an alternative proof of Theorem 2.17, and suggest that being bad is a local property.

2^{nd} proof of Theorem 2.17. Identify $G \cong \mathbb{Z}_n$ with a subgroup of $\mathbb{C}^\times$ and suppose $a_1, a_2, \ldots, a_k$ are all n^{th} roots of unity, n odd. Note that $\mathrm{Per}V(a_1, \ldots, a_k) = \mathrm{Det}V(a_1, \ldots, a_k) + 2A = \prod_{1 \le j < i \le k}(a_i - a_j) + 2A$, where $A \in \mathbb{Q}_n$ is an algebraic integer. Were $\mathrm{Per}V(a_1, \ldots, a_k) = 0$ we would have $\prod_{1 \le j < i \le k}(1 - a_j/a_i) = 2B$ with an algebraic integer $B \in \mathbb{Q}_n$. The norm of the right hand side in $\mathbb{Q}_n$ is divisible by $N_{\mathbb{Q}_n/\mathbb{Q}}(2) = 2^{\phi(n)}$. On the other hand, if a_j/a_i is a primitive m^{th} root of unity for some divisor m of n, then $N_{\mathbb{Q}_n/\mathbb{Q}}(1 - a_j/a_i) = (N_{\mathbb{Q}_m/\mathbb{Q}}(1 - a_j/a_i))^{\phi(n)/\phi(m)}$ is an odd integer, unless $m = 1$. Consequently, $(a_1, a_2, \ldots, a_k)$ cannot be a bad sequence, unless there are indices $1 \le j < i \le k$ with $a_i = a_j$. □

Problem 3.7. *Is it true that, if $A = (a_1, a_2, \ldots, a_k)$ is a bad sequence in an abelian group G, then there exists a subgroup $H \le G$ with $|H| = k$, a bad sequence $A' = (a'_1, a'_2, \ldots, a'_k)$ in H, and an element $c \in G$ such that $a_i = a'_i + c$ for every $1 \le i \le k$?*

If true, it would settle down Snevily's other conjectures mentioned in the introduction. Indeed, assume that the answer is yes. Let first G be any abelian group of odd order which contains a bad set $A = \{a_1, \ldots a_k\}$. It follows that $\{a'_1, \ldots, a'_k\}$ is a bad set in a k-element subgroup H of G. That is, H itself is a bad set in H, a contradiction, since k is odd. Thus, Snevily's conjecture [83, Conjecture 3] follows. Next, let $A = \{a_1, \ldots a_k\}$ be a bad set in $\mathbb{Z}_n$, n even. Then again, $A' = H$ is a bad set in $H \cong \mathbb{Z}_k$, which can only happen if k is even. Moreover, A is a translate of $A' = H$, implying [83, Conjecture 2] as well.

3.2 Restricted Addition in Cyclic Groups of Prime Power Order

In this section we prove that Statement 1.3 is valid in cyclic groups whose order is a power of a prime $p \geq k + \ell - 3$.

Theorem 3.8. *Let $A, B \subseteq \mathbb{Z}/q\mathbb{Z}$, where $q = p^\alpha$ is a power of a prime p. Then*

$$|A\dot{+}B| \geq \min\{p, |A| + |B| - 3\}.$$

Proof. We may clearly assume that $|A| = k \geq 2$ and $|B| = \ell \geq 2$. Since $A' \supseteq A$ and $B' \supseteq B$ implies $|A'\dot{+}B'| \geq |A\dot{+}B|$, we also may assume that $k+\ell-3 \leq p$. Our proof will again depend on the polynomial lemma.

Like in the previous section, we will use this lemma in a multiplicative setting. Let $\varepsilon = e^{2\pi i/q}$ and consider the unique embedding $\varphi : G \hookrightarrow \mathbb{C}^\times$ of G into the multiplicative group of the field of complex numbers with the property $\varphi(1) = \varepsilon$. Write $C = A\dot{+}B$ and define

$$\tilde{A} = \{\varphi(a) \mid a \in A\},\ \tilde{B} = \{\varphi(b)^{-1} \mid b \in B\},\ \tilde{C} = \{\varphi(c) \mid c \in C\}.$$

Observe that for $a \in A$ and $b \in B$,

$$a = b \Longleftrightarrow \varphi(a)\varphi(b)^{-1} - 1 = 0$$

and

$$a + b = c \Longleftrightarrow \varphi(a) - \varphi(c)\varphi(b)^{-1} = 0.$$

Thus, if $x \in \tilde{A}$ and $y \in \tilde{B}$, then either $xy - 1 = 0$, or there exists a $c \in \tilde{C}$ such that $x - cy = 0$.

We wish to prove that $|C| \geq k + \ell - 3$. Assume that on the contrary, $|C| = |\tilde{C}| \leq k + \ell - 4$, and choose any set $\tilde{C}' \subseteq G$, of cardinality $k + \ell - 4$, that contains $\tilde{C}$. Consider the polynomial $P \in \mathbb{C}[x, y]$ defined as

$$P(x, y) = (xy - 1) \prod_{c \in \tilde{C}'} (x - cy),$$

then $P(x, y) = 0$ for every $x \in \tilde{A}$, $y \in \tilde{B}$. Since the degree of P is clearly not greater than $k + \ell - 2$, in view of Lemma 2.27, the desired contradiction comes from the fact that the coefficient of the monomial $x^{k-1}y^{\ell-1}$ in P is different from 0.

To verify this fact, observe that writing $\tilde{C}' = \{c_1, c_2, \ldots, c_{k+\ell-4}\}$, this coefficient is

$$\mathrm{coeff}_P(x^{k-1}y^{\ell-1}) = (-1)^{\ell-2}Q(c_1, c_2, \ldots, c_{k+\ell-4}),$$

where $Q(x_1, x_2, \ldots, x_{k+\ell-4})$ is the $(\ell-2)^{nd}$ elementary symmetric polynomial in the variables $x_1, \ldots, x_{k+\ell-4}$. In particular, $Q(c_1, c_2, \ldots, c_{k+\ell-4})$ is the sum of $\binom{k+\ell-4}{\ell-2}$ numbers, each of which is a product of $\ell - 2$ terms. These terms, each being equal to some c_i, are all elements

of $\varphi(G)$. Consequently, each of the $\binom{k+\ell-4}{\ell-2}$ summands is an element of $\varphi(G)$, hence equals some q^{th} root of unity. As $p > k + \ell - 4$, the binomial coefficient $\binom{k+\ell-4}{\ell-2}$ is not divisible by p. Thus, it follows from Lemma 3.2 that $Q(c_1, c_2, \ldots, c_{k+\ell-4})$ cannot be zero. Accordingly, $\text{coeff}_P(x^{k-1}y^{\ell-1}) \neq 0$, which completes the proof of Theorem 3.8. □

Chapter 4

The Combinatorial Nullstellensatz

In the present chapter we demonstrate the strength of the Combinatorial Nullstellensatz. A relatively simple application will prove Theorem 2.12. This is done in the first section, whereas the second section is devoted to the proof of the inverse theorem related to the Erdős–Heilbronn problem (Theorem 2.14). Not only the main idea is more striking in this case, but also the technical details are a lot more formidable. The same idea leads to a new proof of Theorem 2.3 whose difficulty is intermediate.

4.1 The Exceptional Case of the Erdős–Heilbronn Conjecture

Here we prove Theorem 2.12. Denote the field of characteristic p at issue by F. If $|A|+|B|-2 > p$, then there exist nonempty subsets $A' \subseteq A$ and $B' \subseteq B$ such that $|A| + |B| - 2 = p$ and $A' \neq B'$. Since $A' \dot{+} B' \subseteq A \dot{+} B$, it is enough to prove the theorem for the pair A', B'. Thus we may assume that $p \geq |A| + |B| - 2$. The statement is obvious if $p = 2$, so we also assume that p is an odd prime, or $p = \infty$.

If A and B are *arbitrary* nonempty subsets of F with $p \geq |A| + |B| - 2$, then $|A \dot{+} B| \geq |A| + |B| - 3$. Indeed, if $|A| \neq |B|$, then in fact $|A \dot{+} B| \geq |A| + |B| - 2$ as it was proven by Alon, Nathanson, and Ruzsa in [7], see Theorem 1 therein. Although it is formally stated only for prime fields, the proof works in arbitrary fields, as they mention it at the end of the paper. If $|A| = |B| \geq 2$, then this applied for the sets A and $B' = B \setminus \{b\}$ for any $b \in B$ gives

$$|A \dot{+} B| \geq |A \dot{+} B'| \geq |A| + |B'| - 2 = |A| + |B| - 3.$$

If one of the sets has only one element, then the statement is obvious. Accordingly, we only have to prove the following 'inverse' version of Theorem 2.12.

Theorem 4.1. *Let A, B be subsets of a field F of characteristic $p > 2$ such that $|A| = |B| = k \geq 2$ and $p \geq 2k-1$. If $|A\dot{+}B| = 2k-3$, then $A = B$.*

Assume that $A = \{a_1, a_2, \ldots, a_k\}$, $B = \{b_1, b_2, \ldots, b_k\}$, and put

$$C = A\dot{+}B = \{c_1, c_2, \ldots, c_{2k-3}\}.$$

The polynomial $f \in F[x, y]$ defined as

$$f(x, y) = (x - y) \prod_{i=1}^{2k-3} (x + y - c_i)$$

has the property that $f(a_i, b_j) = 0$ for any $1 \leq i, j \leq k$.
In order to apply the Combinatorial Nullstellensatz (Theorem 2.26), we introduce the polynomials

$$g(x) = \prod_{i=1}^{k}(x - a_i) = x^k - \alpha_1 x^{k-1} + \alpha_2 x^{k-2} - \ldots + (-1)^k \alpha_k$$

and

$$h(y) = \prod_{i=1}^{k}(y - b_i) = y^k - \beta_1 y^{k-1} + \beta_2 y^{k-2} - \ldots + (-1)^k \beta_k,$$

where $\alpha_i = \sigma_i(A)$ and $\beta_i = \sigma_i(B)$ are the elementary symmetric functions of $a_1, a_2, \ldots, a_k$ resp. $b_1, b_2, \ldots, b_k$. In view of Theorem 2.26, there exist polynomials $q, r \in F[x, y]$ of degree at most $k-2$ such that

$$f(x, y) = q(x, y)g(x) - r(y, x)h(y). \tag{4.1}$$

Writing

$$q(x, y) = \sum_{i=0}^{k-2} q_i(x, y), \quad r(x, y) = \sum_{i=0}^{k-2} r_i(x, y) \quad \text{and} \quad f_i(x, y) = (x - y)(x + y)^{i-1},$$

where q_i, r_i, f_i are homogeneous polynomials of degree i, with the additional notations $\gamma_i = \sigma_i(C)$ $(1 \leq i \leq 2k-3)$ and

$$q_{-1} = q_{-2} = r_{-1} = r_{-2} = 0, \quad \alpha_0 = \beta_0 = \gamma_0 = 1,$$

Eq. (4.1) implies the following equations of homogeneous polynomials of degree $2k-2-t$ for every integer $0 \leq t \leq k$:

$$\begin{aligned}(-1)^t \gamma_t f_{2k-2-t}(x, y) \quad = \quad & \sum_{j=0}^{t} (-1)^{t-j} \big\{ \alpha_{t-j} q_{k-2-j}(x, y) x^{k-t+j} \\ & - \beta_{t-j} r_{k-2-j}(y, x) y^{k-t+j} \big\}. \end{aligned} \tag{4.2}$$

Finally writing

$$q_i(x, y) = \sum_{u+v=i} A_{uv} x^u y^v \qquad \text{and} \qquad r_i(x, y) = \sum_{u+v=i} B_{uv} x^u y^v$$

we find that the equations (4.2) encode certain relations between the coefficients A_{uv}, B_{uv} and the numbers $\alpha_i, \beta_i, \gamma_i$. The careful study of these relations, after a technical elimination process that we postpone until the next section, results in the following

Lemma 4.2. *For every integer* $1 \le t \le k$, $\alpha_t = \beta_t$ *and* $u+v = k-2-t$ *implies* $A_{uv} = B_{uv}$.

Consequently, $g(z) = h(z)$. It means that $a_1, a_2, \ldots, a_k$ and $b_1, b_2, \ldots, b_k$ are the roots of the same polynomial of degree k, hence $A = B$ as claimed. It only remains to prove Lemma 4.2.

Details I: Proof of Lemma 4.2

For $1 \le i \le 2k-3$, let

$$f_i(x,y) = (x-y)(x+y)^{i-1} = \sum_{u+v=i} C_{uv} x^u y^v.$$

Then $C_{i,0} = 1$, $C_{0,i} = -1$, and in case $u, v \neq 0$ we have

$$C_{uv} = -C_{vu} = \binom{i-1}{u-1} - \binom{i-1}{u} = \frac{2u-i}{u}\binom{i-1}{u-1}.$$

Since $i < p$, $C_{uv} = 0$ if and only if i is even and $u = v = i/2$. Consider $C_{uv} + C_{u-1,v+1}$. If $u = i$, then it is

$$C_{i,0} + C_{i-1,1} = 1 + \binom{i-1}{i-2} - \binom{i-1}{i-1} = i-1,$$

a nonzero element in F if $i > 1$. Similarly in the case $u = 1$,

$$C_{1,i-1} + C_{0,i} = 1 - i \neq 0.$$

In general, if $2 \le u \le i-1$, then

$$\begin{aligned} C_{uv} + C_{u-1,v+1} &= \frac{2u-i}{u}\binom{i-1}{u-1} + \frac{2u-2-i}{u-1}\binom{i-1}{u-2} \\ &= \left\{\frac{2u-i}{u}\cdot\frac{i-u+1}{u-1} + \frac{2u-2-i}{u-1}\right\}\binom{i-1}{u-2} \\ &= \frac{i(i-2v-1)}{u(u-1)}\binom{i-1}{u-2}. \end{aligned}$$

Thus we proved:

Claim 4.3. *If* $i > 1$, *then* $C_{uv} + C_{u-1,v+1} = 0$ *if and only if* $i - 2v - 1 = 0$.

We prove Lemma 4.2 by induction on t. Note that if $t > k-2$, then by definition $u+v = k-2-t$ implies $A_{uv} = B_{uv} = 0$. For the initial step, $\alpha_0 = \beta_0 = 1$ by definition. Let $u+v = k-2$. To see that $A_{uv} = B_{uv}$, consider Eq. (4.2) for $t = 0$. It reads as

$$\sum_{u+v=2k-2} C_{uv} x^u y^v = \sum_{u+v=k-2} A_{uv} x^{u+k} y^v - \sum_{u+v=k-2} B_{uv} y^{u+k} x^v.$$

It follows that

$$B_{uv} = -C_{v,u+k} = C_{u+k,v} = A_{uv}. \tag{4.3}$$

For complete induction, let $1 \le t \le k$, and suppose that Lemma 4.2 has been already proved for smaller values of t. We start with the first statement. First we verify $\alpha_t = \beta_t$ in the case when t is even, that is, $t = 2s$ for some $s \ge 1$. We have $k-1-s \ge k-1-(t-1) \ge 0$. Consider the coefficient of the term $x^{k-1-s}y^{k-1-s}$ in Eq. (4.2). On the left hand side this coefficient is $(-1)^t\gamma_t C_{k-1-s,k-1-s} = 0$. In the polynomial $q_{k-2-j}(x,y)x^{k-t+j}$, the coefficient of $x^{k-1-s}y^{k-1-s}$ is $A_{s-1-j,k-1-s}$ if $j \le s-1$ and 0 otherwise, whereas in $r_{k-2-j}(y,x)y^{k-t+j}$, the coefficient of the same term is $B_{s-1-j,k-1-s}$ if $j \le s-1$ and 0 otherwise. Thus Eq. (4.2) implies

$$\sum_{j=0}^{s-1}(-1)^{t-j}\big\{\alpha_{t-j}A_{s-1-j,k-1-s} - \beta_{t-j}B_{s-1-j,k-1-s}\big\} = 0.$$

Since $(s-1-j)+(k-1-s) = k-2-j$ and $s-1 < t$, based on the induction hypothesis we have $A_{s-1-j,k-1-s} = B_{s-1-j,k-1-s}$ and $\alpha_{t-j} = \beta_{t-j}$ for every $1 \le j \le s-1$. The summation can thus be reduced to the first term and we obtain

$$\alpha_t A_{s-1,k-1-s} - \beta_t B_{s-1,k-1-s} = 0.$$

Here $(s-1)+(k-1-s) = k-2$, and in view of Eq. (4.3)

$$A_{s-1,k-1-s} = B_{s-1,k-1-s} = C_{s-1+k,k-1-s} \ne 0,$$

since $s-1+k \ne k-1-s$, given that $s \ge 1$. It follows that $\alpha_t = \beta_t$.

If t is odd, that is, $t = 2s+1$ with some $s \ge 0$, then in Eq. (4.2) we consider the sum of the coefficients of the terms $x^{k-1-s}y^{k-2-s}$ and $x^{k-2-s}y^{k-1-s}$. (Note that $k-2-s \ge k-2-(t-2) \ge 0$, unless $k = t = 1$, which is excluded by $k \ge 2$.) On the left hand side it is

$$(-1)^t\gamma_t(C_{k-1-s,k-2-s} + C_{k-2-s,k-1-s}) = 0.$$

Therefore Eq. (4.2) implies

$$\begin{aligned} 0 \;=\; & \sum_{j=0}^{s}(-1)^{t-j}\alpha_{t-j}A_{s-j,k-2-s} + \sum_{j=0}^{s-1}(-1)^{t-j}\alpha_{t-j}A_{s-1-j,k-1-s} \\ & - \sum_{j=0}^{s}(-1)^{t-j}\beta_{t-j}B_{s-j,k-2-s} - \sum_{j=0}^{s-1}(-1)^{t-j}\beta_{t-j}B_{s-1-j,k-1-s}. \end{aligned}$$

Since $(s-j)+(k-2-s) = (s-1-j)+(k-1-s) = k-2-j$ and $s < t$, the induction hypothesis once again allows us to reduce the above equation to

$$\begin{aligned} 0 \;=\; & (-1)^t\alpha_t A_{s,k-2-s} + (-1)^t\alpha_t A_{s-1,k-1-s} \\ & -(-1)^t\beta_t B_{s,k-2-s} - (-1)^t\beta_t B_{s-1,k-1-s}. \end{aligned}$$

In view of Eq. (4.3) this equation can be rewritten as

$$(\alpha_t - \beta_t)(C_{s+k,k-2-s} + C_{s-1+k,k-1-s}) = 0.$$

Since $(2k-2) - 2(k-2-s) - 1 = 2s + 1 = t$ is not zero in F, in view of Claim 4.3 it follows that the second term is not zero, and we conclude that $\alpha_t - \beta_t = 0$, $\alpha_t = \beta_t$.

It remains to verify the second statement of the lemma under the additional assumption that the first statement has been already verified. Accordingly, we assume $t \leq k-2$, $\alpha_t = \beta_t$, and let $u + v = k - 2 - t$. On the left hand side of Eq. (4.2), the coefficient of $x^{u+k}y^v$ is $(-1)^t\gamma_t C_{u+k,v}$. If $0 \leq j \leq t$, then $v \leq k-2-t < k-t+j$, thus in $r_{k-2-j}(y,x)y^{k-t+j}$ the coefficient of $x^{u+k}y^v$ is 0. Therefore on the right hand side of Eq. (4.2), the coefficient of $x^{u+k}y^v$ is

$$\sum_{j=0}^{t}(-1)^{t-j}\alpha_{t-j}A_{t-j+u,v}.$$

Consequently, Eq. (4.2) implies

$$\sum_{j=0}^{t}(-1)^{t-j}\alpha_{t-j}A_{t-j+u,v} = (-1)^t\gamma_t C_{u+k,v}.$$

Looking at the coefficient of $x^v y^{u+k}$ the same way we obtain

$$-\sum_{j=0}^{t}(-1)^{t-j}\beta_{t-j}B_{t-j+u,v} = (-1)^t\gamma_t C_{v,u+k}.$$

Since $C_{v,u+k} = -C_{u+k,v}$, it follows that

$$\sum_{j=0}^{t}(-1)^{t-j}\alpha_{t-j}A_{t-j+u,v} = \sum_{j=0}^{t}(-1)^{t-j}\beta_{t-j}B_{t-j+u,v}.$$

Because $(t-j+u)+v = k-2-j$, the induction hypothesis implies $A_{t-j+u,v} = B_{t-u+j,v}$ for $0 \leq j < t$. We have furthermore assumed $\alpha_{t-j} = \beta_{t-j}$ for all $0 \leq j \leq t$, therefore the above equality can be reduced to

$$(-1)^{t-t}\alpha_{t-t}A_{t-t+u,v} = (-1)^{t-t}\beta_{t-t}B_{t-t+u,v}.$$

Since $\alpha_0 = \beta_0 = 1$, we obtain $A_{uv} = B_{uv}$.

4.2 Inverse Theorems

Now we are ready for more serious applications of the Combinatorial Nullstellensatz. First we describe the main idea behind the proofs of Theorem 2.3 and Theorem 2.14. The complicated technical details are worked in the subsequent subsections.

The Main Idea

We start with the more interesting Theorem 2.14. The 'if' part of the theorem being obvious, we only focus on the proof of the reverse implication. The group $\mathbb{Z}/p\mathbb{Z}$ can be embedded into the additive group of any field $\mathbb{F}$ of characteristic p. In particular, if $\mathbb{F}$ is the algebraic closure of the Galois field of order p, then every element of $\mathbb{F}$ has a square root in $\mathbb{F}$. Therefore Theorem 2.14 follows directly from the more general

Theorem 4.4. *Given any integer $k \geq 5$, let $p > 2k-3$ be a prime number and let $\mathbb{F}$ be any field of characteristic p in which every element has a square root. Then every k-element subset A of $\mathbb{F}$ satisfying $|A\dot{+}A| = 2k-3$ is an arithmetic progression in $\mathbb{F}$.*

Proof. Let us remark in advance that throughout most part of the proof we can work without the assumption that every element of $\mathbb{F}$ has a square root in $\mathbb{F}$; this condition is only needed in the proof of Lemma 4.8.

We assume that

$$C = A\dot{+}A = \{c_1, c_2, \ldots, c_{2k-3}\},$$

and the elements of A are $a_1, a_2, \ldots, a_k$. We define the polynomial

$$\dot{f}(x,y) = (x-y)\prod_{c\in C}(x+y-c)$$

and also an auxiliary polynomial

$$g(z) = \prod_{i=1}^{k}(z-a_i).$$

Notice that $\dot{f}(x,y) = 0$ for arbitrary $x, y \in A$. Thus once again we may apply the Combinatorial Nullstellensatz (Theorem 2.26). Accordingly, there exist polynomials $\dot{h}', \dot{h}'' \in \mathbb{F}[x,y]$ of degree at most $k-2$ such that

$$\dot{f}(x,y) = \dot{h}'(x,y)g(x) + \dot{h}''(x,y)g(y).$$

Since the polynomial $\dot{f}$ alternates we can write

$$\dot{f}(x,y) = -\dot{f}(y,x) = -\dot{h}'(y,x)g(y) - \dot{h}''(y,x)g(x)$$

to obtain that

$$\dot{f}(x,y) = \dot{h}(x,y)g(x) - \dot{h}(y,x)g(y), \tag{4.4}$$

where $\dot{h}(x,y) = (1/2)(\dot{h}'(x,y) - \dot{h}''(y,x))$ is a polynomial of degree at most $k-2$. Thus we can write

$$\dot{h}(x,y) = \sum_{i=0}^{k-2}\dot{h}_i(x,y),$$

where

$$\dot{h}_i(x,y) = \sum_{j=0}^{i} \dot{A}_{ij} x^j y^{i-j}.$$

We can also rewrite $\dot{f}(x,y)$ in the form

$$\dot{f}(x,y) = \sum_{i=0}^{2k-3} (-1)^i \dot{\tau}_i \dot{p}_{2k-2-i}(x,y).$$

Here $\dot{\tau}_0 = 1$ and, for $1 \leq i \leq 2k-3$, $\dot{\tau}_i$ is the i^{th} elementary symmetric polynomial of $c_1, c_2, \ldots, c_{2k-3}$, while

$$\dot{p}_i(x,y) = (x-y)(x+y)^{i-1} = \sum_{j=0}^{i} \dot{B}_{ij} x^j y^{i-j},$$

where $\dot{B}_{ii} = 1$, $\dot{B}_{i,0} = -1$, and otherwise

$$\dot{B}_{ij} = \binom{i-1}{j-1} - \binom{i-1}{j} = \frac{2j-i}{j}\binom{i-1}{j-1} = \frac{2j-i}{j}\binom{i-1}{i-j}.$$

If we also denote, for $0 \leq i \leq k$, by $\sigma_i = \sigma_i(A)$ the i^{th} elementary symmetric polynomial in $a_1, a_2, \ldots, a_k$, after comparing coefficients we arrive at certain relations between the numbers $\dot{\tau}_i$, the numbers σ_i and the coefficients $\dot{A}_{ij}$. To have an idea of what is going on, we refer to [53] where all the calculations are carried out in the special case $k = 5$.

After a lengthy argument we obtain the following lemma whose proof we postpone until the very end of this chapter.

Lemma 4.5. *Given any integer $k \geq 5$, let $p > 2k-3$ be a prime number and let $\mathbb{F}$ be any field of characteristic p. There exist polynomials $\dot{q}_3, \dot{q}_4, \ldots, \dot{q}_k \in \mathbb{F}[x,y]$ whose coefficients only depend on k and p with the following property. For every integer $3 \leq i \leq k$, $\dot{q}_i(x, y^2)$ is a homogeneous polynomial of degree i in $\mathbb{F}[x,y]$ such that, if A is any set of k distinct elements of $\mathbb{F}$ satisfying $|A \dot{+} A| = 2k-3$, then*

$$\sigma_i(A) = \dot{q}_i(\sigma_1(A), \sigma_2(A)).$$

In view of this lemma we can conclude that the values of σ_1 and σ_2 uniquely determine those of $\sigma_3, \sigma_4, \ldots, \sigma_k$, and in turn also the elements of A, since they are the k solutions of the equation

$$g(z) = z^k - \sigma_1 z^{k-1} + \sigma_2 z^{k-2} - \ldots + (-1)^k \sigma_k = 0.$$

This means that each k-element subset A of $\mathbb{F}$ for which $|A \dot{+} A| = 2k-3$ is uniquely determined by some pair

$$(\sigma_1, \sigma_2) \in \mathbb{F} \times \mathbb{F}.$$

This is true in particular if A is a (non-constant) arithmetic progression of length k.

Lemma 4.6. *Let $\mathcal{A} = (a_1, a_2, \ldots, a_k)$ be any arithmetic progression in a field $\mathbb{F}$ of characteristic $p > 2k - 3 \geq 7$. Keeping the notation of Lemma 4.5, we have*

$$\sigma_i(\mathcal{A}) = \dot{q}_i(\sigma_1(\mathcal{A}), \sigma_2(\mathcal{A}))$$

for every $i = 3, 4, \ldots, k$.

Proof. Note that if the arithmetic progression $\mathcal{A}$ is not constant, then $|A\dot{+}A| = 2k - 3$ and the assertion follows from Lemma 4.5. Fix the values of k and p. For any $a, d \in \mathbb{F}$, let $\mathcal{A}(a, d)$ denote the arithmetic progression

$$a_1 = a,\ a_i = a + (i-1)d \quad (i = 2, 3, \ldots, k).$$

For any arithmetic progression $\mathcal{A}$ in $\mathbb{F}$ there is a unique pair $(a, d) \in \mathbb{F} \times \mathbb{F}$ such that $\mathcal{A} = \mathcal{A}(a, d)$. Note that, for $1 \leq i \leq k$, there exist homogeneous polynomials $r_i \in \mathbb{F}[x, y]$ of degree i such that

$$\sigma_i(\mathcal{A}(a, d)) = r_i(a, d).$$

Introducing the polynomial

$$\tilde{r}_i(x, y) = \dot{q}_i(r_1(x, y), r_2(x, y))$$

for $i = 3, 4, \ldots, k$, we find that $\tilde{r}_i \in \mathbb{F}[x, y]$ is again a homogeneous polynomial of degree i. Moreover, it follows from Lemma 4.5 that

$$r_i(a, d) = \sigma_i(\mathcal{A}(a, d)) = \dot{q}_i(\sigma_1(\mathcal{A}(a, d)), \sigma_2(\mathcal{A}(a, d))) = \tilde{r}_i(a, d)$$

holds for every $(a, d) \in \mathbb{F} \times (\mathbb{F} \setminus \{0\})$. Recall the following simple lemma (see e.g. [2]).

Lemma 4.7. *If $f = f(x_1, x_2, \ldots, x_k)$ is a polynomial over a field F, whose degree as a polynomial in x_i is at most t_i for $1 \leq i \leq k$, and $f(s_1, s_2, \ldots, s_k) = 0$ for all $s_1 \in S_1, s_2 \in S_2, \ldots, s_k \in S_k$ where, for $1 \leq i \leq k$, $S_i \subseteq F$ such that $|S_i| > t_i$, then f is the zero polynomial.*

Noting that $|\mathbb{F}| - 1 \geq p - 1 > k \geq i$, we can conclude that $r_i = \tilde{r}_i$. Consequently,

$$\sigma_i(\mathcal{A}(a, d)) = \dot{q}_i(\sigma_1(\mathcal{A}(a, d)), \sigma_2(\mathcal{A}(a, d)))$$

holds for every $a, d \in \mathbb{F}$, and the assertion is proved. □

On the other hand, every pair $(\sigma_1, \sigma_2) \in \mathbb{F} \times \mathbb{F}$ determines a unique arithmetic progression:

Lemma 4.8. *Let $k \geq 3$ be any integer and let $\mathbb{F}$ be a field of characteristic $p > k + 1$ in which every element has a square root. For every pair $(\sigma_1, \sigma_2) \in \mathbb{F} \times \mathbb{F}$ there is an arithmetic progression $\mathcal{A} = (a_1, a_2, \ldots, a_k)$ such that $\sigma_1(\mathcal{A}) = \sigma_1$ and $\sigma_2(\mathcal{A}) = \sigma_2$. Moreover, this progression is unique up to the reversal of the order of its elements.*

Proof. Let m be the unique element of $\mathbb{F}$ satisfying $km = \sigma_1$, that is, $m = \sigma_1/k$. If $k = 2\ell + 1$ is odd, then the arithmetic progression $\mathcal{A} = (a_1, a_2, \ldots, a_k)$ satisfies $\sigma_1(\mathcal{A}) = \sigma_1$ if and only if

$$a_1 = m - \ell d, a_2 = m - (\ell - 1)d, \ldots, a_{\ell+1} = m, \ldots, a_k = m + \ell d$$

for some element $d \in \mathbb{F}$. As

$$2\sigma_2(\mathcal{A}) = \sigma_1(\mathcal{A})^2 - \sum_{i=1}^{k} a_i^2 = \sigma_1^2 - km^2 - 2d^2 \sum_{i=1}^{\ell} i^2,$$

$\sigma_2(\mathcal{A}) = \sigma_2$ holds if and only if

$$2\frac{k\ell(\ell+1)}{6}d^2 = \sigma_1^2 - km^2 - 2\sigma_2.$$

Note that $\mathrm{char}(\mathbb{F}) > k + 1 > 3$ guarantees that division by the numbers $2, 3, \ell$, $\ell + 1, k - 1, k$ and $k + 1$ is possible in $\mathbb{F}$. Similarly, if $k = 2\ell$ is even, then the arithmetic progression $\mathcal{A} = (a_1, a_2, \ldots, a_k)$ satisfies $\sigma_1(\mathcal{A}) = \sigma_1$ if and only if

$$a_1 = m - (2\ell - 1)(d/2), a_2 = m - (2\ell - 3)(d/2), \ldots, a_k = m + (2\ell - 1)(d/2)$$

for some element $d \in \mathbb{F}$. As in the previous case, $\sigma_2(\mathcal{A}) = \sigma_2$ holds if and only if

$$km^2 + 2(d/2)^2(1^2 + 3^2 + \ldots + (2\ell - 1)^2) = \sigma_1^2 - 2\sigma_2.$$

In each case, the arithmetic progression $\mathcal{A}$ satisfies the conditions if and only if

$$d^2 = \frac{12}{k^2(k-1)(k+1)}\Big((k-1)\sigma_1^2 - 2k\sigma_2\Big).$$

Since by our assumption on $\mathbb{F}$, every element of $\mathbb{F}$ has a square root, there is indeed an arithmetic progression $\mathcal{A}$ that satisfies the two requirements. The uniqueness of $\mathcal{A}$ follows from the fact that square roots in $\mathbb{F}$ are unique up to a multiplicative factor ± 1. □

Now it is straightforward to complete the proof of Theorem 4.4. Given the k-element subset A of $\mathbb{F}$ with $|A \dot{+} A| = 2k - 3$, Lemma 4.8 guarantees the existence of a k-term arithmetic progression $\mathcal{A}$ such that $\sigma_1(\mathcal{A}) = \sigma_1(A)$ and $\sigma_2(\mathcal{A}) = \sigma_2(A)$. It follows from Lemmas 4.5 and 4.6 that $\sigma_i(\mathcal{A}) = \sigma_i(A)$ is valid for every $1 \le i \le k$. Consequently, there is a bijection between the elements of A and the terms of $\mathcal{A}$, that is, the elements of A indeed form an arithmetic progression. □

Turning to the proof of Theorem 2.3, note that if $k = 1$ or $k = 2$, then A is a priori an arithmetic progression. Similarly to the previous case, Theorem 2.3 is an immediate consequence of

Theorem 4.9. *Given any integer $k \ge 3$, let $p > 2k - 1$ be a prime number and let $\mathbb{F}$ be any field of characteristic p in which every element has a square root. Then every k-element subset A of $\mathbb{F}$ satisfying $|A + A| = 2k - 1$ is an arithmetic progression in $\mathbb{F}$.*

Proof. Keeping the notations from the previous proof, the key lemma in this case is

Lemma 4.10. *Given any integer $k \geq 3$, let $p > 2k-1$ be a prime number and let $\mathbb{F}$ be any field of characteristic p. There exist polynomials $q_3, q_4, \ldots, q_k \in \mathbb{F}[x,y]$ whose coefficients only depend on k and p with the following property. For every integer $3 \leq i \leq k$, $q_i(x, y^2)$ is a homogeneous polynomial of degree i in $\mathbb{F}[x,y]$ such that, if A is any set of k distinct elements of $\mathbb{F}$ satisfying $|A+A| = 2k-1$, then*

$$\sigma_i(A) = q_i(\sigma_1(A), \sigma_2(A)).$$

We will prove this lemma in the following subsection. Based on this lemma one only has to mimic the proof of Lemma 4.6 to obtain

Lemma 4.11. *Let $\mathcal{A} = (a_1, a_2, \ldots, a_k)$ be any arithmetic progression in a field $\mathbb{F}$ of characteristic $p > 2k-1 \geq 5$. Keeping the notation of Lemma 4.10, we have*

$$\sigma_i(\mathcal{A}) = q_i(\sigma_1(\mathcal{A}), \sigma_2(\mathcal{A}))$$

for every $i = 3, 4, \ldots, k$.

Now given the k-element subset A of $\mathbb{F}$ with $|A+A| = 2k-1$, replacing Lemmas 4.5 and 4.6 by Lemmas 4.10 and 4.11, respectively, the proof of Theorem 4.9 can be completed along the same lines as that of Theorem 4.4. It only remains to prove Lemma 4.10. □

Details II: Proof of Lemma 4.10

The proof of this lemma is very similar to that of Lemma 4.5, but technically it is considerably more simple. Therefore we begin with the proof of this lemma and postpone the proof of the more interesting Lemma 4.5 to the next subsection.

Again, let the elements of A be $a_1, a_2, \ldots, a_k$ and assume that

$$D = A + A = \{d_1, d_2, \ldots, d_{2k-1}\}.$$

Introduce the polynomial

$$f(x,y) = \prod_{d \in D} (x + y - d).$$

This time we find that $f(x,y) = 0$ for arbitrary $x, y \in A$. It follows from the Combinatorial Nullstellensatz (Lemma 2.26) that there exist polynomials $h', h'' \in \mathbb{F}[x,y]$ of degree at most $k-1$ such that

$$f(x,y) = h'(x,y)g(x) + h''(x,y)g(y),$$

where

$$g(z) = \prod_{i=1}^{k}(z - a_i)$$

is the same auxiliary polynomial as in the previous proof.

Since the polynomial f is symmetric we can write

$$f(x, y) = f(y, x) = h'(y, x)g(y) + h''(y, x)g(x)$$

to obtain that

$$f(x, y) = h(x, y)g(x) + h(y, x)g(y), \tag{4.5}$$

where $h(x, y) = (1/2)(h'(x, y) + h''(y, x))$ is a polynomial of degree at most $k - 1$. Thus we can write

$$h(x, y) = \sum_{i=0}^{k-1} h_i(x, y),$$

where

$$h_i(x, y) = \sum_{j=0}^{i} A_{ij}x^j y^{i-j}.$$

We can also rewrite $f(x, y)$ in the form

$$f(x, y) = \sum_{i=0}^{2k-1} (-1)^i \tau_i p_{2k-1-i}(x, y).$$

Here $\tau_0 = 1$ and, for $1 \leq i \leq 2k - 1$, τ_i is the i^{th} elementary symmetric polynomial of $d_1, d_2, \ldots, d_{2k-1}$, while

$$p_i(x, y) = (x + y)^i = \sum_{j=0}^{i} B_{ij}x^j y^{i-j},$$

where this time $B_{ij} = \binom{i}{j}$. Now the coefficients $A_{k-1,i}$ for $0 \leq i \leq k - 1$ can be easily determined if one compares in Equation 4.5 the terms of degree $2k - 1$. With our notations, this equation implies

$$p_{2k-1}(x, y) = h_{k-1}(x, y)x^k + h_{k-1}(y, x)y^k,$$

from which we conclude that

$$A_{k-1,i} = B_{2k-1,i+k} = \binom{2k-1}{k+i},$$

which is a nonzero element of $\mathbb{F}$ for $\mathrm{char}(\mathbb{F}) = p > 2k - 1$.

Now we are ready to prove the following extension of Lemma 4.10.

Lemma 4.12. *There exist polynomials q_t ($0 \leq t \leq k$) and q_{ti} ($0 \leq t \leq k - 1$, $0 \leq i \leq k - 1 - t$) in $\mathbb{F}[x, y]$ whose coefficients only depend on k and p with the following property. The polynomials $q_t(x, y^2)$ and $q_{ti}(x, y^2)$ are homogeneous polynomials of degree t such that*

$$\sigma_t(A) = q_t(\sigma_1(A), \sigma_2(A))$$

and

$$A_{k-1-t,i} = q_{ti}(\sigma_1(A), \sigma_2(A)).$$

Proof. We prove this lemma by induction on t. The statement is clearly valid with $q_0 = 1$ and $q_{0,i} = \binom{2k-1}{k+i}$. Thus we may assume that $1 \le t \le k$, and the polynomials q_s, q_{si} have been already found for $0 \le s \le t-1$ and for all appropriate values of i. To prove the statement for t we will compare in Equation 4.5 the terms of degree $2k-1-t$. That is, we consider the following consequence of Equation 4.5:

$$\begin{aligned}(-1)^t &\tau_t p_{2k-1-t}(x,y) \\ &= \sum_{j=0}^{t} (-1)^{t-j}\sigma_{t-j}\Big(h_{k-1-j}(x,y)x^{k-t+j} + h_{k-1-j}(y,x)y^{k-t+j}\Big), \end{aligned} \tag{4.6}$$

where we use the convenient notation $h_{-1}(x,y) = 0$, and as before, we write $\sigma_i = \sigma_i(A)$. First we determine the polynomial q_t. If $t = 1$ or $t = 2$, then $q_1(x,y) = x$, resp. $q_2(x,y) = y$ will obviously have the desired properties. Next, if $t = 2s+1$ where $s \ge 1$, we compare the coefficients of $x^{k-s-1}y^{k-s-1}$ in the above equation, and also that of $x^{k-s}y^{k-s-2}$, to obtain the relations

$$\tau_t B_{2k-1-t,k-s-1} = 2\sum_{j=0}^{s} (-1)^j \sigma_{t-j} A_{k-1-j,s-j} \tag{4.7}$$

and

$$\tau_t B_{2k-1-t,k-s} = \sum_{j=0}^{s+1} (-1)^j \sigma_{t-j} A_{k-1-j,s+1-j} + \sum_{j=0}^{s-1} (-1)^j \sigma_{t-j} A_{k-1-j,s-1-j}.$$

Eliminating τ_t from these equations we find that

$$\begin{aligned} &2\binom{2k-2s-2}{k-s} \sum_{j=0}^{s} (-1)^j \sigma_{t-j} A_{k-1-j,s-j} = \\ &= \binom{2k-2s-2}{k-s-1} \left\{ \sum_{j=0}^{s+1} (-1)^j \sigma_{t-j} A_{k-1-j,s+1-j} + \sum_{j=0}^{s-1} (-1)^j \sigma_{t-j} A_{k-1-j,s-1-j} \right\}. \end{aligned}$$

It follows that

$$\begin{aligned} &\binom{2k-2s-2}{k-s-1} \left\{ \sum_{j=1}^{s+1} (-1)^j \sigma_{t-j} A_{k-1-j,s+1-j} + \sum_{j=1}^{s-1} (-1)^j \sigma_{t-j} A_{k-1-j,s-1-j} \right\} - \\ &-2\binom{2k-2s-2}{k-s} \sum_{j=1}^{s} (-1)^j \sigma_{t-j} A_{k-1-j,s-j} = \gamma_t \sigma_t, \end{aligned}$$

where

$$\gamma_t = 2\binom{2k-2s-2}{k-s}\binom{2k-1}{k+s} - \binom{2k-2s-2}{k-s-1}\left\{ \binom{2k-1}{k+s+1} + \binom{2k-1}{k+s-1} \right\}.$$

To see that γ_t is a nonzero element of $\mathbb{F}$, we express it as

$$\gamma_t = \binom{2k-2s-2}{k-s-1}\binom{2k-1}{k+s-1}\delta_t,$$

where the binomial coefficients $\binom{2k-2s-2}{k-s-1}$ and $\binom{2k-1}{k+s-1}$ are nonzero elements of $\mathbb{F}$ due to the assumption $p > 2k-1$, as well as

$$\begin{aligned}
\delta_t &= 2 \cdot \frac{k-s-1}{k-s} \cdot \frac{k-s}{k+s} - \left\{ \frac{(k-s)(k-s-1)}{(k+s+1)(k+s)} + 1 \right\} \\
&= \frac{2(k-s-1)(k+s+1) - (k-s)(k-s-1) - (k+s+1)(k+s)}{(k+s+1)(k+s)} \\
&= -\frac{2(s+1)^2 + 2s(s+1)}{(k+s+1)(k+s)} \\
&= -\frac{2(s+1)t}{(k+s+1)(k+s)}.
\end{aligned}$$

Since $s+1 < t$, it follows from the induction hypothesis that

$$A_{k-1-j,s-j} = q_{j,s-j}(\sigma_1, \sigma_2) \quad \text{for} \quad 1 \le j \le s,$$

$$A_{k-1-j,s+1-j} = q_{j,s+1-j}(\sigma_1, \sigma_2) \quad \text{for} \quad 1 \le j \le s+1,$$

$$A_{k-1-j,s-1-j} = q_{j,s-1-j}(\sigma_1, \sigma_2) \quad \text{for} \quad 1 \le j \le s-1,$$

whereas

$$\sigma_{t-j} = q_{t-j}(\sigma_1, \sigma_2) \quad \text{for} \quad 1 \le j \le s+1,$$

and that $(q_{t-j}q_{j,s-j})(x, y^2)$, $(q_{t-j}q_{j,s+1-j})(x, y^2)$, $(q_{t-j}q_{j,s-1-j})(x, y^2)$ are homogeneous polynomials of degree t, for all relevant values of j. Therefore the polynomial

$$q_t = \gamma_t^{-1}\left(\sum_{j=1}^{s+1} (-1)^j r_j + \sum_{j=1}^{s-1} (-1)^j r'_j - 2\sum_{j=1}^{s} (-1)^j r''_j \right),$$

where

$$r_j = \binom{2k-2s-2}{k-s-1} q_{t-j}q_{j,s+1-j}, \quad r'_j = \binom{2k-2s-2}{k-s-1} q_{t-j}q_{j,s-1-j}$$

and

$$r''_j = \binom{2k-2s-2}{k-s} q_{t-j}q_{j,s-j}$$

will certainly satisfy all the requirements.

A similar procedure can be taken also if $t = 2s$ for some integer $s \ge 2$. It is done by comparing the coefficients of $x^{k-s}y^{k-s-1}$ and also that of $x^{k-s+1}y^{k-s-2}$ in Equation 4.6. This leads to the relations

$$\tau_t B_{2k-1-t,k-s} = \sum_{j=0}^{s} (-1)^j \sigma_{t-j} A_{k-1-j,s-j} + \sum_{j=0}^{s-1} (-1)^j \sigma_{t-j} A_{k-1-j,s-1-j} \tag{4.8}$$

and

$$\tau_t B_{2k-1-t,k-s+1} = \sum_{j=0}^{s+1}(-1)^j \sigma_{t-j} A_{k-1-j,s+1-j} + \sum_{j=0}^{s-2}(-1)^j \sigma_{t-j} A_{k-1-j,s-2-j}.$$

After eliminating τ_t from these equations and rearranging the terms we find that

$$\gamma_t \sigma_t = \binom{2k-2s-1}{k-s}\left\{\sum_{j=1}^{s+1}(-1)^j \sigma_{t-j} A_{k-1-j,s+1-j} + \sum_{j=1}^{s-2}(-1)^j \sigma_{t-j} A_{k-1-j,s-2-j}\right\}$$

$$-\binom{2k-2s-1}{k-s+1}\left\{\sum_{j=1}^{s}(-1)^j \sigma_{t-j} A_{k-1-j,s-j} + \sum_{j=1}^{s-1}(-1)^j \sigma_{t-j} A_{k-1-j,s-1-j}\right\},$$

where this time

$$\begin{aligned}\gamma_t &= \binom{2k-2s-1}{k-s+1}\left\{\binom{2k-1}{k+s} + \binom{2k-1}{k+s-1}\right\} \\ &\quad -\binom{2k-2s-1}{k-s}\left\{\binom{2k-1}{k+s+1} + \binom{2k-1}{k+s-2}\right\}.\end{aligned}$$

Again we want to prove that γ_t is a nonzero element of $\mathbb{F}$, so we write

$$\gamma_t = \binom{2k-2s-1}{k-s}\binom{2k-1}{k+s-2}\delta_t,$$

where the binomial coefficients $\binom{2k-2s-1}{k-s}$ and $\binom{2k-1}{k+s-2}$ are nonzero elements of $\mathbb{F}$ due to the assumption $p > 2k-1$, and so is

$$\begin{aligned}\delta_t &= \frac{k-s-1}{k-s+1}\left\{\frac{(k-s+1)(k-s)}{(k+s)(k+s-1)} + \frac{k-s+1}{k+s-1}\right\} \\ &\quad -\left\{\frac{(k-s+1)(k-s)(k-s-1)}{(k+s+1)(k+s)(k+s-1)} + 1\right\} \\ &= \frac{(k-s-1)(k-s)(k+s+1) + (k-s-1)(k+s)(k+s+1)}{(k+s+1)(k+s)(k+s-1)} \\ &\quad -\frac{(k-s+1)(k-s)(k-s-1) + (k+s+1)(k+s)(k+s-1)}{(k+s+1)(k+s)(k+s-1)} \\ &= \frac{2k(k^2-(s+1)^2)}{(k+s+1)(k+s)(k+s-1)} \\ &\quad -\frac{2k^3 + 2k(s(s+1)+s(s-1)+(s+1)(s-1))}{(k+s+1)(k+s)(k+s-1)} \\ &= -\frac{2k(2s)(2s+1)}{(k+s+1)(k+s)(k+s-1)} \\ &= -\frac{2kt(t+1)}{(k+s+1)(k+s)(k+s-1)}.\end{aligned}$$

Therefore we may introduce the polynomial

$$q_t = \gamma_t^{-1}\left(\sum_{j=1}^{s+1}(-1)^j r_j + \sum_{j=1}^{s-2}(-1)^j r'_j - \sum_{j=1}^{s}(-1)^j r''_j - \sum_{j=1}^{s-1}(-1)^j r'''_j\right),$$

where, referring only to polynomials q_i, q_{ij} we have already defined,

$$r_j = \binom{2k-2s-1}{k-s} q_{t-j}q_{j,s+1-j}, \quad r'_j = \binom{2k-2s-1}{k-s} q_{t-j}q_{j,s-2-j}$$

and

$$r''_j = \binom{2k-2s-1}{k-s+1} q_{t-j}q_{j,s-j}, \quad r'''_j = \binom{2k-2s-1}{k-s+1} q_{t-j}q_{j,s-1-j}.$$

According to the induction hypothesis, $q_t(x, y^2)$ is a homogeneous polynomial of degree t, and $\sigma_t = q_t(\sigma_1, \sigma_2)$.

Now we are in the position to define the polynomials q_{ti}, assuming also that $t < k$. We start with an intermediate result about the number τ_t.

Lemma 4.13. *There exists a polynomial $q_t^* \in \mathbb{F}[x, y]$ whose coefficients only depend on k and p, such that $q_t^*(x, y^2)$ is a homogeneous polynomial of degree t with the property*

$$\tau_t = q_t^*(\sigma_1, \sigma_2).$$

Proof. If $t = 2s + 1$, $s \geq 0$, then we can use Equation 4.7 to find that the polynomial

$$q_t^* = 2\binom{2k-2s-2}{k-s-1}^{-1} \sum_{j=0}^{s}(-1)^j q_{t-j}q_{j,s-j}$$

will have the desired properties. Similarly, in the case when $t = 2s$, $s \geq 1$, it follows from Equation 4.8 that

$$q_t^* = \binom{2k-2s-1}{k-s}^{-1}\left\{\sum_{j=0}^{s}(-1)^j q_{t-j}q_{j,s-j} + \sum_{j=0}^{s-1}(-1)^j q_{t-j}q_{j,s-1-j}\right\}$$

is an appropriate polynomial. □

Returning to the polynomials q_{ti}, to express the coefficients $A_{k-1-t,k-1-t-i}$ $(0 \leq i \leq k-t-1)$ in the desired form we compare the coefficients of $x^{2k-1-t-i}y^i$ in Equation 4.6. Since $2k-1-t-i \geq k-t+j$ for every $0 \leq j \leq t$, whereas $i < k-t+j$ for every $0 \leq j \leq t$, we obtain that

$$\tau_t B_{2k-1-t,2k-1-t-i} = \sum_{j=0}^{t}(-1)^j \sigma_{t-j} A_{k-1-j,k-1-j-i},$$

which implies that

$$A_{k-1-t,k-1-t-i} = (-1)^t \tau_t \binom{2k-1-t}{2k-1-t-i} - \sum_{j=0}^{t-1} (-1)^{t-j} \sigma_{t-j} A_{k-1-j,k-1-j-i}.$$

Given that $t < k$, our induction hypothesis, the already proved properties of q_t and Lemma 4.13 imply that, for every $0 \le i \le k-t-1$, the polynomial

$$q_{t,k-1-t-i} = (-1)^t \binom{2k-1-t}{2k-1-t-i} q_t^* - \sum_{j=0}^{t-1} (-1)^{t-j} q_{t-j} q_{j,k-1-j-i}$$

is such that $q_{t,k-1-t-i}(x, y^2)$ is homogeneous of degree t and

$$A_{k-1-t,k-1-t-i} = q_{t,k-1-t-i}(\sigma_1, \sigma_2).$$

This completes the proof of the induction step and also that of Lemma 4.10. □

Details III: Proof of Lemma 4.5

We intend to carry the proof of Lemma 4.10 through as far as it is possible. Note first of all, that although $\dot{B}_{ij} = 0$ for $i = 2j$, in the case $i/2 < j \le i \le 2k-2$ we have that

$$\dot{B}_{ij} = \frac{2j-i}{j} \binom{i-1}{i-j}$$

is a nonzero element of $\mathbb{F}$ for $\mathrm{char}(\mathbb{F}) = p > 2k-3$ implies $\mathrm{char}(\mathbb{F}) \ge 2k-1 > \max\{2j-i, j, i-1\}$.

Collecting the terms of degree $2k-2$ in Equation 4.4 results in the polynomial equation

$$\dot{p}_{2k-2}(x, y) = \dot{h}_{k-2}(x, y) x^k - \dot{h}_{k-2}(y, x) y^k.$$

Looking at the coefficient of $x^{k+i} y^{k-i-2}$ on each side we find that

$$\dot{A}_{k-2,i} = \dot{B}_{2k-2,k+i} = \frac{2i+2}{k+i} \binom{2k-3}{k-i-2}$$

is a nonzero element of $\mathbb{F}$ for $i = 0, 1, \ldots, k-2$.

The analogue of Lemma 4.12, which is a direct extension of the lemma we are about to prove is the following

Lemma 4.14. *There exist polynomials $\dot{q}_t$ ($0 \le t \le k$) and $\dot{q}_{ti}$ ($0 \le t \le k-2$, $0 \le i \le k-2-t$) in $\mathbb{F}[x, y]$ whose coefficients only depend on k and p with the following property. The polynomials $\dot{q}_t(x, y^2)$ and $\dot{q}_{ti}(x, y^2)$ are homogeneous polynomials of degree t such that*

$$\sigma_t(A) = \dot{q}_t(\sigma_1(A), \sigma_2(A))$$

and

$$\dot{A}_{k-2-t,i} = \dot{q}_{ti}(\sigma_1(A), \sigma_2(A)).$$

Proof. We prove this lemma by induction on t. The statement is clearly valid with $\dot{q}_0 = 1$ and $\dot{q}_{0,i} = \frac{2i+2}{k+i}\binom{2k-3}{k-i-2}$. Thus we may assume that $1 \le t \le k$, and the polynomials q_s, q_{si} have been already found for $0 \le s \le t-1$ and for all appropriate values of i. To prove the statement for t we will compare in Equation 4.4 the terms of degree $2k-2-t$. That is, we consider the following consequence of Equation 4.4:

$$(-1)^t \dot{\tau}_t \dot{p}_{2k-2-t}(x,y) = \sum_{j=0}^{t} (-1)^{t-j} \sigma_{t-j} \Big(\dot{h}_{k-2-j}(x,y) x^{k-t+j} - \dot{h}_{k-2-j}(y,x) y^{k-t+j} \Big), \qquad (4.9)$$

where we conveniently rely on the notation $\dot{h}_{-1}(x,y) = \dot{h}_{-2}(x,y) = 0$, and also $\sigma_i = \sigma_i(A)$.

Again, the main difficulty is to define the polynomial $\dot{q}_t$, whereas the polynomials $\dot{q}_{ti}$ that we only need for the purpose of induction can be easily constructed afterwards. If $t = 1$ or $t = 2$, then $\dot{q}_1(x,y) = x$, resp. $\dot{q}_2(x,y) = y$ have the desired properties. Next we try to determine $\dot{q}_t$ in the case when $t = 2s+1$, $s \ge 1$. For this end we compare the coefficients of $x^{k-s-1}y^{k-s-2}$ resp. $x^{k-s}y^{k-s-3}$ in Equation 4.9 to obtain the relations

$$\dot{\tau}_t \dot{B}_{2k-2-t,k-s-1} = \sum_{j=0}^{s} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-j} - \sum_{j=0}^{s-1} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-1-j} \qquad (4.10)$$

and

$$\dot{\tau}_t \dot{B}_{2k-2-t,k-s} = \sum_{j=0}^{s+1} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s+1-j} - \sum_{j=0}^{s-2} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-2-j}. \qquad (4.11)$$

After eliminating $\dot{\tau}_t$ from these equations and rearranging the terms we find that

$$\dot{\gamma}_t \sigma_t = \dot{B}_{2k-2-t,k-s-1} \left\{ \sum_{j=1}^{s+1} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s+1-j} - \sum_{j=1}^{s-2} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-2-j} \right\}$$
$$- \dot{B}_{2k-2-t,k-s} \left\{ \sum_{j=1}^{s} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-j} - \sum_{j=1}^{s-1} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-1-j} \right\},$$

where

$$\begin{aligned} \dot{\gamma}_t &= \dot{B}_{2k-2-t,k-s}(\dot{A}_{k-2,s} - \dot{A}_{k-2,s-1}) - \dot{B}_{2k-2-t,k-s-1}(\dot{A}_{k-2,s+1} - \dot{A}_{k-2,s-2}) \\ &= \frac{3}{k-s}\binom{2k-t-3}{k-s-1}\left\{ \frac{t+1}{k+s}\binom{2k-3}{k-s-2} - \frac{t-1}{k+s-1}\binom{2k-3}{k-s-1} \right\} \\ &\quad - \frac{1}{k-s-1}\binom{2k-t-3}{k-s-2}\left\{ \frac{t+3}{k+s+1}\binom{2k-3}{k-s-3} - \frac{t-3}{k+s-2}\binom{2k-3}{k-s} \right\}. \end{aligned}$$

We should mention that in the case $s = 1$ the term $\dot{A}_{k-2,s-2}$ is meaningless and in fact does not occur in the above expression for $\dot{\gamma}_t$. Nevertheless, the final formula is valid even in this

case, since $s = 1$ implies that $\frac{t-3}{k+s-2}\binom{2k-3}{k-s} = 0$. In an attempt to prove that $\dot{\gamma}_t \neq 0$ we express it as

$$\dot{\gamma}_t = \binom{2k-t-3}{k-s-2}\binom{2k-3}{k-s-3}\dot{\delta}_t,$$

where the binomial coefficients $\binom{2k-t-3}{k-s-2}$ and $\binom{2k-3}{k-s-3}$ are nonzero elements of $\mathbb{F}$ due to the assumption $p > 2k-3$, whereas

$$\begin{aligned}
\dot{\delta}_t &= \frac{3}{k-s}\cdot\frac{k-s-2}{k-s-1}\left\{\frac{t+1}{k+s}\cdot\frac{k+s}{k-s-2} - \frac{t-1}{k+s-1}\cdot\frac{(k+s)(k+s-1)}{(k-s-1)(k-s-2)}\right\} \\
&\quad -\frac{1}{k-s-1}\left\{\frac{t+3}{k+s+1} - \frac{t-3}{k+s-2}\cdot\frac{(k+s)(k+s-1)(k+s-2)}{(k-s)(k-s-1)(k-s-2)}\right\} \\
&= \frac{1}{k-s-1}\cdot\frac{1}{(k+s+1)(k-s)(k-s-1)(k-s-2)}\cdot\dot{\epsilon}_t,
\end{aligned}$$

where $(k-s-1)(k+s+1)(k-s)(k-s-1)(k-s-2) \neq 0$ and

$$\begin{aligned}
\dot{\epsilon}_t &= 3(t+1)(k+s+1)(k-s-1)(k-s-2) \\
&\quad -3(t-1)(k+s+1)(k+s)(k-s-2) \\
&\quad -(t+3)(k-s)(k-s-1)(k-s-2) \\
&\quad +(t-3)(k+s+1)(k+s)(k+s-1) \\
&= -4t(s+1)\Big\{3k-(2s^2+4s+3)\Big\}.
\end{aligned}$$

We can conclude that $\dot{\gamma}_t$ is a nonzero element of $\mathbb{F}$ if and only if $3k-(2s^2+4s+3) \neq 0$ in $\mathbb{F}$. This is indeed the case when $s = 1$ for then $3k-(2s^2+4s+3) = 3(k-3) \neq 0$, and also when $s = 2$ and $k = 5$. Unfortunately it is not the case in general, thus we cannot really proceed along the lines of the previous proof. However, if $s \geq 2$ and $k > 5$, then $k-s-4 \geq 0$ and we may compare the coefficients of $x^{k-s+1}y^{k-s-4}$ in Equation 4.9 to obtain a new relation

$$\dot{\tau}_t\dot{B}_{2k-2-t,k-s+1} = \sum_{j=0}^{s+2}(-1)^j\sigma_{t-j}\dot{A}_{k-2-j,s+2-j} - \sum_{j=0}^{s-3}(-1)^j\sigma_{t-j}\dot{A}_{k-2-j,s-3-j}. \tag{4.12}$$

Now we can eliminate $\dot{\tau}_t$ from Equations 4.11 and 4.12 to get

$$\begin{aligned}
\dot{\gamma}'_t\sigma_t &= \dot{B}_{2k-2-t,k-s+1}\left\{\sum_{j=1}^{s+1}(-1)^j\sigma_{t-j}\dot{A}_{k-2-j,s+1-j} - \sum_{j=1}^{s-2}(-1)^j\sigma_{t-j}\dot{A}_{k-2-j,s-2-j}\right\} \\
&\quad -\dot{B}_{2k-2-t,k-s}\left\{\sum_{j=1}^{s+2}(-1)^j\sigma_{t-j}\dot{A}_{k-2-j,s+2-j} - \sum_{j=1}^{s-3}(-1)^j\sigma_{t-j}\dot{A}_{k-2-j,s-3-j}\right\},
\end{aligned}$$

where

$$\dot{\gamma}'_t = \dot{B}_{2k-2-t,k-s}(\dot{A}_{k-2,s+2} - \dot{A}_{k-2,s-3}) - \dot{B}_{2k-2-t,k-s+1}(\dot{A}_{k-2,s+1} - \dot{A}_{k-2,s-2})$$

$$
\begin{aligned}
&= \frac{3}{k-s}\binom{2k-t-3}{k-s-1}\left\{\frac{t+5}{k+s+2}\binom{2k-3}{k-s-4}-\frac{t-5}{k+s-3}\binom{2k-3}{k-s+1}\right\}\\
&\quad -\frac{5}{k-s+1}\binom{2k-t-3}{k-s}\left\{\frac{t+3}{k+s+1}\binom{2k-3}{k-s-3}-\frac{t-3}{k+s-2}\binom{2k-3}{k-s}\right\}.
\end{aligned}
$$

Again, if $s=2$, then the term $\dot{A}_{k-2,s-3}$ is meaningless, but the final formula is nevertheless correct for $t-5=0$ in this case. Therefore we can write

$$\dot{\gamma}'_t = \binom{2k-t-3}{k-s-1}\binom{2k-3}{k-s-4}\dot{\delta}'_t,$$

where the binomial coefficients $\binom{2k-t-3}{k-s-1}$ and $\binom{2k-3}{k-s-4}$ are nonzero elements of $\mathbb{F}$ due to the assumption $p > 2k-3$, whereas

$$
\begin{aligned}
\dot{\delta}'_t &= \frac{3}{k-s}\left\{\frac{t+5}{k+s+2}-\right.\\
&\qquad \left.-\frac{t-5}{k+s-3}\cdot\frac{(k+s+1)(k+s)(k+s-1)(k+s-2)(k+s-3)}{(k-s+1)(k-s)(k-s-1)(k-s-2)(k-s-3)}\right\}\\
&\quad -\frac{5}{k-s+1}\cdot\frac{k-s-3}{k-s}\left\{\frac{t+3}{k+s+1}\cdot\frac{k+s+1}{k-s-3}-\right.\\
&\qquad \left.-\frac{t-3}{k+s-2}\cdot\frac{(k+s+1)(k+s)(k+s-1)(k+s-2)}{(k-s)(k-s-1)(k-s-2)(k-s-3)}\right\}.
\end{aligned}
$$

That is,

$$\dot{\delta}'_t = \frac{1}{(k+s+2)(k-s+1)(k-s)^2(k-s-1)(k-s-2)(k-s-3)}\cdot\dot{\epsilon}'_t$$

where $(k+s+2)(k-s+1)(k-s)^2(k-s-1)(k-s-2)(k-s-3)\neq 0$ and

$$
\begin{aligned}
\dot{\epsilon}'_t &= 3(t+5)(k-s+1)(k-s)(k-s-1)(k-s-2)(k-s-3)\\
&\quad -3(t-5)(k+s+2)(k+s+1)(k+s)(k+s-1)(k+s-2)\\
&\quad -5(t+3)(k+s+2)(k-s)(k-s-1)(k-s-2)(k-s-3)\\
&\quad +5(t-3)(k+s+2)(k+s+1)(k+s)(k+s-1)(k-s-3)\\
&= 8t(s+1)\Big\{15k^3-(10s^2+20s+30)k^2+(25s^2+50s+15)k-\\
&\qquad -(2s^4+8s^3+17s^2+18s)\Big\}.
\end{aligned}
$$

Thus we can conclude that $\dot{\gamma}'_t$ is a nonzero element of $\mathbb{F}$ if and only if the integer

$$15k^3-(10s^2+20s+30)k^2+(25s^2+50s+15)k-(2s^4+8s^3+17s^2+18s)$$

is not divisible by p.

Now we can prove that either $\dot{\gamma}_t$ or $\dot{\gamma}'_t$ is a nonzero element of $\mathbb{F}$. Were it not the case, the prime p would divide the integers $3k-(2s^2+4s+3)$ and

$$15k^3-(10s^2+20s+30)k^2+(25s^2+50s+15)k-(2s^4+8s^3+17s^2+18s).$$

Thus in turn, by the division algorithm p would also divide the integers

$$-15k^2+(25s^2+50s+15)k-(2s^4+8s^3+17s^2+18s),$$

$$(15s^2+30s)k-(2s^4+8s^3+17s^2+18s),$$

and finally also the integer

$$5s(s+2)(2s^2+4s+3)-(2s^4+8s^3+17s^2+18s)=2s(s+2)(2s+1)(2s+3)$$

which is absurd since $2s+3=t+2\leq 2k-3<p$.

Accordingly, if $\dot{\gamma}_t\neq 0$ (this is the case if, for example $s=1$, or $s=2$ and $k=5$) we can define the polynomial $\dot{q}_t$ as

$$\dot{q}_t=\dot{\gamma}_t^{-1}\left(\sum_{j=1}^{s+1}(-1)^j\dot{r}_j-\sum_{j=1}^{s-2}(-1)^j\dot{r}'_j-\sum_{j=1}^{s}(-1)^j\dot{r}''_j+\sum_{j=1}^{s-1}(-1)^j\dot{r}'''_j\right),$$

where

$$\dot{r}_j=\frac{1}{k-s-1}\binom{2k-t-3}{k-s-2}\dot{q}_{t-j}\dot{q}_{j,s+1-j},\ \dot{r}'_j=\frac{1}{k-s-1}\binom{2k-t-3}{k-s-2}\dot{q}_{t-j}\dot{q}_{j,s-2-j}$$

and

$$\dot{r}''_j=\frac{3}{k-s}\binom{2k-t-3}{k-s-1}\dot{q}_{t-j}\dot{q}_{j,s-j},\quad \dot{r}'''_j=\frac{3}{k-s}\binom{2k-t-3}{k-s-1}\dot{q}_{t-j}\dot{q}_{j,s-1-j}.$$

Note that since $s+1<t$ and also $s+1\leq k-2$, all the polynomials $\dot{q}_i,\dot{q}_{ij}$ that occur in the above expressions have been already defined.

On the other hand, if $s\geq 2$, $k>5$ and $\dot{\gamma}'_t\neq 0$, then we can define the polynomial $\dot{q}_t$ as

$$\dot{q}_t=(\dot{\gamma}'_t)^{-1}\left(\sum_{j=1}^{s+1}(-1)^j\dot{r}_j^{(4)}-\sum_{j=1}^{s-2}(-1)^j\dot{r}_j^{(5)}-\sum_{j=1}^{s+2}(-1)^j\dot{r}_j^{(6)}+\sum_{j=1}^{s-3}(-1)^j\dot{r}_j^{(7)}\right),$$

where

$$\dot{r}_j^{(4)}=\frac{5}{k-s+1}\binom{2k-t-3}{k-s}\dot{q}_{t-j}\dot{q}_{j,s+1-j},$$

$$\dot{r}_j^{(5)}=\frac{5}{k-s+1}\binom{2k-t-3}{k-s}\dot{q}_{t-j}\dot{q}_{j,s-2-j},$$

$$\dot{r}_j^{(6)}=\frac{3}{k-s}\binom{2k-t-3}{k-s-1}\dot{q}_{t-j}\dot{q}_{j,s+2-j}$$

and

$$\dot{r}_j^{(7)}=\frac{3}{k-s}\binom{2k-t-3}{k-s-1}\dot{q}_{t-j}\dot{q}_{j,s-3-j}.$$

Again, all the polynomials $\dot{q}_i, \dot{q}_{ij}$ that occur in the above expressions have been already defined, as in this case clearly $s+2 < t$ and also $s+2 \leq k-2$. According to the induction hypothesis, in each case $\dot{q}_t(x, y^2)$ is a homogeneous polynomial of degree t, and $\sigma_t = \dot{q}_t(\sigma_1, \sigma_2)$.

We still have to determine the polynomial $\dot{q}_t$ in the case when $t = 2s$, $s \geq 2$. Comparing in Equation 4.9 the coefficients of $x^{k-s-1}y^{k-s-1}$ would yield the trivial equation $0 = 0$, therefore we rather proceed on with comparing the coefficients of $x^{k-s}y^{k-s-2}$ and $x^{k-s+1}y^{k-s-3}$, respectively. Thus we obtain the relations

$$\dot{\tau}_t \dot{B}_{2k-2-t,k-s} = \sum_{j=0}^{s} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-j} - \sum_{j=0}^{s-2} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-2-j} \tag{4.13}$$

and

$$\dot{\tau}_t \dot{B}_{2k-2-t,k-s+1} = \sum_{j=0}^{s+1} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s+1-j} - \sum_{j=0}^{s-3} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-3-j}. \tag{4.14}$$

After eliminating $\dot{\tau}_t$ from these equations and rearranging the terms we find that

$$\begin{aligned} \dot{\gamma}_t \sigma_t = \dot{B}_{2k-2-t,k-s+1} &\left\{ \sum_{j=1}^{s} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-j} - \sum_{j=1}^{s-2} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-2-j} \right\} \\ -\dot{B}_{2k-2-t,k-s} &\left\{ \sum_{j=1}^{s+1} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s+1-j} - \sum_{j=1}^{s-3} (-1)^j \sigma_{t-j} \dot{A}_{k-2-j,s-3-j} \right\}, \end{aligned}$$

where

$$\begin{aligned} \dot{\gamma}_t &= \dot{B}_{2k-2-t,k-s}(\dot{A}_{k-2,s+1} - \dot{A}_{k-2,s-3}) - \dot{B}_{2k-2-t,k-s+1}(\dot{A}_{k-2,s} - \dot{A}_{k-2,s-2}) \\ &= \frac{2}{k-s}\binom{2k-t-3}{k-s-1}\left\{ \frac{t+4}{k+s+1}\binom{2k-3}{k-s-3} - \frac{t-4}{k+s-3}\binom{2k-3}{k-s+1} \right\} \\ &\quad -\frac{4}{k-s+1}\binom{2k-t-3}{k-s}\left\{ \frac{t+2}{k+s}\binom{2k-3}{k-s-2} - \frac{t-2}{k+s-2}\binom{2k-3}{k-s} \right\}. \end{aligned}$$

Again, the formula is valid even in the case of $s = 2$, because then $t - 4 = 0$. We further express $\dot{\gamma}_t$ as

$$\dot{\gamma}_t = \binom{2k-t-3}{k-s-1}\binom{2k-3}{k-s-3}\dot{\delta}_t,$$

where the binomial coefficients $\binom{2k-t-3}{k-s-1}$ and $\binom{2k-3}{k-s-3}$ are nonzero elements of $\mathbb{F}$ due to the assumption $p > 2k-3$, whereas

$$\begin{aligned} \dot{\delta}_t &= \frac{2}{k-s}\left\{ \frac{t+4}{k+s+1} - \right. \\ &\qquad \left. -\frac{t-4}{k+s-3} \cdot \frac{(k+s)(k+s-1)(k+s-2)(k+s-3)}{(k-s+1)(k-s)(k-s-1)(k-s-2)} \right\} \end{aligned}$$

$$-\frac{4}{k-s+1}\cdot\frac{k-s-2}{k-s}\Big\{\frac{t+2}{k+s}\cdot\frac{k+s}{k-s-2}-$$
$$-\frac{t-2}{k+s-2}\cdot\frac{(k+s)(k+s-1)(k+s-2)}{(k-s)(k-s-1)(k-s-2)}\Big\}.$$

That is,

$$\dot\delta_t=\frac{2}{(k+s+1)(k-s+1)(k-s)^2(k-s-1)(k-s-2)}\cdot\dot\epsilon_t$$

where $(k+s+1)(k-s+1)(k-s)^2(k-s-1)(k-s-2)\neq 0$ and

$$\begin{aligned}\dot\epsilon_t &= (t+4)(k-s+1)(k-s)(k-s-1)(k-s-2)\\ &\quad -(t-4)(k+s+1)(k+s)(k+s-1)(k+s-2)\\ &\quad -2(t+2)(k+s+1)(k-s)(k-s-1)(k-s-2)\\ &\quad +2(t-2)(k+s+1)(k+s)(k+s-1)(k-s-2)\\ &= 4t(t+1)\Big\{3k^2-(2s^2+2s+3)k+(2s^2+2s)\Big\}\\ &= 4t(t+1)(k-1)\Big\{3k-(2s^2+2s)\Big\}.\end{aligned}$$

We can conclude that $\dot\gamma_t$ is a nonzero element of $\mathbb{F}$ if and only if $3k-(2s^2+2s)\neq 0$ in $\mathbb{F}$. This is indeed the case when $s=2$ for then $3k-(2s^2+2s)=3(k-4)\neq 0$, and also when $s=3$ and $k=6$; but not in general. However, if $s\geq 3$ and $k>6$, then $k-s-4\geq 0$ and we may compare the coefficients of $x^{k-s+2}y^{k-s-4}$ in Equation 4.9 to obtain a new relation

$$\dot\tau_t\dot B_{2k-2-t,k-s+2}=\sum_{j=0}^{s+2}(-1)^j\sigma_{t-j}\dot A_{k-2-j,s+2-j}-\sum_{j=0}^{s-4}(-1)^j\sigma_{t-j}\dot A_{k-2-j,s-4-j}. \tag{4.15}$$

Now we can eliminate $\dot\tau_t$ from Equations 4.14 and 4.15 to get that

$$\dot\gamma_t'\sigma_t=\dot B_{2k-2-t,k-s+2}\Big\{\sum_{j=1}^{s+1}(-1)^j\sigma_{t-j}\dot A_{k-2-j,s+1-j}-\sum_{j=1}^{s-3}(-1)^j\sigma_{t-j}\dot A_{k-2-j,s-3-j}\Big\}$$
$$-\dot B_{2k-2-t,k-s+1}\Big\{\sum_{j=1}^{s+2}(-1)^j\sigma_{t-j}\dot A_{k-2-j,s+2-j}-\sum_{j=1}^{s-4}(-1)^j\sigma_{t-j}\dot A_{k-2-j,s-4-j}\Big\},$$

where

$$\begin{aligned}\dot\gamma_t' &= \dot B_{2k-2-t,k-s+1}(\dot A_{k-2,s+2}-\dot A_{k-2,s-4})\\ &\quad -\dot B_{2k-2-t,k-s+2}(\dot A_{k-2,s+1}-\dot A_{k-2,s-3})\\ &= \frac{4}{k-s+1}\binom{2k-t-3}{k-s}\Big\{\frac{t+6}{k+s+2}\binom{2k-3}{k-s-4}-\frac{t-6}{k+s-4}\binom{2k-3}{k-s+2}\Big\}\\ &- \frac{6}{k-s+2}\binom{2k-t-3}{k-s+1}\Big\{\frac{t+4}{k+s+1}\binom{2k-3}{k-s-3}-\frac{t-4}{k+s-3}\binom{2k-3}{k-s+1}\Big\}.\end{aligned}$$

Note that the formula is valid even in the case of $s = 3$, because then $t - 6 = 0$. We further express $\dot{\gamma}'_t$ as

$$\dot{\gamma}'_t = \binom{2k-t-3}{k-s}\binom{2k-3}{k-s-4}\dot{\delta}'_t$$

where the binomial coefficients $\binom{2k-t-3}{k-s}$ and $\binom{2k-3}{k-s-4}$ are nonzero elements of $\mathbb{F}$, whereas

$$\begin{aligned}\dot{\delta}'_t &= \frac{4}{k-s+1}\Bigg\{\frac{t+6}{k+s+2} - \frac{t-6}{k+s-4}\cdot \\ &\qquad \cdot\frac{(k+s+1)(k+s)(k+s-1)(k+s-2)(k+s-3)(k+s-4)}{(k-s+2)(k-s+1)(k-s)(k-s-1)(k-s-2)(k-s-3)}\Bigg\} \\ &\quad -\frac{6}{k-s+2}\cdot\frac{k-s-3}{k-s+1}\Bigg\{\frac{t+4}{k+s+1}\cdot\frac{k+s+1}{k-s-3} - \\ &\qquad -\frac{t-4}{k+s-3}\cdot\frac{(k+s+1)(k+s)(k+s-1)(k+s-2)(k+s-3)}{(k-s+1)(k-s)(k-s-1)(k-s-2)(k-s-3)}\Bigg\}.\end{aligned}$$

That is,

$$\dot{\delta}'_t = \frac{2}{(k+s+2)(k-s+2)(k-s+1)^2(k-s)(k-s-1)(k-s-2)(k-s-3)}\cdot\dot{\epsilon}'_t$$

where $(k+s+2)(k-s+2)(k-s+1)^2(k-s)(k-s-1)(k-s-2)(k-s-3) \neq 0$ and

$$\begin{aligned}\dot{\epsilon}'_t &= 2(t+6)(k-s+2)(k-s+1)(k-s)(k-s-1)(k-s-2)(k-s-3) \\ &\quad -2(t-6)(k+s+2)(k+s+1)(k+s)(k+s-1)(k+s-2)(k+s-3) \\ &\quad -3(t+4)(k+s+2)(k-s+1)(k-s)(k-s-1)(k-s-2)(k-s-3) \\ &\quad +3(t-4)(k+s+2)(k+s+1)(k+s)(k+s-1)(k+s-2)(k-s-3) \\ &= t(k-5s+9)(k+s+2)(k+s+1)(k+s)(k+s-1)(k+s-2) \\ &\quad -t(k+5s+14)(k-s+1)(k-s)(k-s-1)(k-s-2)(k-s-3) \\ &= 4t(t+1)\Big\{15k^4 - (10s^2+10s+30)k^3 + (45s^2+45s-15)k^2 - \\ &\qquad -(6s^4+12s^3+29s^2+23s-30)k + (6s^4+12s^3-6s^2-12s)\Big\} \\ &= 4t(t+1)(k-1)\Big\{15k^3 - (10s^2+10s+15)k^2 + (35s^2+35s-30)k - \\ &\qquad -(6s^4+12s^3-6s^2-12s)\Big\}.\end{aligned}$$

Thus we can conclude that $\dot{\gamma}'_t$ is a nonzero element of $\mathbb{F}$ if and only if the integer

$$15k^3 - (10s^2+10s+15)k^2 + (35s^2+35s-30)k - (6s^4+12s^3-6s^2-12s)$$

is not divisible by p.

Now we can prove that either $\dot{\gamma}_t$ or $\dot{\gamma}'_t$ is a nonzero element of $\mathbb{F}$. Were it not the case, the prime p would divide the integers $M = 3k - (2s^2+2s)$ and

$$15k^3 - (10s^2+10s+15)k^2 + (35s^2+35s-30)k - (6s^4+12s^3-6s^2-12s).$$

Consequently, p would also divide the integers

$$-15k^2+(35s^2+35s-30)k-(6s^4+12s^3-6s^2-12s),$$

$$N=(25s^2+25s-30)k-(6s^4+12s^3-6s^2-12s),$$

and finally also the integer

$$\begin{aligned}3N-(25s^2+25s-30)M &= 2s(s+1)\Big\{(25s^2+25s-30)-9(s-1)(s+2)\Big\}\\ &= 8s(s+1)(2s-1)(2s+3),\end{aligned}$$

which is absurd since $2s+3=t+3\leq 2k-3<p$.

Accordingly, if $\dot\gamma_t\neq 0$ (this is the case if, for example $s=2$, or $s=3$ and $k=6$) we can define the polynomial $\dot q_t$ as

$$\dot q_t=\dot\gamma_t^{-1}\left(\sum_{j=1}^{s}(-1)^j\dot r_j-\sum_{j=1}^{s-2}(-1)^j\dot r'_j-\sum_{j=1}^{s+1}(-1)^j\dot r''_j+\sum_{j=1}^{s-3}(-1)^j\dot r'''_j\right),$$

where

$$\dot r_j=\frac{4}{k-s+1}\binom{2k-t-3}{k-s}\dot q_{t-j}\dot q_{j,s-j},\ \dot r'_j=\frac{4}{k-s+1}\binom{2k-t-3}{k-s}\dot q_{t-j}\dot q_{j,s-2-j}$$

and

$$\dot r''_j=\frac{2}{k-s}\binom{2k-t-3}{k-s-1}\dot q_{t-j}\dot q_{j,s+1-j},\quad \dot r'''_j=\frac{2}{k-s}\binom{2k-t-3}{k-s-1}\dot q_{t-j}\dot q_{j,s-3-j}.$$

Note that since $s+1<t$ and also $s+1\leq k-2$, all the polynomials $\dot q_i,\dot q_{ij}$ that occur in the above expressions have been already defined.

On the other hand, if $s\geq 3$, $k>6$ and $\dot\gamma'_t\neq 0$, then we can define the polynomial $\dot q_t$ as

$$\dot q_t=(\dot\gamma'_t)^{-1}\left(\sum_{j=1}^{s+1}(-1)^j\dot r_j^{(4)}-\sum_{j=1}^{s-3}(-1)^j\dot r_j^{(5)}-\sum_{j=1}^{s+2}(-1)^j\dot r_j^{(6)}+\sum_{j=1}^{s-4}(-1)^j\dot r_j^{(7)}\right),$$

where

$$\begin{aligned}\dot r_j^{(4)}&=\frac{6}{k-s+2}\binom{2k-t-3}{k-s+1}\dot q_{t-j}\dot q_{j,s+1-j},\\ \dot r_j^{(5)}&=\frac{6}{k-s+2}\binom{2k-t-3}{k-s+1}\dot q_{t-j}\dot q_{j,s-3-j},\\ \dot r_j^{(6)}&=\frac{4}{k-s+1}\binom{2k-t-3}{k-s}\dot q_{t-j}\dot q_{j,s+2-j}\end{aligned}$$

and

$$\dot r_j^{(7)}=\frac{4}{k-s+1}\binom{2k-t-3}{k-s}\dot q_{t-j}\dot q_{j,s-4-j}.$$

Again, all the polynomials $\dot q_i,\dot q_{ij}$ that occur in the above expressions have been already defined, as in this case clearly $s+2<t$ and also $s+2\leq k-2$. According to the induction hypothesis, in each case $\dot q_t(x,y^2)$ is a homogeneous polynomial of degree t, and $\sigma_t=\dot q_t(\sigma_1,\sigma_2)$.

Having thus found the polynomial $\dot{q}_t$, we proceed on with the definition of the polynomials q_{ti}, under the additional assumption that $t \leq k-2$. First we need the following analogue of Lemma 4.13.

Lemma 4.15. *There exists a polynomial $\dot{q}_t^* \in \mathbb{F}[x,y]$ whose coefficients only depend on k and p, such that $\dot{q}_t^*(x,y^2)$ is a homogeneous polynomial of degree t with the property*

$$\dot{\tau}_t = \dot{q}_t^*(\sigma_1, \sigma_2).$$

Proof. If $t = 2s+1$, $s \geq 0$, then we can use Equation 4.10 to find that the polynomial

$$\dot{q}_t^* = (k-s-1)\binom{2k-t-3}{k-s-2}^{-1}\left\{\sum_{j=0}^{s}(-1)^j \dot{q}_{t-j}\dot{q}_{j,s-j} - \sum_{j=0}^{s-1}(-1)^j \dot{q}_{t-j}\dot{q}_{j,s-1-j}\right\}$$

will have the desired properties. Similarly, in the case when $t = 2s$, $s \geq 1$, it follows from Equation 4.13 that

$$\dot{q}_t^* = \frac{k-s+1}{4}\binom{2k-t-3}{k-s}^{-1}\left\{\sum_{j=0}^{s}(-1)^j \dot{q}_{t-j}\dot{q}_{j,s-j} - \sum_{j=0}^{s-2}(-1)^j \dot{q}_{t-j}\dot{q}_{j,s-2-j}\right\}$$

is an appropriate polynomial. □

Returning to the polynomials $\dot{q}_{ti}$, to express the coefficients $\dot{A}_{k-2-t,k-2-t-i}$ ($0 \leq i \leq k-t-2$) in the desired form we compare the coefficients of $x^{2k-2-t-i}y^i$ in Equation 4.9. Since $2k-2-t-i \geq k \geq k-t+j$ for every $0 \leq j \leq t$, whereas $i \leq k-2-t < k-t+j$ for every $0 \leq j \leq t$, we obtain that

$$\dot{\tau}_t \dot{B}_{2k-2-t,2k-2-t-i} = \sum_{j=0}^{t}(-1)^j \sigma_{t-j}\dot{A}_{k-2-j,k-2-j-i},$$

which implies that

$$\begin{aligned}\dot{A}_{k-2-t,k-2-t-i} &= (-1)^t \dot{\tau}_t \frac{2k-2-t-2i}{2k-2-t-i}\binom{2k-3-t}{i} \\ &\quad - \sum_{j=0}^{t-1}(-1)^{t-j}\sigma_{t-j}\dot{A}_{k-2-j,k-2-j-i}.\end{aligned}$$

Given that $t \leq k-2$, our induction hypothesis, the already proved properties of $\dot{q}_t$ and Lemma 4.15 imply that, for every $0 \leq i \leq k-t-2$, the polynomial

$$\dot{q}_{t,k-2-t-i} = (-1)^t \frac{2k-2-t-2i}{2k-2-t-i}\binom{2k-3-t}{i}\dot{q}_t^* - \sum_{j=0}^{t-1}(-1)^{t-j}\dot{q}_{t-j}\dot{q}_{j,k-2-j-i}$$

is such that $\dot{q}_{t,k-2-t-i}(x,y^2)$ is homogeneous of degree t and

$$\dot{A}_{k-2-t,k-2-t-i} = \dot{q}_{t,k-2-t-i}(\sigma_1,\sigma_2).$$

This completes the proof of the inductional step and also that of Lemma 4.5. □

Chapter 5

The Method of Group Extensions

In the first two sections of the present chapter we extend the Dias da Silva–Hamidoune theorem and our corresponding inverse theorem to arbitrary abelian groups. The third section concerns general finite groups; we prove the noncommutative analogues of the Cauchy–Davenport theorem and Vosper's inverse theorem.

5.1 The Erdős–Heilbronn Problem in Abelian Groups

First we show that Theorem 2.11 is sharp. Assume that $p(G)$ is finite and $p(G)/2 + 1 < k \leq p(G)$. Let P be a subgroup of G with $|P| = p(G)$ and assume that $P = \langle g \rangle$. If

$$A = \{0, g, 2g, \ldots, (k-1)g\},$$

then clearly $A \dot{+} A = P$, indicating that the bound is tight.

Turning to the proof, we note that, since dealing with a finite problem, we may assume that G is finitely generated. We have already seen that the result is valid if $G \cong \mathbb{Z}$ (Statement 1.5), and also when G is a cyclic group of prime power order (Section 3.2). In view of the structure theorem of finitely generated abelian groups, it only remains to prove that if the statement of Theorem 2.11 is true for two abelian groups G^1 and G^2, then it is also valid for their direct sum $G^1 \oplus G^2$. Accordingly, suppose that we have already proved Theorem 2.11 for the abelian groups G^1 and G^2. Let

$$G = G^1 \oplus G^2 = \{(g, h) \mid g \in G^1, h \in G^2\},$$

where addition in G is defined by

$$(g, h) + (g', h') = (g + g', h + h').$$

Note that $p(G^i) \geq p(G)$ for $i = 1, 2$. For a set $X \subseteq G$ write

$$X^1 = \{g \in G^1 \mid \text{there exists } h \in G^2 \text{ with } (g, h) \in X\}.$$

We define X^2 in a similar way. An immediate consequence of this definition is the following statement.

Proposition 5.1. *For arbitrary $X, Y \subseteq G$ we have $(X \setminus Y)^1 \supseteq X^1 \setminus Y^1$ and $X^1 \dot{+} X^1 \subseteq (X \dot{+} X)^1 \subseteq X^1 + X^1$.*

We have to prove that $|A \dot{+} A| \geq \min\{p(G), 2k - 3\}$ holds for every $A \subseteq G$ with $|A| = k$. This is easy to check if $p(G) = 2$, and we may assume that $2k - 3 \leq p(G)$ otherwise. Then

$$2|A^i| - 3 \leq 2k - 3 \leq p(G) \leq p(G^i)$$

for $i = 1, 2$. Write $A = A_0 \cup C$, where $C = C_1 \cup \ldots \cup C_t$,

$$A_0 = \{(a_i, b_i) \mid 1 \leq i \leq s\},\ C_i = \{(c_i, d_{ij}) \mid 1 \leq j \leq k_i\}$$

for $1 \leq i \leq t$ such that $2 \leq k_1 \leq k_2 \leq \ldots \leq k_t$, and $a_1, \ldots, a_s, c_1, \ldots, c_t$ are pairwise different elements of G^1. Note that $k = s + k_1 + \ldots + k_t$. The following easy lemma will be used frequently throughout the proof.

Lemma 5.2. *For $1 \leq \alpha, \beta \leq t$, $\alpha \neq \beta$ we have*

$$|C_\alpha \dot{+} C_\alpha| \geq 2k_\alpha - 3$$

and

$$|C_\alpha \dot{+} C_\beta| \geq k_\alpha + k_\beta - 1.$$

Proof. Since $|C_\alpha \dot{+} C_\alpha| = |C_\alpha^2 \dot{+} C_\alpha^2|$ and

$$2|C_\alpha^2| - 3 = 2k_\alpha - 3 \leq 2k - 3 \leq p(G) \leq p(G^2),$$

the first estimate follows directly from our hypothesis on G^2. On the other hand we have

$$|C_\alpha^2| + |C_\beta^2| - 1 = k_\alpha + k_\beta - 1 \leq 2k - 5 < p(G) \leq p(G^2),$$

and thus Theorem 1.1, applied to G^2, immediately implies

$$|C_\alpha \dot{+} C_\beta| = |C_\alpha^2 + C_\beta^2| \geq k_\alpha + k_\beta - 1.$$

□

Turning back to the proof of the estimate $|A\dot{+}A| \geq 2k-3$, assume first that $t=0$. In this case $|A_0^1| = s = k$ and

$$|A\dot{+}A| \geq |A_0^1\dot{+}A_0^1| \geq 2k-3$$

based on our assumption on the group G^1.

Assume next that $t \geq 4$. Consider the t numbers $c_i + c_t \in G^1$ for $1 \leq i \leq t$. Based on the hypothesis on G^1 we have $|C^1\dot{+}C^1| \geq 2t-3 \geq t+1$, and thus there exist indices $\alpha \neq \beta$ different from t such that $c_\alpha + c_\beta \in G^1$ differs from each number $c_i + c_t$. Then

$$|C_\alpha\dot{+}C_\beta| \geq k_\alpha + k_\beta - 1 \geq 3$$

by Lemma 5.2. Since $m = |C^1 + C^1| \geq 2t-1 > t+1$ by Theorem 1.1, there is a set I of $m-t-1$ pairs (γ,δ) such that the numbers

$$c_\alpha + c_\beta,\ c_i + c_t\ (1 \leq i \leq t),\ c_\gamma + c_\delta\ ((\gamma,\delta) \in I)$$

are all different. Lemma 5.2 implies $|C_\gamma\dot{+}C_\delta| \geq 1$ for these pairs (γ,δ). Based on Proposition 5.1, we can argue that

$$((A\dot{+}A) \setminus (C\dot{+}C))^1 \supseteq (A\dot{+}A)^1 \setminus (C\dot{+}C)^1 \supseteq (A^1\dot{+}A^1) \setminus (C^1 + C^1)$$

and consequently

$$\begin{aligned} |A\dot{+}A| &= |(A\dot{+}A) \setminus (C\dot{+}C)| + |C\dot{+}C| \\ &\geq |((A\dot{+}A) \setminus (C\dot{+}C))^1| + |C\dot{+}C| \\ &\geq |A^1\dot{+}A^1| - |C^1 + C^1| + |C\dot{+}C| \\ &\geq (2(s+t)-3) - m + |C\dot{+}C|, \end{aligned}$$

according to our hypothesis concerning $A^1 \subseteq G^1$. Based on our previous remarks and Lemma 5.2, we have

$$\begin{aligned} |C\dot{+}C| &\geq |C_\alpha\dot{+}C_\beta| + \sum_{(\gamma,\delta)\in I} |C_\gamma\dot{+}C_\delta| + \sum_{i=1}^{t} |C_i\dot{+}C_t| \\ &\geq 3 + (m-t-1) + \sum_{i=1}^{t-1}(k_i + k_t - 1) + (2k_t - 3) \\ &\geq (m-t+2) + 2\sum_{i=1}^{t} k_i - (t-1) - 3 = (m-2t) + 2(k-s). \end{aligned}$$

Consequently,

$$|A\dot{+}A| \geq (2s + 2t - 3 - m) + (m - 2t + 2k - 2s) = 2k-3,$$

as is was intended to prove. This completes the proof of the generic case $t \geq 4$.

The last case we study here is that of $t = 1$. As the remaining cases $t = 2$ and $t = 3$ require some more delicate analysis, these we postpone to the following two subsections, respectively. First we note that if $s = 0$, then $k_1 = k$, $A = C_1$ and

$$|A\dot{+}A| = |C_1\dot{+}C_1| \geq 2k_1 - 3 = 2k - 3$$

by Lemma 5.2. Otherwise we have $3 \leq s+2 \leq (k+2)-2$. Note that in this case $(A \setminus C)\dot{+}C = A_0\dot{+}C$ and $C\dot{+}C$ are disjoint, since $(g,h) \in C\dot{+}C$ implies $g = c_1 + c_1$, while $g = a_i + c_1$ for some $1 \leq i \leq s$ if $(g,h) \in A_0\dot{+}C$. Moreover, the elements $(a_i + c_1, b_i + d_{1j})$ are pairwise different for $1 \leq i \leq s$, $1 \leq j \leq k_1$, thus we obtain the estimate

$$\begin{aligned} |A\dot{+}A| &\geq |A\dot{+}C| = |A_0\dot{+}C| + |C\dot{+}C| \\ &\geq sk_1 + (2k_1 - 3) = s(k-s) + 2(k-s) - 3 \\ &= ((k+2) - (s+2))(s+2) - 3 \geq 2k - 3, \end{aligned}$$

as it was to be proved.

The Case $t = 2$

If $s = 0$, then $k = k_1 + k_2 \geq 4$. Since the numbers $c_1 + c_1$, $c_1 + c_2$ and $c_2 + c_2$ are pairwise distinct, we have

$$\begin{aligned} |A\dot{+}A| &\geq |C_1\dot{+}C_1| + |C_1\dot{+}C_2| + |C_2\dot{+}C_2| \\ &\geq (2k_1 - 3) + (k_1 + k_2 - 1) + (2k_2 - 3) = 3k - 7 \geq 2k - 3 \end{aligned}$$

by Lemma 5.2. Thus we may assume that $s \geq 1$. Then the numbers $a_i + c_2$ $(1 \leq i \leq s)$, $c_1 + c_2$ and $c_2 + c_2$ are all different, and thus

$$\begin{aligned} |A\dot{+}A| &\geq |A\dot{+}C_2| = |A_0\dot{+}C_2| + |C_1\dot{+}C_2| + |C_2\dot{+}C_2| \\ &\geq sk_2 + (k_1 + k_2 - 1) + (2k_2 - 3) \\ &\geq 2s + (k_2 - 2)s + 2(k_1 + k_2) - 4 \\ &= (2k - 4) + (k_2 - 2)s \geq 2k - 3, \end{aligned}$$

if $k_2 \geq 3$. Thus, in the sequel we will assume that $s \geq 1$ and $k_1 = k_2 = 2$. In particular, $k = s + 4$.

Consider the $2s + 1 = 2k - 7$ numbers $(a_i + c_2, b_i + d_{21})$, $(a_i + c_2, b_i + d_{22})$ $(1 \leq i \leq s)$, and $(c_2 + c_2, d_{21} + d_{22})$; they are all distinct, and also differ from the numbers $(c_1 + c_2, d_{11} + d_{21})$, $(c_1 + c_2, d_{11} + d_{22})$, $(c_1 + c_2, d_{12} + d_{21})$, $(c_1 + c_2, d_{12} + d_{22})$. Out of the latter four numbers at least 3 must be pairwise different. Thus we have found $2k - 3$ or $2k - 4$ different elements of $|A\dot{+}A|$ so far, denote the set of these elements by X.

If, for some $1 \le i \le s$,

$$a_i + c_1 \notin \{a_1 + c_2, \ldots, a_s + c_2, c_1 + c_2, c_2 + c_2\},$$

then $(a_i + c_1, b_i + d_{11}) \in (A\dot{+}A) \setminus X$, and therefore $|A\dot{+}A| \ge |X| + 1 \ge 2k - 3$. If $a_i + c_1 = c_2 + c_2$, then we may replace in X the element $(c_2 + c_2, d_{21} + d_{22})$ by the two new elements $(a_i + c_1, b_i + d_{11})$ and $(a_i + c_1, b_i + d_{12})$ to obtain at least $2k - 3$ different elements of $A\dot{+}A$. Since $a_i + c_1 = c_1 + c_2$ cannot occur, in any other case we conclude that

$$\{a_i + c_1 \mid 1 \le i \le s\} = \{a_i + c_2 \mid 1 \le i \le s\}.$$

This, however, is not possible, because in this case we would get $A_0^1 + c = A_0^1$ with $c = c_2 - c_1 \neq 0$, yielding

$$A_0^1 + (p(G) - 1)c = A_0^1 + (p(G) - 2)c = \ldots = A_0^1 + 2c = A_0^1 + c = A_0^1,$$

that in turn implies $p(G) \le |A_0^1| = s = k - 4 < 2k - 3 \le p(G)$, a contradiction.

Since we have considered all possibilities, the study of the case $t = 2$ is now complete.

The Case $t = 3$

The numbers $a_i + c_3$ $(1 \le i \le s)$, $c_1 + c_3$, $c_2 + c_3$ and $c_3 + c_3$ are all different, and thus

$$\begin{aligned} |A\dot{+}A| &\ge |A\dot{+}C_3| = |A_0\dot{+}C_3| + |C_1\dot{+}C_3| + |C_2\dot{+}C_3| + |C_3\dot{+}C_3| \\ &\ge sk_3 + (k_1 + k_3 - 1) + (k_2 + k_3 - 1) + (2k_3 - 3) \\ &= 2(s + k_1 + k_2 + k_3) - 5 + s(k_3 - 2) + (2k_3 - k_2 - k_1). \end{aligned}$$

Therefore $|A\dot{+}A| \ge 2k - 3$, whenever $s(k_3 - 2) \ge 2$. This is indeed the case if $k_3 \ge 3$ and $s \ge 2$.

Next, if $s \le 1$, then $k_1 + k_2 + k_3 \ge k - 1$, and $p(G) \ge 2k - 3 \ge 9$. The numbers $c_1 + c_2$, $c_1 + c_3$, $c_2 + c_3$ are pairwise different. By Theorem 1.1 we have

$$|\{c_1, c_2, c_3\} + \{c_1, c_2, c_3\}| \ge 5.$$

Consequently, there exist two indices $i \neq j$ such that the five numbers $c_1 + c_2$, $c_1 + c_3$, $c_2 + c_3$, $c_i + c_i$, $c_j + c_j$ are still pairwise different. Then, according to Lemma 5.11,

$$\begin{aligned} |A\dot{+}A| &\ge |C_1\dot{+}C_2| + |C_1\dot{+}C_3| + |C_2\dot{+}C_3| + |C_i\dot{+}C_i| + |C_j\dot{+}C_j| \\ &\ge (k_1 + k_2 - 1) + (k_1 + k_3 - 1) + (k_2 + k_3 - 1) + 1 + 1 \\ &= 2(k_1 + k_2 + k_3) - 1 \ge 2k - 3. \end{aligned}$$

It only remains to handle the case $k_1 = k_2 = k_3 = 2$, $s \ge 2$. Now we have $k = s + 6 \ge 8$, and then $p(G) \ge 2k - 3 \ge 13 > 2$.

Assume that there is no $1 \leq i \leq s$ such that $a_i + c_3 = c_1 + c_2$. Then the numbers $a_i + c_3$ $(1 \leq i \leq s)$, $c_1 + c_2$, $c_1 + c_3$ and $c_2 + c_3$ are all different, and

$$\begin{aligned} |A\dot{+}A| &\geq |A_0\dot{+}C_3| + |C_1\dot{+}C_2| + |C_1\dot{+}C_3| + |C_2\dot{+}C_3| \\ &\geq 2s + 3 + 3 + 3 = 2k - 3. \end{aligned}$$

Thus, we may assume that $a_i + c_3 = c_1 + c_2$ for some $1 \leq i \leq s$. By symmetry we may also suppose that $a_j + c_2 = c_1 + c_3$ for some $1 \leq j \leq s$. Were $i = j$, it would follow that

$$c_1 + c_2 - c_3 = a_i = a_j = c_1 + c_3 - c_2,$$

implying $2(c_3 - c_2) = 0$, in contradiction with $p(G) > 2$. Consequently, $i \neq j$.

Note that the numbers $a_\alpha + c_3$ $(1 \leq \alpha \leq s,\ \alpha \neq i)$, $c_1 + c_2$, $c_1 + c_3$ and $c_2 + c_3$ are still all different. If there is an index $1 \leq \beta \leq s$, $\beta \neq j$, such that

$$a_\beta + c_2 \notin \{a_1 + c_3, \ldots, a_s + c_3, c_1 + c_3, c_2 + c_3\},$$

then

$$\begin{aligned} |A\dot{+}A| &\geq |(a_\beta, b_\beta)\dot{+}C_2| + |(A_0 \setminus \{(a_i, b_i)\})\dot{+}C_3| \\ &+ |C_1\dot{+}C_2| + |C_1\dot{+}C_3| + |C_2\dot{+}C_3| \\ &\geq 2 + 2(s-1) + 3 + 3 + 3 = 2k - 3. \end{aligned}$$

Since for $1 \leq \beta \leq s$, $\beta \neq j$,

$$a_\beta + c_2 \notin \{a_i + c_3 = c_1 + c_2, c_1 + c_3, c_2 + c_3\},$$

in every other case we can conclude that

$$\{a_\alpha + c_3 \mid 1 \leq \alpha \leq s, \alpha \neq i\} = \{a_\beta + c_2 \mid 1 \leq \beta \leq s, \beta \neq j\}.$$

In particular, for every $\alpha \neq i$, $a_\alpha + (c_3 - c_2) \in A_0^1$.

Consider now the sequence defined recursively by

$$x_0 = a_i,\ x_{n+1} = x_n + c_3 - c_2\ (n \geq 0).$$

Then $x_1 = c_1$, $x_2 = a_j \in A_0^1 \setminus \{a_i\}$, and if $x_n \in A_0^1 \setminus \{a_i\}$, then $x_{n+1} \in A_0^1$ holds. It follows that there is a smallest positive integer n for which there exists an integer $0 \leq m < n$ such that $x_n = x_m$, and in this case $x_{m+1}, x_{m+2}, \ldots, x_n$ are all different elements of $A_0^1 \cup \{c_i\}$. Consequently,

$$1 \leq n - m \leq |A_0^1| + 1 = s + 1 < k < p(G),$$

which contradicts the fact that

$$(n - m)(c_3 - c_2) = x_n - x_m = 0.$$

This completes the investigation of the case $t = 3$ and also the proof of Theorem 2.11.

A more simple proof of Theorem 2.11 can be found in [51]. To avoid repetitions we do not include it here. However, it is not clear how to apply the method of the previous chapter in a multiplicative setting. Thus, to prove Theorem 2.15 we will need an additional idea. That is exactly the novelty contained in [51], which will be clear from the following section.

5.2 Inverse Theorems in Abelian Groups

Since A is contained in a finitely generated subgroup H of G, and obviously $p(H) \geq p(G)$, it is enough to prove Theorems 2.4 and 2.15 in the case when G is finitely generated. In this case we can write

$$G = G^1 \oplus G^2 \oplus \ldots \oplus G^m,$$

where each group G^i is isomorphic either to the infinite cyclic group $\mathbb{Z}$ or to a cyclic group $\mathbb{Z}/p^\alpha\mathbb{Z}$ with some prime number $p \geq p(G)$ and positive integer α. Note that here $p(\mathbb{Z}) = \infty$ while $p(G) = p$ if $G \cong \mathbb{Z}/p^\alpha\mathbb{Z}$. Moreover,

$$p(G^1 \oplus G^2) = \min\{p(G^1), p(G^2)\}.$$

If a set G is equipped with a binary operation '+', then we can naturally talk about arithmetic progressions in G: the sequence $(a_1, a_2, \ldots, a_k)$ is an arithmetic progression in G, if there exists $d \in G$ such that $a_i = a_{i-1} + d$ for $i = 2, \ldots, k$. For simplicity we will call $\langle G, + \rangle$ an *additive structure*. The notations $A + B$ and $A\dot{+}B$ can also be naturally extended to such structures.

Definition 5.3. *Let ℓ denote a positive integer. We say that the additive structure $\langle G, + \rangle$ has* property Π_ℓ *if*

(i) for any positive integer $k \leq \ell$ and a k-element subset A of G, $|A + A| \geq 2k - 1$ with equality if and only if A is an arithmetic progression in G;

(ii) for any positive integer $k \leq \ell + 1$ and a k-element subset A of G, $|A\dot{+}A| \geq 2k - 3$ with equality (in case of $k \geq 5$) if and only if A is an arithmetic progression in G.

We have seen that the group $\mathbb{Z}$ has property Π_ℓ for every positive integer ℓ. According to the Cauchy–Davenport theorem and Theorems 2.3, 2.9 and 2.14, the group $\mathbb{Z}/p\mathbb{Z}$ has property Π_ℓ whenever p is a prime number greater than $2\ell - 1$. In view of all this, to prove Theorems 2.4 and 2.15 it is enough to verify the following two statements. Note that Theorem 2.4 is obvious if $k = 1$.

Statement 5.4. *Let G^1 and G^2 be two abelian groups such that*

$$\min\{p(G^1), p(G^2)\} > 2\ell - 1 \geq 3.$$

If G^1 and G^2 have property Π_ℓ, then so does their direct sum $G^1 \oplus G^2$.

Statement 5.5. *Let $\alpha \geq 1$ and $\ell \geq 2$ be integers and let $p > 2\ell - 1$ be a prime number. If the group $\mathbb{Z}/p^\alpha\mathbb{Z}$ has property Π_ℓ, then so does the group $\mathbb{Z}/p^{\alpha+1}\mathbb{Z}$.*

The key observation is that we can verify both statements using the same argument, based on the following notion. Let G^1 and G^2 be two abelian groups, and let $\varphi : G^1 \times G^1 \to G^2$ be any map. On the set of all ordered pairs (g^1, g^2) $(g^1 \in G^1, g^2 \in G^2)$, define an additive structure $\langle G_\varphi, +_\varphi \rangle$ by introducing a binary operation $+_\varphi$ as follows:

$$(g^1, g^2) +_\varphi (h^1, h^2) =: (g^1 + h^1, g^2 + h^2 + \varphi(g^1, h^1)).$$

Note that if the map φ is symmetric, then the operation $+_\varphi$ is commutative. Now Statements 5.4 and 5.5 can be easily derived from the following lemma.

Lemma 5.6. *Let $\ell \geq 2$ be any integer and assume that the abelian groups G^1 and G^2 satisfy*

$$\min\{p(G^1), p(G^2)\} > 2\ell - 1 \geq 3.$$

Let furthermore $\varphi : G^1 \times G^1 \to G^2$ be any symmetric map satisfying $\varphi(g, 0) = 0$ for every $g \in G^1$ such that the additive structure $G_\varphi = \langle G_\varphi, +_\varphi \rangle$ is a group. If G^1 and G^2 have property Π_ℓ, then the abelian group G_φ also has property Π_ℓ.

Indeed, letting $\varphi \equiv 0$ we get back the notion of direct sum: $G_\varphi \cong G^1 \oplus G^2$. Thus, Statement 5.4 follows immediately. On the other hand, if we choose $G^1 = \mathbb{Z}/p\mathbb{Z}$, $G^2 = \mathbb{Z}/p^\alpha\mathbb{Z}$ for a prime $p > 2\ell - 1$, and we define

$$\varphi(x + p\mathbb{Z}, y + p\mathbb{Z}) = \begin{cases} 0 & \text{if } x + y < p \\ 1 & \text{otherwise} \end{cases}$$

for $x, y \in \{0, 1, \ldots, p-1\}$, then $G_\varphi \cong \mathbb{Z}/p^{\alpha+1}\mathbb{Z}$. Namely, if we define

$$f(a + p/\mathbb{Z}, b + p^\alpha/\mathbb{Z}) = (pb + a) + p^{\alpha+1}/\mathbb{Z}$$

for $a \in \{0, 1, \ldots, p-1\}$ and $b \in \{0, 1, \ldots, p^\alpha - 1\}$, then f maps the set $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p^\alpha\mathbb{Z}$ bijectively onto the set $\mathbb{Z}/p^{\alpha+1}\mathbb{Z}$, and clearly is a homomorphism from G_φ to $\mathbb{Z}/p^{\alpha+1}\mathbb{Z}$. Since $\mathbb{Z}/p\mathbb{Z}$ has property Π_ℓ, Lemma 5.6 implies Statement 5.5 as well. It only remains to prove Lemma 5.6.

Proof of Lemma 5.6. Note that the condition $\varphi(g, 0) = \varphi(0, g) = 0$ implies

Proposition 5.7. *If $(a_1, a_2, \ldots, a_k)$ is an arithmetic progression in G^2, then*

$$((g, a_1), (g, a_2), \ldots, (g, a_k))$$

is an arithmetic progression in the abelian group G_φ for any $g \in G^1$.

For a set $X \subseteq G_\varphi$ write

$$X^1 = \{g^1 \in G^1 \mid \text{there exists } g^2 \in G^2 \text{ with } (g^1, g^2) \in X\}.$$

We define X^2 in a similar way. For $A, B \subseteq G_\varphi$ we also introduce

$$A + B =: \{a +_\varphi b \mid a \in A, b \in B\}$$

and

$$A\dot{+}B =: \{a +_\varphi b \mid a \in A, b \in B, a \neq b\}.$$

In the sequel we will simply write '+' for '$+_\varphi$'. An immediate consequence of these definitions is the following statement.

Proposition 5.8. *For arbitrary $X, Y \subseteq G_\varphi$ we have $(X \setminus Y)^1 \supseteq X^1 \setminus Y^1$ and $X^1\dot{+}X^1 \subseteq (X\dot{+}X)^1 \subseteq X^1 + X^1$.*

The careful reader may observe that the second part of the statement does not remain valid in general if, instead of the projection to the first coordinate, one considers the projection to the second one. We will also need the following easy lemma.

Lemma 5.9. *Let G^1, G^2, φ and ℓ as in Lemma 5.6. Assume that $(a_1, a_2, \dots, a_k)$ is a non-constant arithmetic progression in G^1 and let $b_1, b_2, \dots, b_k \in G^2$. Consider the set*

$$A = \{g_i = (a_i, b_i) \mid 1 \leq i \leq k\} \subset G_\varphi.$$

(i) If $k \leq \ell$ and $|A + A| = 2k - 1$, then A is an arithmetic progression in G_φ.

(ii) If $5 \leq k \leq \ell + 1$ and $|A\dot{+}A| = 2k - 3$, then A is an arithmetic progression in G_φ.

Proof. For $1 \leq i \leq k$, introduce $d_i = g_{i+1} - g_i \in G_\varphi$. Write $a_1 = a$ and $a_2 - a_1 = d$, then in case (i)

$$(A + A)^1 = \{2a, 2a + d, 2a + 2d, \dots, 2a + (2k - 2)d\}$$

whereas in case (ii)

$$(A\dot{+}A)^1 = \{2a + d, 2a + 2d, \dots, 2a + (2k - 3)d\},$$

the containment $\supseteq$ being obvious from the definition and the assumption $p(G^1) > 2\ell - 1$. To prove the first statement we may assume that $k \geq 3$. For every $1 \leq i \leq k - 2$, $g_i + g_{i+2}$ and $g_{i+1} + g_{i+1}$ have the same first coordinate $2a + 2id$. According to the assumption $|A + A| = 2k - 1$, these elements of G_φ must be equal. Consequently,

$$2g_i + d_i + d_{i+1} = 2g_i + 2d_i.$$

It follows that $d_1 = d_2 = \ldots = d_{k-1}$, and $g_1, g_2, \dots, g_k$ is indeed an arithmetic progression.

Similarly, in case (ii) we can argue that

$$g_i + g_{i+3} = g_{i+1} + g_{i+2}$$

for every $1 \le i \le k-3$, implying

$$2g_i + d_i + d_{i+1} + d_{i+2} = 2g_i + 2d_i + d_{i+1}.$$

Consequently, we have that $d_{i+2} = d_i$ for every $1 \le i \le k-3$. Moreover, since $k \ge 5$, we have $g_1 + g_5 = g_2 + g_4$, that is,

$$2g_1 + d_1 + d_2 + d_3 + d_4 = 2g_1 + 2d_1 + d_2 + d_3.$$

Therefore $d_1 = d_4$, which completes the proof of the second statement. □

We conclude this subsection by proving that $\langle G_\varphi, +_\varphi \rangle$ satisfies condition (i) of property Π_ℓ. Remark that the proof below does not depend on the hypothesis that the groups G^1, G^2 satisfy condition (ii) as well, thus it can be read as a self-contained proof of Theorem 2.4. That is, we prove that if G^1, G^2 satisfy (i) of Π_ℓ, then so does G_φ.

Thus let A denote a k-element subset of G_φ. The cases $k = 1, 2$ being obvious, assume that $3 \le k \le \ell$. Write $A = A_0 \cup C$, where $C = C_1 \cup \ldots \cup C_t$,

$$A_0 = \{(a_i, b_i) \mid 1 \le i \le s\},\ C_i = \{(c_i, d_{ij}) \mid 1 \le j \le k_i\}$$

for $1 \le i \le t$ such that $2 \le k_1 \le k_2 \le \ldots \le k_t$, and $a_1, \ldots, a_s, c_1, \ldots, c_t$ are pairwise different elements of G^1. In particular, $k = s + k_1 + \ldots + k_t$ and $|A^1| = s + t$. The following easy lemma will be used frequently throughout the proof.

Lemma 5.10. *For $1 \le \alpha, \beta \le t$ we have $|C_\alpha + C_\beta| \ge k_\alpha + k_\beta - 1$. Moreover, in the case $\alpha = \beta$, equality holds if and only if C_α^2 is an arithmetic progression in G^2.*

Proof. Adding $\varphi(c_\alpha, c_\beta)$ to each element of $C_\alpha^2 + C_\beta^2$, we obtain the set $(C_\alpha + C_\beta)^2$. Consequently, $|C_\alpha + C_\beta| = |(C_\alpha + C_\beta)^2| = |C_\alpha^2 + C_\beta^2|$. Since

$$|C_\alpha^2| + |C_\beta^2| - 1 = k_\alpha + k_\beta - 1 \le 2k - 1 \le 2\ell - 1 < p(G^2),$$

the estimate follows from Theorem 1.1. Since $k_\alpha \le k \le \ell$, in the case $|C_\alpha^2 + C_\alpha^2| = 2k_\alpha - 1$ it follows from our hypothesis on G^2 that C_α^2 is an arithmetic progression in G^2. On the other hand, if this is the case, then Proposition 5.7 implies that C_α itself is an arithmetic progression in G_φ, consequently $|C_\alpha + C_\alpha| \le 2k_\alpha - 1$. □

Assume first that $t \ge 2$. The numbers $c_i + c_t$ $(1 \le i \le t)$ are t distinct elements of $C^1 + C^1$. It follows from Theorem 1.1 that $|C^1 + C^1| \ge 2t - 1$, and thus there is a set I of $t-1$ pairs (γ, δ) such that the numbers

$$c_i + c_t\ (1 \le i \le t),\ c_\gamma + c_\delta\ ((\gamma, \delta) \in I)$$

are all different. Lemma 5.10 implies $|C_\gamma + C_\delta| \geq 3$ for these pairs (γ, δ). It follows that the sets

$$C_i + C_t \ (1 \leq i \leq t), \ C_\gamma + C_\delta \ ((\gamma, \delta) \in I)$$

are pairwise disjoint subsets of $A + A$. Moreover, since $s + t \leq k \leq \ell$, we have $|A^1 + A^1| \geq 2(s + t) - 1$ and thus there exist at least $2s$ elements of $A + A$ whose first coordinates are different from the numbers

$$c_i + c_t \ (1 \leq i \leq t), \ c_\gamma + c_\delta \ ((\gamma, \delta) \in I).$$

Based on Lemma 5.10 and the inequalities $k_i \leq k_t$ for $1 \leq i \leq t$, we then indeed obtain

$$\begin{aligned} |A + A| &\geq 2s + \sum_{(\gamma,\delta)\in I} |C_\gamma + C_\delta| + \sum_{i=1}^{t} |C_i + C_t| \\ &\geq 2s + 3(t-1) + \sum_{i=1}^{t} (k_i + k_t - 1) \\ &\geq 2s + 2\sum_{i=1}^{t} k_i + 2t - 3 > 2k - 1. \end{aligned}$$

Next assume that $t = 0$, that is, $|A_0^1| = s = k$. Then we have

$$|A + A| \geq |A_0^1 + A_0^1| \geq 2k - 1$$

according to our assumption on the group G^1. Moreover, $|A_0^1 + A_0^1| = 2k - 1$ if and only if A_0^1 is an arithmetic progression in G^1. Consequently, if $|A + A| = 2k - 1$, we can apply Lemma 5.9 (i) to find that A is an arithmetic progression in G_φ.

If $t = 1$ and $s = 0$, then it follows from Lemma 5.10 that

$$|A + A| = |C_1 + C_1| \geq 2k_1 - 1 = 2k - 1,$$

where equality holds if and only if C_1^2 is an arithmetic progression in G^2. Note that in this case $A = C_1$ is an arithmetic progression in G_φ, according to Proposition 5.7.

Suppose finally that $t = 1$ and $s \geq 1$, then we have $3 \leq s + 2 \leq (k + 2) - 2$. Note that in this case $(A \setminus C) + C = A_0 + C$ and $C + C$ are disjoint, since $(g^1, g^2) \in C + C$ implies $g^1 = c_1 + c_1$, while $g^1 = a_i + c_1$ for some $1 \leq i \leq s$ if $(g^1, g^2) \in A_0 + C$. Moreover, the elements $(a_i + c_1, b_i + d_{1j})$ are pairwise different for $1 \leq i \leq s$, $1 \leq j \leq k_1$, thus we obtain the inequality

$$\begin{aligned} |A + A| &\geq |A + C| = |A_0 + C| + |C + C| \\ &\geq sk_1 + (2k_1 - 1) = s(k - s) + 2(k - s) - 1 \\ &= ((k + 2) - (s + 2))(s + 2) - 1 \geq 2k - 1, \end{aligned}$$

proving the estimate. Now we prove that $|A + A| = 2k - 1$ is not possible in this case. Indeed, it only could happen if it were $s + 2 = k$, that is, $k_1 = 2$, in which case we could argue as

follows. Since $|A^1| = k-1$, we have $|A^1 + A^1| \geq 2k-3$, according to our assumption on the group G^1. Therefore the elements of $A + A$ have at least $2k-3$ different first coordinates. One of those is $c_1 + c_1$, to which correspond (at least) three different second coordinates:

$$d_{11} + d_{11} + \varphi(c_1, c_1), d_{11} + d_{12} + \varphi(c_1, c_1), d_{12} + d_{12} + \varphi(c_1, c_1).$$

Another one is $a_1 + c_1$, with two different second coordinates

$$b_1 + d_{11} + \varphi(a_1, c_1), b_1 + d_{12} + \varphi(a_1, c_1).$$

This way we found at least $2k$ different elements of $A + A$.

Thus we have overviewed all possible cases and found that in every case $|A + A| \geq 2k-1$ and $|A + A| = 2k-1$ can only happen if A is an arithmetic progression in G_φ. Noting that if A is an arithmetic progression then obviously $|A + A| \leq 2k-1$, we find that $\langle G_\varphi, +_\varphi \rangle$ indeed satisfies condition (i) of property Π_ℓ.

Proof of Lemma 5.6, continued

The aim of this section is to prove that $\langle G_\varphi, +_\varphi \rangle$ satisfies condition (ii) of property Π_ℓ, thus completing the proof of Lemma 5.6. For this end let A denote a k-element subset of G_φ. Since we have already discussed the case $k \leq 4$ on Page 17, we will assume that $5 \leq k \leq \ell + 1$. Keeping the notation of the previous section, we first verify the following analogue of Lemma 5.10.

Lemma 5.11. *Let $1 \leq \alpha, \beta \leq t$, $\alpha \neq \beta$. Then $|C_\alpha \dot{+} C_\beta| \geq k_\alpha + k_\beta - 1$. Moreover, $|C_\alpha \dot{+} C_\alpha| \geq 2k_\alpha - 3$, where in the case $k_\alpha \geq 5$ equality holds if and only if C_α^2 is an arithmetic progression in G^2.*

Proof. Since $C_\alpha \dot{+} C_\beta = C_\alpha + C_\beta$, the first estimate follows as in the proof of Lemma 5.10, noting that this time

$$k_\alpha + k_\beta - 1 \leq k - 1 \leq \ell < p(G^2).$$

On the other hand, adding $\varphi(c_\alpha, c_\alpha)$ to each element of $C_\alpha^2 \dot{+} C_\alpha^2$, we obtain the set $(C_\alpha \dot{+} C_\alpha)^2$. Consequently, $|C_\alpha \dot{+} C_\alpha| = |(C_\alpha \dot{+} C_\alpha)^2| = |C_\alpha^2 \dot{+} C_\alpha^2|$. Since $k_\alpha \leq k \leq \ell + 1$, the second statement follows directly from our hypothesis on G^2. □

Assume first that $t = 0$, that is, $|A_0^1| = s = k$. Then we have

$$|A + A| \geq |A_0^1 + A_0^1| \geq 2k - 3$$

according to our assumption on the group G^1. Moreover, $|A_0^1 + A_0^1| = 2k-3$ if and only if A_0^1 is an arithmetic progression in G^1. Consequently, if $|A + A| = 2k-3$, we can apply Lemma 5.9 (ii) to find that A is an arithmetic progression in G_φ.

Next we assume that $t \geq 4$. Consider the t numbers $c_i + c_t \in G^1$ for $1 \leq i \leq t$. Based on the hypothesis on G^1 we have $|C^1 \dot{+} C^1| \geq 2t - 3 \geq t + 1$, and thus there exist indices $\alpha \neq \beta$ different from t such that $c_\alpha + c_\beta \in G^1$ differs from each number $c_i + c_t$. Then

$$|C_\alpha \dot{+} C_\beta| \geq k_\alpha + k_\beta - 1 \geq 3$$

by Lemma 5.11. Since $m = |C^1 + C^1| \geq 2t - 1 > t + 1$ by Theorem 1.1, there is a set I of $m - t - 1$ pairs (γ, δ) such that the numbers

$$c_\alpha + c_\beta,\ c_i + c_t\ (1 \leq i \leq t),\ c_\gamma + c_\delta\ ((\gamma, \delta) \in I)$$

are all different. Lemma 5.11 implies $|C_\gamma \dot{+} C_\delta| \geq 1$ for these pairs (γ, δ). Based on Proposition 5.8, we can argue that

$$((A \dot{+} A) \setminus (C \dot{+} C))^1 \supseteq (A \dot{+} A)^1 \setminus (C \dot{+} C)^1 \supseteq (A^1 \dot{+} A^1) \setminus (C^1 + C^1)$$

and consequently

$$\begin{aligned} |A \dot{+} A| &= |(A \dot{+} A) \setminus (C \dot{+} C)| + |C \dot{+} C| \\ &\geq |((A \dot{+} A) \setminus (C \dot{+} C))^1| + |C \dot{+} C| \\ &\geq |A^1 \dot{+} A^1| - |C^1 + C^1| + |C \dot{+} C| \\ &\geq (2(s + t) - 3) - m + |C \dot{+} C|, \end{aligned}$$

according to our hypothesis concerning $A^1 \subseteq G^1$. Based on our previous remarks and Lemma 5.11, we have

$$\begin{aligned} |C \dot{+} C| &\geq |C_\alpha \dot{+} C_\beta| + \sum_{(\gamma,\delta) \in I} |C_\gamma \dot{+} C_\delta| + \sum_{i=1}^{t} |C_i \dot{+} C_t| \\ &\geq 3 + (m - t - 1) + \sum_{i=1}^{t-1} (k_i + k_t - 1) + (2k_t - 3) \\ &\geq (m - t + 2) + 2 \sum_{i=1}^{t} k_i - (t - 1) - 3 = (m - 2t) + 2(k - s). \end{aligned}$$

Putting these estimates together we obtain that

$$|A \dot{+} A| \geq (2s + 2t - 3 - m) + (m - 2t + 2k - 2s) = 2k - 3,$$

as it was intended to prove. Now we proceed to show that in fact $|A \dot{+} A| > 2k - 3$ in this case. If $k_1 < k_t$, then we can immediately increase the estimate on $|C \dot{+} C|$ and thus on $|A \dot{+} A|$ as well. On the other hand, if $k_1 = k_2 = \ldots = k_t$, then we can argue as follows. First, since $|C_1 \dot{+} C_1| \geq 2t - 3$, there is a set J of $2t - 3$ pairs $(\alpha, \beta), \alpha \neq \beta$ such that the numbers $c_\alpha + c_\beta,\ ((\alpha, \beta) \in J)$ are all different. It follows from Lemma 5.11 that $|C_\alpha \dot{+} C_\beta| \geq 2k_t - 1$ for

$(\alpha,\beta)\in J$. Next, since $m\geq 2t-1>|J|$, there is a set K of $m-2t+3$ pairs (γ,δ) such that the numbers

$$c_\alpha+c_\beta,\ ((\alpha,\beta)\in J),\ c_\gamma+c_\delta\ ((\gamma,\delta)\in K)$$

are all different. For $(\gamma,\delta)\in K$ we have the estimate $|C_\gamma\dot{+}C_\delta|\geq 2k_t-3$. Consequently,

$$\begin{aligned}|C\dot{+}C| &\geq \sum_{(\alpha,\beta)\in J}|C_\alpha\dot{+}C_\beta|+\sum_{(\gamma,\delta)\in K}|C_\gamma\dot{+}C_\delta|\\ &\geq (2t-3)(2k_t-1)+(m-2t+3)(2k_t-3)\\ &= 2mk_t-3m+4t-6.\end{aligned}$$

It follows that

$$\begin{aligned}|A\dot{+}A| &\geq (2(s+t)-3)-m+|C\dot{+}C|\\ &\geq 2(k-tk_t)+2t-3-m+(2mk_t-3m+4t-6)\\ &= 2k+(m-t)2k_t-4m+6t-9\\ &= 2k+(m-t)(2k_t-4)+(2t-9)\geq 2k-1.\end{aligned}$$

This completes the proof for the generic case $t\geq 4$.

The next case we study is that of $t=1$. If $s=0$, then it follows from Lemma 5.11 that

$$|A\dot{+}A|=|C_1\dot{+}C_1|\geq 2k_1-3=2k-3,$$

where equality holds if and only if C_1^2 is an arithmetic progression in G^2. Note that in this case $A=C_1$ is an arithmetic progression in G_φ, according to Proposition 5.7. If $t=1$ and $s\geq 1$, then we have $3\leq s+2\leq (k+2)-2$. Note that in this case $(A\setminus C)\dot{+}C=A_0\dot{+}C$ and $C\dot{+}C$ are disjoint. Moreover, the elements (a_i+c_1,b_i+d_{1j}) are pairwise different for $1\leq i\leq s$, $1\leq j\leq k_1$, thus we obtain the estimate

$$\begin{aligned}|A\dot{+}A| &\geq |A\dot{+}C|=|A_0\dot{+}C|+|C\dot{+}C|\\ &\geq sk_1+(2k_1-3)=s(k-s)+2(k-s)-3\\ &= ((k+2)-(s+2))(s+2)-3\geq 2k-3,\end{aligned}$$

proving the estimate. Now we prove that $|A\dot{+}A|=2k-3$ is not possible in this case. Indeed, it only could happen if it were $s+2=k$, that is, $k_1=2$, in which case we could argue as follows. Since $|A^1|=k-1$, we have $|A^1\dot{+}A^1|\geq 2k-5$, according to our assumption on the group G^1. Therefore the elements of $A\dot{+}A$ have at least $2k-5$ different first coordinates. Since $k\geq 5$, that is, $s\geq 3$, at least three of these first coordinates are in the form a_i+c_1 for some $1\leq i\leq s$. To each of these correspond two different second coordinates

$$b_i+d_{11}+\varphi(a_i,c_1),b_i+d_{12}+\varphi(a_i,c_1).$$

This way we found at least $2k-2$ different elements of $A\dot{+}A$.

Next we will show that if $t = 2$, then $|A\dot{+}A| \geq 2k-2$. Assume first that $s = 0$, that is, $k = k_1 + k_2 \geq 5$. Since the numbers $c_1 + c_1$, $c_1 + c_2$ and $c_2 + c_2$ are pairwise distinct, we have

$$\begin{aligned} |A\dot{+}A| &\geq |C_1\dot{+}C_1| + |C_1\dot{+}C_2| + |C_2\dot{+}C_2| \\ &\geq (2k_1 - 3) + (k_1 + k_2 - 1) + (2k_2 - 3) = 3k - 7 \geq 2k - 2 \end{aligned}$$

by Lemma 5.11. Thus we may assume that $s \geq 1$. Then the numbers $a_i + c_2$ $(1 \leq i \leq s)$, $c_1 + c_2$ and $c_2 + c_2$ are all different, and thus

$$\begin{aligned} |A\dot{+}A| &\geq |A\dot{+}C_2| = |A_0\dot{+}C_2| + |C_1\dot{+}C_2| + |C_2\dot{+}C_2| \\ &\geq sk_2 + (k_1 + k_2 - 1) + (2k_2 - 3) \\ &= 2s + (k_2 - 2)s + 2(k_1 + k_2) + (k_2 - k_1) - 4 \\ &= (2k - 4) + (k_2 - 2)s + (k_2 - k_1) \geq 2k - 2, \end{aligned}$$

unless $k_1 = k_2 = 2$, or $s = 1$ and $k_1 = k_2 = 3$. In the latter case either $a_1 + c_1$ or $a_1 + c_2$ does not belong to the set that consists of the three distinct numbers $c_1 + c_1, c_1 + c_2, c_2 + c_2$. Indeed, otherwise we would have $a_1 + c_1 = c_2 + c_2$ and $a_1 + c_2 = c_1 + c_1$, which implies $3(c_2 - c_1) = 0$, contradicting $p(G^1) > 3$. Hence we may assume without any loss of generality that the numbers $a_1 + c_1, c_1 + c_1, c_1 + c_2, c_2 + c_2$ are pairwise distinct, in which case by Lemma 5.11

$$\begin{aligned} |A\dot{+}A| &\geq |A_0\dot{+}C_1| + |C_1\dot{+}C_1| + |C_1\dot{+}C_2| + |C_2\dot{+}C_2| \\ &\geq 2 + 3 + 5 + 3 > 12 = 2k - 2. \end{aligned}$$

If $k_1 = k_2 = 2$ and $s \geq 3$, then $c_1 + c_2, a_1 + c_2, a_2 + c_2$ and $a_3 + c_2$ are 4 pairwise disjoint elements of $A^1\dot{+}A^1$. These elements are first coordinates of at least $3, 2, 2$ and 2 elements of $A\dot{+}A$, respectively. Since $|A^1| = k - 2$, we have $|A^1\dot{+}A^1| \geq 2k - 7$, based on our hypothesis on the group G^1. Given that $2k - 7 > 4$, there are at least $(2k - 7) - 4$ elements of $A\dot{+}A$ whose first coordinates do not belong to the set

$$\{c_1 + c_2, a_1 + c_2, a_2 + c_2, a_3 + c_2\}.$$

This way we found $3+2+2+2+(2k-11) = 2k-2$ different elements of $A\dot{+}A$. If $k_1 = k_2 = 2$ and $s = 1$, that is, $k = 5$, then in $A\dot{+}A$ we can respectively find $3, 2$ and 2 elements whose first coordinates are $c_1 + c_2, a_1 + c_1$ and $a_1 + c_2$, in this order. It cannot happen that both $c_1 + c_1$ and $c_2 + c_2$ belong to the set $\{c_1 + c_2, a_1 + c_1, a_1 + c_2\}$, since it would imply that $a_1 + c_1 = c_2 + c_2$ and $a_1 + c_2 = c_1 + c_1$, and we have already seen the contradiction arising from that. Therefore, in addition to the 7 elements of $A\dot{+}A$ we have already found, there is at least one more element of $A\dot{+}A$ whose first coordinate is either $c_1 + c_1$ or $c_2 + c_2$, that is, $|A\dot{+}A| \geq 8 = 2k - 2$, as claimed. If $k_1 = k_2 = 2$ and $s = 2$, that is, $k = 6$, then $|A^1| = 4$ and thus $|A^1\dot{+}A^1| \geq 5$. The number $c_1 + c_2$ is among the elements of $A^1\dot{+}A^1$ as well as the

four numbers $a_i + c_j$ $(1 \le i, j \le 2)$. At least three of the last four numbers must be different, otherwise we would have $a_1 + c_1 = a_2 + c_2$ and $a_1 + c_2 = a_2 + c_1$, leading to the contradiction $2(c_1 - c_2) = 0$. Thus we can choose three such numbers; each of which is the first coordinate of at least 2 elements of $A\dot{+}A$. On the other hand, the number $c_1 + c_2$, which is definitely different from the previous three numbers, is the first coordinate of at least 3 elements of $A\dot{+}A$. So far we have found at least 9 elements of $A\dot{+}A$, but they only have 4 different first coordinates. Since $|(A\dot{+}A)^1| \ge |A^1\dot{+}A^1| \ge 5$, there must be at least one more element in $A\dot{+}A$, that is, $|A\dot{+}A| \ge 10 = 2k - 2$ follows in this case, too.

Finally we discuss the case $t = 3$. First suppose that $s \ge 2$. The numbers $c_1 + c_2, c_1 + c_3, c_2 + c_3$ are pairwise distinct. Among the numbers $a_i + c_1$ $(1 \le i \le s)$ at most one can be equal to $c_2 + c_3$ and none is equal to $c_1 + c_2$ or $c_1 + c_3$. Thus there is a set I of $s - 1$ indices such that the numbers

$$c_1 + c_2,\ c_1 + c_3,\ c_2 + c_3,\ a_i + c_1\ (i \in I)$$

are $s + 2$ different elements of $A^1\dot{+}A^1$. Since $|A^1| = s + 3$, based on the assumption on the group G^1 we have that $|A^1\dot{+}A^1| \ge 2s + 3$, and thus there are at least $s + 1$ elements of $A\dot{+}A$ whose first coordinates are not among the above numbers. It follows that

$$\begin{aligned} |A\dot{+}A| &\ge (s+1) + \sum_{i\in I} |\{a_i\} + C_1| + |C_1\dot{+}C_2| + |C_1\dot{+}C_3| + |C_2\dot{+}C_3| \\ &\ge (s+1) + 2(s-1) + (k_1 + k_2 - 1) + (k_1 + k_3 - 1) + (k_2 + k_3 - 1) \\ &= 2(k-s) + 3s - 4 = 2k + s - 4 \ge 2k - 2. \end{aligned}$$

If $s \le 1$, then we can do the following. The numbers $c_1 + c_2$, $c_1 + c_3$, $c_2 + c_3$ are pairwise different. By Theorem 1.1 we have

$$|\{c_1, c_2, c_3\} + \{c_1, c_2, c_3\}| \ge 5.$$

Consequently, there exist two indices $i \ne j$ such that the five numbers $c_1 + c_2$, $c_1 + c_3$, $c_2 + c_3$, $c_i + c_i$, $c_j + c_j$ are still pairwise different. Then, according to Lemma 5.11,

$$\begin{aligned} |A\dot{+}A| &\ge |C_1\dot{+}C_2| + |C_1\dot{+}C_3| + |C_2\dot{+}C_3| + |C_i\dot{+}C_i| + |C_j\dot{+}C_j| \\ &\ge (k_1 + k_2 - 1) + (k_1 + k_3 - 1) + (k_2 + k_3 - 1) + 1 + 1 \\ &= 2(k_1 + k_2 + k_3) - 1. \end{aligned}$$

Thus we have $|A\dot{+}A| \ge 2k - 1$ if $s = 0$, and $|A\dot{+}A| \ge 2k - 3$ if $s = 1$. In the latter case we can immediately increase the estimate, whenever $|(A\dot{+}A)^1| > 5$. On the other hand, if $|(A\dot{+}A)^1| = 5$, then the numbers $a_1 + c_1, a_1 + c_2, a_1 + c_3$ belong to the set

$$\{c_1 + c_2, c_1 + c_3, c_2 + c_3, c_i + c_i, c_j + c_j\}.$$

If $a_1 + c_\alpha = c_\beta + c_\beta$ for some $\alpha \in \{1,2,3\}$ and $\beta \in \{i,j\}$, then we can replace $|C_\beta \dot{+} C_\beta|$ by $|\{a_1\} + C_\alpha| = k_\alpha \geq 2$ in the above estimate to conclude that $|A \dot{+} A| \geq 2k-2$. Were it not the case we would obtain that

$$a_1 + c_1 = c_2 + c_3, a_1 + c_2 = c_1 + c_3, a_1 + c_3 = c_1 + c_2,$$

resulting in the contradiction $2(c_1 - c_2) = 0$. Therefore we have $|A \dot{+} A| \geq 2k-2$ whenever $t = 3$.

All in all, we found that in every case $|A \dot{+} A| \geq 2k-3$ and $|A \dot{+} A| = 2k-3$ can only happen if A is an arithmetic progression in G_φ. Noting that if A is an arithmetic progression then obviously $|A \dot{+} A| \leq 2k-3$, we find that $\langle G_\varphi, +_\varphi \rangle$ indeed satisfies condition (ii) of property Π_ℓ. This completes the proof of Lemma 5.6, and in turn that of Theorem 2.15 as well.

5.3 Noncommutative Groups

Hidden behind the previous proof is the fact that for any prime p, the group $\mathbb{Z}_{p^{\alpha+1}}$ can be obtained as a cyclic extension of the group $\mathbb{Z}_{p^\alpha}$ by the group $\mathbb{Z}_p$. This eluded us previously and only became clear later, leading to the proof of Theorem 2.1 first and then to that of Theorem 2.6. Based on the theory of group extensions, the proof of the former result is surprisingly simple. Most of this section is devoted to the study of critical pairs A, B for which equality is attained in Theorem 2.1.

A Brief Outline of the Proofs

Note that the assertions of both Theorems 2.1 and 2.6 are obvious if $p(G) = 2$. Thus in view of the Feit–Thompson theorem [33], it is enough to prove the assertions for solvable groups. Given that the results hold for cyclic groups of prime order, the natural approach is then to transfer the results to group extensions. In the case of Theorem 2.1 it is relatively simple, and only depends mildly on the structure of the extension, see Lemma 5.14. We prove this result in the next subsection. The proof of Theorem 2.6 is more delicate, in this case we cannot directly transfer the result to group extensions. In the third subsection we study how much the general approach of the second subsection can contribute towards the characterization of critical pairs if we also assume that the group H in Lemma 5.14 is a cyclic group of prime order, meaning that we can also take advantage of Vosper's inverse theorem. We complete the proof of Theorem 2.6 in the last subsection, where we finally take into account the specific structure of cyclic extensions. The proof also relies on Hamidoune's result Theorem 2.5.

Finally we note that the following alternative proof of Theorem 2.1 has been suggested by Hamidoune [46]. Let A and S denote nonempty finite subsets of an arbitrary group G. Denote by $\langle S \rangle$ the subgroup generated by S and by $\nu(S)$ the minimum order of an element in

S. According to a result of Hamidoune [43], if $A \cup AS \neq A\langle S\rangle$, then

$$|A \cup AS| \geq |A| + \min\{|S|, \nu(S)\}.$$

Now let A and B be arbitrary nonempty finite subsets of G satisfying $|A| + |B| - 1 \leq p(G)$. If $|B| = 1$, then obviously $|AB| = |A| + |B| - 1$. Otherwise, replacing A by Ab and B by $b^{-1}B$ for some element $b \in B$, we may assume that $1 \in B$. Let $S = B \setminus \{1\}$, then $\nu(S) \geq p(G)$ and $|\langle S\rangle| \geq p(G)$. Moreover, $A \cup AS = AB$. Thus either $AB = A\langle S\rangle$, in which case

$$|AB| \geq |\langle S\rangle| \geq p(G) \geq |A| + |B| - 1,$$

or the above theorem implies

$$|AB| = |A \cup AS| \geq |A| + \min\{|S|, \nu(S)\} = |A| + |S| = |A| + |B| - 1.$$

Even though this argument extends Theorem 2.1 to infinite groups, we feel that our direct approach is more transparent. We also depend on our proof in order to derive Theorem 2.6.

Proof of Theorem 2.1

For simplicity, we say that the group G possesses the Cauchy–Davenport property if for any pair of nonempty subsets A, B of G with $p(G) \geq |A| + |B| - 1$, we have $|AB| \geq |A| + |B| - 1$. In view of our previous remarks, Theorem 2.1 can be reduced to the following

Theorem 5.12. *Every finite solvable group G possesses the Cauchy–Davenport property.*

Let $G = G_0 \triangleright G_1 \triangleright \ldots \triangleright G_r = \{1\}$ be a composition series of G. Here every composition factor G_i/G_{i+1} is a cyclic group of prime order, and the length of the series $r = r(G)$, being equal to the total number of prime divisors of the order of G, does not depend on the particular choice of the composition series. If $G/N = H$ for some proper normal subgroup N of G, then $|G| = |N| \cdot |H|$ and thus $p(G) = \min\{p(N), p(H)\}$. We just remark that even if the group G is not finite, the inequality $p(G) \geq \min\{p(N), p(H)\}$ is not difficult to verify. Since every cyclic group of prime order has the Cauchy–Davenport property, Theorem 5.12 follows easily by induction on r from the following lemma.

Lemma 5.13. *Let G be an arbitrary group with a proper normal subgroup N. Assume that $p(G) = \min\{p(N), p(G/N)\}$. If both N and G/N possess the Cauchy–Davenport property, then so does G.*

Before we indicate how this lemma follows from a more general statement, we briefly recall the structure of general group extensions, following the terminology of [50]. Namely, if $H = G/N$, then the group G can be reconstructed from N and H as follows. There exist a map $f : H \times H \to N$ and for every $h \in H$ an automorphism $\vartheta_h \in \mathrm{Aut}(N)$ such that the following conditions hold for every $n \in N$ and $h_1, h_2, h_3 \in H$:

(i) $f(1,h_1) = f(h_1,1) = 1$;

(ii) $f(h_1,h_2)f(h_1h_2,h_3) = \vartheta_{h_1}(f(h_2,h_3))f(h_1,h_2h_3)$;

(iii) $\vartheta_{h_1}\vartheta_{h_2}(n) = f(h_1,h_2)\vartheta_{h_1h_2}(n)f(h_1,h_2)^{-1}$;

(iv) ϑ_1 is the unit element of $\mathrm{Aut}(N)$.

Then G is isomorphic to the group we obtain if we equip the set of ordered pairs $\{(n,h) \mid n \in N, h \in H\}$ with the multiplication

$$(n_1,h_1)(n_2,h_2) =: (n_1\vartheta_{h_1}(n_2)f(h_1,h_2), h_1h_2).$$

The behavior in the second coordinate is just like in the case of direct product, thus the properties of H can be exploited in a natural way. Note also that for every $h_1, h_2 \in H$, the mapping

$$n \to \vartheta_{h_1}(n)f(h_1,h_2)$$

is an $N \to N$ bijection. This is the key fact that allows us to exploit the properties of N, too. Now it is clear that Lemma 5.13 is a special case of the following statement.

Lemma 5.14. *Let N and H be arbitrary groups that possess the Cauchy–Davenport property. Assume that bijections $\varphi_{h_1,h_2}, \psi_{h_1,h_2} : N \to N$ are given for every $h_1, h_2 \in H$. Define on the set of ordered pairs $G = \{(n,h) \mid n \in N, h \in H\}$ a binary operation as follows:*

$$(n_1,h_1)(n_2,h_2) =: (\varphi_{h_1,h_2}(n_1)\psi_{h_1,h_2}(n_2), h_1h_2).$$

Then $|AB| \geq |A| + |B| - 1$ holds for arbitrary subsets A, B of G which satisfy

$$|A| + |B| - 1 \leq \min\{p(N), p(H)\}.$$

Proof. The assertion is obvious if one of the sets A and B is infinite. Thus we assume that A, B are finite subsets of G such that $|A|+|B|-1 \leq \min\{p(N), p(H)\}$. Write $k = |A|$, $\ell = |B|$ and let $A = C_1 \cup \ldots \cup C_s$ and $B = D_1 \cup \ldots \cup D_t$, where $C_i = \{(a_{ij}, c_i) \mid 1 \leq i \leq k_i\}$ and $D_i = \{(b_{ij}, d_i) \mid 1 \leq i \leq \ell_i\}$. We assume that $C = \{c_1, \ldots, c_s\}$ and $D = \{d_1, \ldots, d_t\}$ are subsets of H of cardinalities s and t, respectively. We will also assume that $k_1 \leq \cdots \leq k_s$ and $\ell_1 \leq \ldots \leq \ell_t$. Thus, $s \leq k$, $t \leq \ell$ and $\sum_{i=1}^{s} k_i = k$, $\sum_{i=1}^{t} \ell_i = \ell$. Introduce also $A_i = \{a_{ij} \mid 1 \leq j \leq k_i\}$ and $B_i = \{b_{ij} \mid 1 \leq j \leq \ell_i\}$, they are subsets of N. In C_iD_j, the second coordinate of each element is c_id_j, whereas the first coordinates form the set $\varphi_{c_i,d_j}(A_i)\psi_{c_i,d_j}(B_j)$. Since φ_{c_i,d_j} and ψ_{c_i,d_j} are $N \to N$ bijections and

$$k_i + \ell_j - 1 \leq k + \ell - 1 \leq \min\{p(N), p(H)\} \leq p(N),$$

our hypothesis on the group N implies that

$$|C_iD_j| = |\varphi_{c_i,d_j}(A_i)\psi_{c_i,d_j}(B_j)| \geq k_i + \ell_j - 1 \geq 1$$

holds for every $1 \leq i \leq s$ and $1 \leq j \leq t$. Due to the symmetry of the multiplication introduced on G, without any loss of generality we may assume that $s \geq t$. Consider the numbers $c_1d_t, c_2d_t, \ldots, c_sd_t \in H$, they are s different elements of the set product CD. Since $s + t - 1 \leq k + \ell - 1 \leq p(H)$, our hypothesis on the group H implies that $|CD| \geq s + t - 1$. Therefore there exists a set I of $t - 1$ pairs (γ, δ) such that the numbers

$$c_id_t \ (1 \leq i \leq s), \ c_\gamma d_\delta \ ((\gamma, \delta) \in I)$$

are all different. Since the sets

$$C_iD_t \ (1 \leq i \leq s), \ C_\gamma D_\delta \ ((\gamma, \delta) \in I)$$

are pairwise disjoint subsets of AB, it follows that

$$\begin{aligned}
|AB| &\geq \sum_{i=1}^{s} |C_iD_t| + \sum_{(\gamma,\delta)\in I} |C_\gamma D_\delta| && (5.1)\\
&\geq \sum_{i=1}^{s} (k_i + \ell_t - 1) + (t - 1) && (5.2)\\
&= k + t\ell_t + (s - t)\ell_t - s + t - 1 && (5.3)\\
&= k + t\ell_t + (s - t)(\ell_t - 1) - 1 && (5.4)\\
&\geq k + \ell - 1, && (5.5)
\end{aligned}$$

as it was to be proved. □

An intermediate step

Now we take a closer look at the proof of Lemma 5.14. For the rest of this subsection we assume that the finite sets A, B satisfy

$$|AB| = |A| + |B| - 1 \leq \min\{p(N), p(H)\} - 1.$$

Then we must have equality in (5.5) which means that $\ell_1 = \ell_2 = \ldots = \ell_t$ and also that either $s = t$ or $\ell_t = 1$ must hold. Note that we have assumed $s \geq t$. In the case $t \geq s$ a similar argument yields that $k_1 = k_2 = \ldots = k_s$ and, in addition, either $s = t$ or $k_s = 1$. Thus, if $s > t = 1$, then $\ell = \ell_1 = 1$, and similarly, if $t > s = 1$, then $k = 1$.

Assume now that $s, t \geq 2$. If H is a cyclic group of order p for some prime number p, then H clearly possesses the Cauchy–Davenport property. In (5.1) we also must have equality, which means that

$$|CD| = s + t - 1 \leq k + \ell - 1 \leq \min\{p(N), p(H)\} - 1 \leq p - 1.$$

Vosper's inverse theorem applied to H leaves us two possibilities, one being that $C = H \setminus hD^{-1}$ for some $h \in H$, but this only can occur if $s = k$, $\ell = t$ and $k + \ell = p \le p(N)$. The other possibility is that $C = \{c'_1, \dots, c'_s\}$ and $D = \{d'_1, \dots, d'_t\}$, where $c'_i = cq^{i-1}$ and $d'_i = dq^{i-1}$ for suitable elements $c, d, q \in H$. There is an index $1 \le \alpha \le s$ such that $c_s = c'_\alpha$. Clearly,

$$\begin{aligned} CD &= \{cd, cdq, cdq^2, \dots, cdq^{s+t-2}\} \\ &= \{c'_1d'_1, c'_2d'_1, \dots, c'_\alpha d'_1, c'_\alpha d'_2, \dots, c'_\alpha d'_t, c'_{\alpha+1}d'_t, \dots, c'_s d'_t\}. \end{aligned}$$

Writing $C'_i = C_j$, $k'_i = k_i$ if $c'_i = c_j$ and $D'_i = D_j$, $\ell'_i = \ell_j$ if $d'_i = d_j$, and noticing that the sets

$$C'_1D'_1, C'_2D'_1, \dots, C'_\alpha D'_1, C'_\alpha D'_2, \dots, C'_\alpha D'_t, C'_{\alpha+1}D'_t, \dots, C'_s D'_t$$

are pairwise disjoint subsets of G that satisfy

$$|C'_iD'_j| \ge k'_i + \ell'_i - 1 \ge k'_i,$$

we may argue that

$$\begin{aligned} |AB| &\ge \sum_{i=1}^{\alpha-1} |C'_iD'_1| + \sum_{i=1}^{t} |C'_\alpha D'_i| + \sum_{i=\alpha+1}^{s} |C'_iD'_t| \\ &\ge \sum_{i=1}^{t} (k_s + \ell_i - 1) + \sum_{i=1}^{s-1} k_i \\ &= \sum_{i=1}^{s} k_i + \sum_{i=1}^{t} \ell_i + (t-1)k_s - t \\ &= k + \ell - 1 + (t-1)(k_s - 1) \\ &\ge k + \ell - 1. \end{aligned}$$

From the conditions $|AB| = |A| + |B| - 1$ and $t \ge 2$ it follows that $k_s = 1$, that is, $s = k$. A similar argument also yields $t = \ell$.

We summarize these observations in the following lemma.

Lemma 5.15. *Let N be an arbitrary group that possesses the Cauchy–Davenport property, and let $H = \mathbb{Z}_p$ for some prime number p. Assume that bijections $\varphi_{h_1,h_2}, \psi_{h_1,h_2} : N \to N$ are given for every $h_1, h_2 \in H$. Define on the set of ordered pairs $G = \{(n, h) \mid n \in N, h \in H\}$ a binary operation as follows:*

$$(n_1, h_1)(n_2, h_2) =: (\varphi_{h_1,h_2}(n_1)\psi_{h_1,h_2}(n_2), h_1h_2).$$

If A, B are subsets of G which satisfy

$$|AB| = |A| + |B| - 1 \le \min\{p(N), p\} - 1,$$

then (using the notations introduced in the proof of Lemma 5.14) one of the following conditions holds:

(a) $k = 1$ *or* $\ell = 1$*;*

(b) $k, \ell \geq 2$ *and* $s = t = 1$*;*

(c) $s = k \geq 2$*,* $t = \ell \geq 2$ *and* C, D *are progressions in* H *of the same common quotient;*

(d) $s = k \geq 2$*,* $t = \ell \geq 2$*,* $k + \ell = p \leq p(N)$ *and* $C = H \setminus hD^{-1}$ *for a suitable element* $h \in H$*.*

Proof of Theorem 2.6

The 'if' part is quite simple. First, if $k = 1$ then $|AB| = |B| = \ell$, and if $\ell = 1$ then $|AB| = |A| = k$. Next, if the second condition holds, then again

$$|AB| = |\{aq^i b \mid 0 \leq i \leq k + \ell - 2\}| = k + \ell - 1,$$

because the order of q is at least $k + \ell$. Finally, in the third case we also have

$$|AB| = |uFv \setminus \{uzv\}| = |F| - 1 = k + \ell - 1.$$

To prove the necessity of the conditions, we may assume that the group G is solvable. We proceed by induction on the length of the composition series of G. If $r(G) = 1$, then G is a cyclic group of prime order and the result is contained in Vosper's theorem. So we assume that $r(G) \geq 2$ and the theorem has been already verified for every finite solvable group G' with $r(G') < r(G)$. Choose a normal subgroup $N \triangleleft G$ such that $H = G/N \cong \mathbb{Z}_p$ for a prime number p. Then G is a cyclic extension of N by H, and can be reconstructed from N and $H = \langle h \rangle$ as follows. There is an element $n_0 \in N$ and an automorphism $\vartheta \in \mathrm{Aut}(N)$ such that $\vartheta(n_0) = n_0$, $\vartheta^p(n) = n_0 n n_0^{-1}$ for every $n \in N$ and the multiplication on the set of ordered pairs

$$G_0 = \{(n, h^i) \mid n \in N, 0 \leq i \leq p - 1\}$$

introduced as

$$(n_1, h^i)(n_2, h^j) = (n_1 \vartheta^i(n_2) f(h^i, h^j), h^{i+j}),$$

where

$$f(h^i, h^j) = \begin{cases} 1 & \text{if } i + j < p \\ n_0 & \text{if } i + j \geq p \end{cases}$$

makes G_0 a group isomorphic to G, which we may as well identify with G. In particular, the function $f : H \times H \to N$ satisfies among others the relations

$$f(h^u, 1) = f(1, h^v) \tag{5.6}$$

and

$$\vartheta^i(f(h^u, h^v)) = f(h^u, h^v) \tag{5.7}$$

for every integer i and $0 \leq u, v \leq p - 1$.

According to Theorem 2.1, N possesses the Cauchy–Davenport property. We also have

$$|A| + |B| - 1 \le p(G) - 1 = \min\{p(N), p\} - 1.$$

Thus we may apply Lemma 5.15 with

$$\varphi_{h^i,h^j} \equiv \mathrm{id} \qquad \text{and} \qquad \psi_{h^i,h^j}(n) = \vartheta^i(n) f(h^i, h^j).$$

Accordingly, we distinguish between four cases.

(a) If $k = 1$ or $\ell = 1$, then condition (i) holds.

(b) If $k, \ell \ge 2$ and $s = t = 1$, then $|A_1| = k_1 = k$ and $|B_1| = \ell_1 = \ell$. Thus,

$$A = \{(a_i, h^\alpha) \mid 1 \le i \le k\} \qquad \text{and} \qquad B = \{(b_j, h^\beta) \mid 1 \le j \le \ell\}$$

with suitable integers $0 \le \alpha, \beta \le p - 1$. Therefore

$$AB = \{(a_i \vartheta^\alpha(b_j) f(h^\alpha, h^\beta), h^{\alpha+\beta}) \mid 1 \le i \le k, 1 \le j \le \ell\}.$$

Put $B_1' = \{\vartheta^\alpha(b_j) \mid 1 \le j \le \ell\}$. Then A_1, B_1' are subsets of N of cardinalities k and ℓ, respectively. Since every element of AB has the same second coordinate $h^{\alpha+\beta}$ and multiplication by $f(h^\alpha, h^\beta)$ is an $N \to N$ bijection, these sets satisfy

$$|A_1 B_1'| = |AB| = k + \ell - 1 \le p(N) - 1.$$

N is a finite solvable group with $r(N) = r(G) - 1$, thus our induction hypothesis implies that either (b1) there exist elements $a, b, q \in N$ such that $A_1 = \{a, aq, \ldots, aq^{k-1}\}$ and $B_1' = \{b, qb, \ldots, q^{\ell-1}b\}$, or (b2) $k + \ell - 1 = p(N) - 1 = p(G) - 1$ and there exist a subgroup F of N of order $p(N)$ and elements $u, v \in N$, $z \in F$ such that $A_1 \subset uF$, $B_1' \subset Fv$ and $A_1 = u(F \setminus zv(B_1')^{-1})$.

We elaborate on these two subcases separately.

(b1) We prove that in this case condition (ii) holds. More precisely, we prove that

$$A = \{a_0, a_0 q_0, \ldots, a_0 q_0^{k-1}\} \qquad \text{and} \qquad B = \{b_0, q_0 b_0, \ldots, q_0^{\ell-1} b_0\}, \tag{5.8}$$

where $a_0 = (a, h^\alpha)$, $b_0 = (\vartheta^{-\alpha}(b), h^\beta)$ and $q_0 = (\vartheta^{-\alpha}(q), 1)$.

We may assume that $a_{i+1} = aq^i$ and $b_{j+1} = \vartheta^{-\alpha}(q^j b)$ holds for $0 \le i \le k - 1$ and $0 \le j \le \ell - 1$. Thus $(a_1, h^\alpha) = a_0$ and $(b_1, h^\beta) = b_0$. We proceed by induction as follows. Assume first that we have already verified that $(a_i, h^\alpha) = a_0 q_0^{i-1}$ holds for some $1 \le i \le k-1$. Then

$$\begin{aligned} a_0 q_0^i &= (a_i, h^\alpha) q_0 = (aq^{i-1}, h^\alpha)(\vartheta^{-\alpha}(q), 1) \\ &= (aq^{i-1} \vartheta^\alpha(\vartheta^{-\alpha}(q)) f(h^\alpha, 1), h^\alpha) = (aq^i, h^\alpha) = (a_{i+1}, h^\alpha). \end{aligned}$$

On the other hand, if we have $(b_j, h^\beta) = q_0^{j-1} b_0$ for some $1 \leq j \leq \ell - 1$, then

$$\begin{aligned} q_0^j b_0 &= q_0(b_j, h^\beta) = (\vartheta^{-\alpha}(q), 1)(\vartheta^{-\alpha}(q^{j-1}b), h^\beta) \\ &= (\vartheta^{-\alpha}(q)\vartheta^0(\vartheta^{-\alpha}(q^{j-1}b))f(1, h^\beta), h^\beta) = (\vartheta^{-\alpha}(q^j b), h^\beta) = (b_{j+1}, h^\beta), \end{aligned}$$

since ϑ, and thus also $\vartheta^{-\alpha}$ is an automorphism of N. This verifies (5.8).

(b2) In this case we can write

$$A_1 = \{u\dot{a}_1, u\dot{a}_2, \ldots, u\dot{a}_k\} \qquad \text{and} \qquad B_1' = \{\dot{b}_1 v, \dot{b}_2 v, \ldots, \dot{b}_\ell v\},$$

where $a_i = u\dot{a}_i$, $\vartheta^\alpha(b_j) = \dot{b}_j v$ and

$$\{\dot{a}_1, \dot{a}_2, \ldots, \dot{a}_k\} = F \setminus z\{\dot{b}_1^{-1}, \dot{b}_2^{-1}, \ldots, \dot{b}_\ell^{-1}\}. \tag{5.9}$$

Let $F_0 = \{(\vartheta^{-\alpha}(f), 1) \mid f \in F\}$, then $|F_0| = |F| = p(N) = p(G)$, and clearly F_0 is a subgroup of G isomorphic to F. Introduce also $u_0 = (u, h^\alpha)$ and $v_0 = (\vartheta^{-\alpha}(v), h^\beta)$, and consider the sets $A_0, B_0 \subset F_0$ defined as follows:

$$A_0 = \{(\vartheta^{-\alpha}(\dot{a}_i), 1) \mid 1 \leq i \leq k\} \qquad \text{and} \qquad B_0 = \{(\vartheta^{-\alpha}(\dot{b}_j), 1) \mid 1 \leq j \leq k\}.$$

Then $A = u_0 A_0 \subset u_0 F_0$, because for any $1 \leq i \leq k$,

$$\begin{aligned} u_0(\vartheta^{-\alpha}(\dot{a}_i), 1) &= (u, h^\alpha)(\vartheta^{-\alpha}(\dot{a}_i), 1) \\ &= (u\vartheta^\alpha(\vartheta^{-\alpha}(\dot{a}_i))f(h^\alpha, 1), h^\alpha) = (u\dot{a}_i, h^\alpha) = (a_i, h^\alpha) \end{aligned}$$

holds. Similarly, for every $1 \leq j \leq \ell$ we have

$$\begin{aligned} (\vartheta^{-\alpha}(\dot{b}_j), 1)v_0 &= (\vartheta^{-\alpha}(\dot{b}_j), 1)(\vartheta^{-\alpha}(v), h^\beta) \\ &= (\vartheta^{-\alpha}(\dot{b}_j)\vartheta^0(\vartheta^{-\alpha}(v))f(1, h^\beta), h^\beta) \\ &= (\vartheta^{-\alpha}(\dot{b}_j v), h^\beta) = (b_j, h^\beta), \end{aligned}$$

implying that $B = B_0 v_0 \subset F_0 v_0$. Finally, applying $\vartheta^{-\alpha}$ to Equation (5.9) and observing that the map $\varphi : N \to G$ defined as $\varphi(x) = (x, 1)$ induces a group isomorphism from $\vartheta^{-\alpha}(F)$ onto F_0, we find that $A_0 = F_0 \setminus z_0 B_0^{-1}$, where $z_0 = (\vartheta^{-\alpha}(z), 1) \in F_0$. Consequently,

$$A = u_0 A_0 = u_0(F_0 \setminus z_0(Bv_0^{-1})^{-1}) = u_0(F_0 \setminus z_0 v_0 B^{-1}),$$

justifying that condition (iii) holds in this case.

(c) $s = k \geq 2$, $t = \ell \geq 2$ and C, D are progressions in H of the same common quotient. In this case we may write

$$A = \{(a_i, c_i) \mid 1 \leq i \leq k\} \qquad \text{and} \qquad B = \{(b_j, d_j) \mid 1 \leq j \leq \ell\},$$

where $c_i = h^{\alpha+(i-1)\gamma}$ and $d_j = h^{\beta+(j-1)\gamma}$ with suitable integers $0 \leq \alpha, \beta, \gamma \leq p-1$, $\gamma \neq 0$. Let $a_0 = (a_1, c_1) = (a_1, h^\alpha)$, $b_0 = (b_1, d_1) = (b_1, h^\beta)$ and $q_0 = (x, h^\gamma)$ where

$$x = \vartheta^{-\alpha}(a_1^{-1} a_2 (f(h^\alpha, h^\gamma))^{-1}).$$

This implies that

$$a_0 q_0 = (a_1, h^\alpha)(x, h^\gamma) = (a_1 \vartheta^\alpha(x) f(h^\alpha, h^\gamma), h^{\alpha+\gamma}) = (a_2, h^{\alpha+\gamma}) = (a_2, c_2).$$

We claim that in general,

$$(a_i, c_i) = a_0 q_0^{i-1} \qquad \text{and} \qquad (b_j, d_j) = q_0^{j-1} b_0$$

holds for every $1 \leq i \leq k$ and $1 \leq j \leq \ell$, indicating that condition (ii) is satisfied in this case.

Let $1 \leq i \leq k$, $1 \leq j \leq \ell$ and $m = i + j - 2$. Then

$$(a_i, c_i)(b_j, d_j) = (a_i \vartheta^{\alpha+(i-1)\gamma}(b_j) f(h^{\alpha+(i-1)\gamma}, h^{\beta+(j-1)\gamma}), h^{\alpha+\beta+m\gamma}).$$

Thus, for each $0 \leq m \leq k + \ell - 2$, there is an element x_m of AB whose second coordinate is $h^{\alpha+\beta+m\gamma}$. Moreover, the facts that p is a prime, $1 \leq \gamma \leq p-1$ and $k + \ell - 1 \leq p$ imply that the numbers $h^{\alpha+\beta+m\gamma}$ $(0 \leq m \leq k + \ell - 2)$ are $k + \ell - 1$ different elements of H, thus the element $x_m \in AB$ must be unique. It follows that

$$(a_i, c_i)(b_j, d_j) = (a_{i'}, c_{i'})(b_{j'}, d_{j'})$$

holds whenever $i + j = i' + j'$. We know that $(a_2, c_2) = (a_1, c_1)q_0$. For arbitrary $1 \leq j \leq \ell - 1$ we have

$$(a_2, c_2)(b_j, d_j) = (a_1, c_1)(b_{j+1}, d_{j+1}),$$

which then implies $q_0(b_j, d_j) = (b_{j+1}, d_{j+1})$. Thus, $(b_j, d_j) = q_0^{j-1} b_0$ follows by induction on j. In particular, $(b_2, d_2) = q_0(b_1, d_1)$. Thus the relation

$$(a_{i+1}, c_{i+1})(b_1, d_1) = (a_i, c_i)(b_2, d_2)$$

implies $(a_{i+1}, c_{i+1}) = (a_i, c_i)q_0$ for every $1 \leq i \leq k-1$, and we also obtain $(a_i, c_i) = a_0 q_0^{i-1}$ by induction on i.

(d) $s = k \geq 2$, $t = \ell \geq 2$, $k + \ell = p \leq p(N)$ and $C = H \setminus hD^{-1}$ for a suitable element $h \in H$. Let us note first, that we may assume $\ell \geq k$. This is because $A = u(F \setminus zvB^{-1})$ is equivalent to $B = (F \setminus A^{-1}uz)v$ and therefore, by reversing the multiplication on G (that is, introducing $a * b = ba$) we may exchange the roles of A and B while not changing the statement of Theorem 2.6. Once again, we may write

$$A = \{(a_i, c_i) \mid 1 \leq i \leq k\} \qquad \text{and} \qquad B = \{(b_j, d_j) \mid 1 \leq j \leq \ell\}.$$

Introduce $\dot{A} = (a_1, c_1)^{-1}A$ and $\dot{B} = B(b_1, d_1)^{-1}$, then we can write

$$\dot{A} = \{(\dot{a}_i, \dot{c}_i) \mid 1 \le i \le k\} \qquad \text{and} \qquad \dot{B} = \{(\dot{b}_j, \dot{d}_j) \mid 1 \le j \le \ell\},$$

where $(\dot{a}_1, \dot{c}_1) = (\dot{b}_1, \dot{d}_1) = (1, 1) \in \dot{A} \cap \dot{B}$, and writing $\dot{C} = \{\dot{c}_i \mid 1 \le i \le k\}$ and $\dot{D} = \{\dot{d}_j \mid 1 \le j \le \ell\}$, we have $|\dot{A}| = |\dot{C}| = k$, $|\dot{B}| = |\dot{D}| = \ell$, and $\dot{C} = H \setminus \dot{h}\dot{D}^{-1}$ holds with $\dot{h} = c_1^{-1}hd_1^{-1}$. In addition, clearly $|\dot{A}\dot{B}| = |AB| = |\dot{A}| + |\dot{B}| - 1$. We distinguish between two cases.

(d1) $G_0 = \langle \dot{B} \rangle \neq G$. Now we claim that $\dot{A} \subset G_0$. Indeed, if $a \in \dot{A} \setminus G_0$ then $(1, 1)\dot{B}$ and $a\dot{B}$ are disjoint subsets of $\dot{A}\dot{B}$, yielding

$$|\dot{A}\dot{B}| \ge 2|\dot{B}| = 2\ell > p > |\dot{A}| + |\dot{B}| - 1,$$

a contradiction. Note that G_0 is a proper subgroup of G, hence solvable with $r(G_0) < r(G)$ and $p(G_0) \ge p(G)$. Thus we may apply our induction hypothesis to conclude that either there exist $\dot{a}, \dot{b}, q_0 \in G_0$ such that

$$\dot{A} = \{\dot{a}, \dot{a}q_0, \dot{a}q_0^2, \ldots, \dot{a}q_0^{k-1}\} \qquad \text{and} \qquad \dot{B} = \{\dot{b}, q_0\dot{b}, q_0^2\dot{b}, \ldots, q_0^{\ell-1}\dot{b}\},$$

or $p(G_0) = p(G)$ and there exist a subgroup F of $G_0 < G$ of order $p(G)$ and elements $u, v \in G_0$, $z \in F$ such that

$$\dot{A} \subset uF, \dot{B} \subset Fv \qquad \text{and} \qquad \dot{A} = u(F \setminus zv\dot{B}^{-1}).$$

In the first case we have

$$A = \{a_0, a_0q_0, a_0q_0^2, \ldots, a_0q_0^{k-1}\} \qquad \text{and} \qquad B = \{b_0, q_0b_0, q_0^2b_0, \ldots, q_0^{\ell-1}b_0\}$$

with $a_0 = (a_1, c_1)\dot{a}$ and $b_0 = \dot{b}(b_1, d_1)$, and thus condition (ii) holds. In the other case, based on $(1, 1) \in \dot{A} \cap \dot{B}$, we may assume $u = v = 1$, and writing $u_0 = (a_1, c_1)$, $v_0 = (b_1, d_1)$ we may conclude that

$$A \subset u_0F, B \subset Fv_0 \qquad \text{and} \qquad A = u_0(F \setminus zv_0B^{-1}),$$

implying condition (iii).

(d2) $G_0 = \langle \dot{B} \rangle = G$. In this case we show that $\dot{B}$ is a Cauchy-subset of G. To see that, let H_0 be any subgroup of G. If $H_0 = G$, then clearly

$$\min\{|\dot{B}H_0|, |H_0\dot{B}|\} = |G| \ge \min\{|G|, |H_0| + |\dot{B}| - 1\}.$$

If $H_0 = \{(1, 1)\}$, then

$$\min\{|\dot{B}H_0|, |H_0\dot{B}|\} = |\dot{B}| = \min\{|G|, |H_0| + |\dot{B}| - 1\}.$$

Otherwise $\dot{B} \not\subseteq H_0$, $|H_0| \ge p(G) > |\dot{B}|$, and thus

$$\min\{|\dot{B}H_0|, |H_0\dot{B}|\} \ge 2|H_0| \ge |H_0| + |\dot{B}| - 1 = \min\{|G|, |H_0| + |\dot{B}| - 1\}.$$

Therefore we can apply Theorem 2.5. Since $|\dot{A}| \neq 1$ and $|\dot{A}| + |\dot{B}| < |G|$, it follows that there are elements $a, b, q \in G$ and a natural number l such that

$$\dot{A} = \{a, aq, aq^2, \ldots, aq^{k-1}\} \quad \text{and} \quad \dot{B} = (G \setminus \langle q \rangle b) \cup \{b, qb, q^2b, \ldots q^{l-1}b\}.$$

Were $\langle q \rangle \neq G$, we would have $|G| \geq p(G)|\langle q \rangle b|$, and thus it would follow that

$$\ell = |\dot{B}| \geq \frac{p(G)-1}{p(G)}|G| \geq \frac{p(G)-1}{p(G)}(p(G))^2 \geq p(G) > \ell,$$

a contradiction. Consequently, $\langle q \rangle = G$, $l = \ell$,

$$\dot{A} = \{a, aq, aq^2, \ldots, aq^{k-1}\} \quad \text{and} \quad \dot{B} = \{b, qb, q^2b, \ldots q^{\ell-1}b\},$$

and with the notation $a_0 = (a_1, c_1)a$, $b_0 = b(b_1, d_1)$ we see that

$$A = \{a_0, a_0q, a_0q^2, \ldots, a_0q^{k-1}\} \quad \text{and} \quad B = \{b_0, qb_0, q^2b_0, \ldots q^{\ell-1}b_0\},$$

implying that condition (ii) must hold.

This concludes the induction step, and the proof of Theorem 2.6 is complete.

Chapter 6

Elementary Methods

In the first section of the present chapter we verify a conjecture of Lev in a very strong sense. Here we use only elementary arguments. In the second section a stronger conjecture of Lev is proved along with a conjecture of Alon. The starting point of those proofs is Theorem 2.10, which is due to Dias da Silva and Hamidoune.

6.1 Balanced Subset Sums in Dense Sets of Integers

In this section we prove Theorems 2.20 and 2.21. The latter can be easily derived from the following result.

Theorem 6.1. *For every $\varepsilon > 0$ there is an integer $n_0 = n_0(\varepsilon)$ with the following property. If $n \geq n_0$, $1 \leq a_1 < a_2 < \ldots < a_n \leq 2n-2$ are integers, and N is an integer such that $|N| \leq (\frac{9}{100} - \varepsilon)n^2$, then there exist $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, +1\}$ such that $|\varepsilon_1 + \ldots + \varepsilon_n| \leq 1$ and $|\varepsilon_1 a_1 + \ldots + \varepsilon_n a_n - N| \leq 1$.*

Indeed, choose $\varepsilon = 9/100 - 1/12$ in the above theorem. If $k = \sigma/2 + x$ is an integer in the prescribed interval, then for the integer $N = 2x$ there exist $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, +1\}$ such that $|\varepsilon_1 + \ldots + \varepsilon_n| \leq 1$ and $|\varepsilon_1 a_1 + \ldots + \varepsilon_n a_n - N| \leq 1$. Since $N = 2x \equiv \sigma \equiv \varepsilon_1 a_1 + \ldots + \varepsilon_n a_n \pmod 2$, it follows that $\varepsilon_1 a_1 + \ldots + \varepsilon_n a_n = N$, and with $I = \{i \mid \varepsilon_i = +1\}$ we have $|I| \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$ and

$$\sum_{i \in I} a_i = \frac{1}{2}\Big(\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} \varepsilon_i a_i\Big) = \frac{\sigma}{2} + x = k.$$

Thus Theorem 2.21 follows. □

Now the first conjecture of Lev we mentioned on Page 18, assumed that $n \geq 89$, follows immediately in a similar way from the Theorem 2.20, unless $a_i = 2i - 1$ for $1 \leq i \leq n$. Even in that case, it is easy to check that the statement of Theorem 2.20 remains valid if $n \equiv 0,\ 1$ or $3 \pmod 4$. This is not the case, however, if $n \equiv 2 \pmod 4$.

Indeed, let $n = 4k + 2$ and suppose that $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, +1\}$ such that $|\varepsilon_1 + \ldots + \varepsilon_n| \leq 1$. Consider $I = \{1 \leq i \leq n \mid \varepsilon_i = +1\}$, then $|I| = 2k + 1$. Therefore $A = \sum_{i \in I} a_i$ and $B = \sum_{i \notin I} a_i$ are odd numbers. However, $A + B = \sum_{i=1}^{n} a_i = (4k+2)^2$ is divisible by 4, hence $A - B \equiv 2 \pmod 4$, and $|\varepsilon_1 a_1 + \ldots + \varepsilon_n a_n| = |A - B| \geq 2$. Nevertheless, choosing

$$I = \{1, 2, 3, 5\} \cup \bigcup_{i=2}^{k} \{4i, 4i + 1\} \subseteq \{1, 2, \ldots, n\}$$

we find that

$$\sum_{i \in I} a_i = \frac{1}{2} \sum_{i=1}^{n} a_i \ ,$$

confirming the conjecture of Lev in this remaining case, too.

The First Conjecture of Lev: Proof of Theorem 2.20

Before turning to the proof we note that although most likely the condition $n \geq 89$ can essentially be relaxed, it is not merely technical. The sequence $(1, 2, 3, 8, 9, 10, 14, 15)$ demonstrates that Theorem 2.20 is not valid with $n_0 = 8$. An other formulation of the condition in the theorem is the requirement that there exists an index $1 \leq \nu \leq n$ such that a_ν is even. Finally, if $a_n \leq 2n - 2$, then the condition is automatically fulfilled.

Turning to the proof, first we note that it is enough to prove Theorem 2.20 when n is an even number. Indeed, let n be odd, and assume that the statement has been proved for $n + 1$. Consider the sequence

$$b_1 = 1 < b_2 = a_1 + 1 < \ldots < b_{n+1} = a_n + 1 < 2(n + 1) - 1.$$

There exist $\eta_1, \ldots, \eta_{n+1} \in \{-1, +1\}$ such that,

$$|\eta_1 + \ldots + \eta_{n+1}| \leq 1 \quad \text{and} \quad |\eta_1 b_1 + \ldots + \eta_{n+1} b_{n+1}| \leq 1.$$

Since $n + 1$ is even, it follows that $\eta_1 + \ldots + \eta_{n+1} = 0$. Let $\varepsilon_i = \eta_{i+1}$, then $|\varepsilon_1 + \ldots + \varepsilon_n| = |-\eta_1| = 1$, and

$$\Big|\sum_{i=1}^{n} \varepsilon_i a_i\Big| = \Big|\sum_{i=1}^{n} \eta_{i+1} a_i + \sum_{i=1}^{n+1} \eta_i\Big| = \Big|\sum_{i=1}^{n+1} \eta_i b_i\Big| \leq 1.$$

Accordingly, we assume that $n = 2m$ with an integer $m \geq 45$. To illustrate the initial idea of the proof, consider the differences $e_i = a_{2i} - a_{2i-1}$ for $i = 1, 2, \ldots, m$. If we found $\delta_1, \ldots, \delta_m \in \{-1, +1\}$ such that $|\sum_{i=1}^{m} \delta_i e_i| < 2$, then the choice $\varepsilon_{2i} = \delta_i$, $\varepsilon_{2i-1} = -\delta_i$ would clearly give the desired result. This is the case, in fact, when $\sum_{i=1}^{m} e_i \leq 2m - 2$, as it can be easily derived from the following two simple lemmas.

Lemma 6.2. *Let $e_1, \ldots, e_k \geq 1$ and suppose that*

$$E = \sum_{i=1}^{k} e_i \leq \beta k - (\beta^2 - \beta)$$

for some positive real number β. Then

$$\sum_{e_i < s+1} e_i \geq s$$

holds for every positive integer $\beta - 1 \leq s \leq k - \beta$.

Proof. If s is a positive integer, then obviously

$$\sum_{e_i < s+1} e_i \geq \sum_{e_i < s+1} 1 = k - \sum_{e_i \geq s+1} 1 \geq k - \frac{E}{s+1} \ .$$

As long as

$$(k-1)^2 - 4(E-k) \geq (k-\alpha)^2 \ , \tag{6.1}$$

we have

$$k - \frac{E}{s+1} \geq s$$

for every $(\alpha - 1)/2 \leq s \leq k - (\alpha + 1)/2$. To complete the proof we only have to notice that (6.1) is satisfied if $\alpha = 2\beta - 1$. □

Lemma 6.3. *Let $e_1, \ldots, e_k \geq 1$ and suppose that*

$$\sum_{e_i < s+1} e_i \geq s \tag{6.2}$$

holds for every integer $1 \leq s \leq \max\{e_i \mid 1 \leq i \leq k\}$. Let F be any number such that

$$|F| < \sum_{i=1}^{k} e_i + 2 \ . \tag{6.3}$$

Then there exist $\varepsilon_1, \ldots, \varepsilon_k \in \{-1, +1\}$ such that

$$\Big| \sum_{i=1}^{k} \varepsilon_i e_i - F \Big| < 2 \ ,$$

in particular $F = \sum_{i=1}^{k} \varepsilon_i e_i$ if the e_i's are integers and $F \equiv \sum_{i=1}^{k} e_i \pmod 2$.

Proof. Without loss of generality, we may suppose that that $e_1 \geq e_2 \geq \ldots \geq e_k$, then $e_k < 2$. The point is, that the condition allows us to construct $\varepsilon_1, \ldots, \varepsilon_k$ sequentially so that the sequence of partial sums $\sum_{j=1}^{i} \varepsilon_j e_j$ oscillates about F with smaller and smaller amplitude, until it eventually approximates F with the desired accuracy.

More precisely, let $\Delta_0 = F$, and define ε_n and Δ_n recursively as follows. Let, for $n = 1, 2, \ldots, k$,

$$\varepsilon_n = \begin{cases} 1 & \text{if } \Delta_{n-1} \geq 0 \\ -1 & \text{if } \Delta_{n-1} < 0 \end{cases}$$

and let $\Delta_n = \Delta_{n-1} - \varepsilon_n e_n$, then

$$\Delta_n = F - \varepsilon_1 e_1 - \varepsilon_2 e_2 - \ldots - \varepsilon_n e_n$$

for every $0 \leq n \leq k$. We prove, by induction, that

$$|\Delta_n| < e_{n+1} + \ldots + e_{k-1} + e_k + 2 \tag{6.4}$$

for $n = 0, 1, \ldots, k$.

This is true for $n = 0$. Thus, let $1 \leq n \leq k$, and suppose that (6.4) is satisfied with $n-1$ in place of n. Assume, w.l.o.g, that $\Delta_{n-1} \geq 0$. Then, by definition,

$$-e_n \leq \Delta_n = \Delta_{n-1} + (-1)e_n < e_{n+1} + \ldots + e_k + 2 \; .$$

Thus, to verify (6.4), it suffices to show that $e_n < e_{n+1} + \ldots + e_k + 2$. This is definitely true, if $e_{n+1} = e_n$ or $n = k$. Otherwise we can write

$$\sum_{i=n+1}^{k} e_i = \sum_{e_i < e_n} e_i \geq \sum_{e_i < \lfloor e_n \rfloor} e_i \geq \lfloor e_n \rfloor - 1 > e_n - 2 \; ,$$

proving the assertion. Letting $n = k$ in (6.4), the statement of the lemma follows. □

The main idea of the proof of Theorem 2.20 is to find a partition

$$\{a_1, a_2, \ldots, a_n\} = \bigcup_{i=1}^{k} \{x_i, y_i\} \cup \{z_1, \ldots, z_{n-2k}\} \tag{6.5}$$

such that $e_i = x_i - y_i$ $(1 \leq i \leq k)$ and $F = \sum_{i=1}^{n-2k} (-1)^i z_i$ satisfy the conditions of Lemma 6.3. Then Theorem 2.20 follows immediately.

To achieve this we will construct the above partition so that

$$\sum_{i=1}^{k} e_i \leq 4k - 12 \quad (\text{or} \quad \sum_{i=1}^{k} e_i \leq 3k - 6), \tag{6.6}$$

$$e_i \leq k - 4 \quad (\text{or} \quad e_i \leq k - 3) \quad \text{for} \quad i = 1, 2, \ldots, k \; , \tag{6.7}$$

$$|F| \leq k + 1 \; , \quad \text{and} \tag{6.8}$$

$$\sum_{e_i \leq s} e_i \geq s \quad \text{if} \quad s = 1 \quad \text{or} \quad s = 2 \; . \tag{6.9}$$

Then an application of Lemma 6.2 with $\beta = 4$ (or with $\beta = 3$) will show that e_i $(1 \leq i \leq k)$ and F satisfy the conditions of Lemma 6.3. More precisely, it follows from (6.6) and (6.9) that

condition (6.2) holds for $s \leq k-\beta$, hence for every integer $1 \leq s \leq \max\{e_i \mid 1 \leq i \leq k\}$ in view of (6.7). Finally, (6.3) follows from (6.8), given that $\sum_{i=1}^{k} e_i \geq k$. Therefore, once we found a partition (6.5) with properties (6.6)–(6.9), the proof of Theorem 2.20 will be complete.

First we take care of the condition (6.9). If we take $x_k = a_{\nu+1}$ and $y_k = a_\nu$, then $e_k = 1$. Moreover, since

$$\sum_{i=1}^{n-1}(a_{i+1} - a_i) \leq 2n - 2,$$

there must be an index $\mu \notin \{\nu-1, \nu, \nu+1, n\}$, such that $a_{\mu+1} - a_\mu \leq 2$. Taking $x_{k-1} = a_{\mu+1}$ and $y_{k-1} = a_\mu$, condition (6.9) will be satisfied. Enumerating the remaining $n-4$ elements of the sequence (a_i) as

$$1 \leq b_1 < b_2 < \ldots < b_{2m-4} \leq 4m - 1,$$

with $f_i = b_{2i} - b_{2i-1}$ we find that

$$\sum_{i=1}^{m-2} f_i = \sum_{i=1}^{m-2}(b_{2i} - b_{2i-1}) \leq (4m-2) - (m-3) = 3m + 1. \tag{6.10}$$

Since $m > 21$, there cannot be 3 different indices i with $f_i \geq m - 5$. We distinguish between three cases.

Case 1) If $f_i \leq m - 6$ for $1 \leq i \leq m-2$, then we can choose $k = m$, $F = 0$. Taking $x_i = b_{2i}$ and $y_i = b_{2i-1}$ for $1 \leq i \leq k-2$, conditions (6.7) and (6.8) are obviously satisfied, whereas (6.6) follows easily form (6.10):

$$\sum_{i=1}^{k} e_i \leq \sum_{i=1}^{m-2} f_i + 3 \leq 3m + 4 \leq 4m - 12,$$

given that $m \geq 16$.

Case 2) There exist indices u, v such that $m - 5 \leq f_u \leq f_v$. In view of (6.10) we have $f_u + f_v \leq (3m+1) - (m-4) = 2m + 5$, and consequently $m - 5 \leq f_u \leq f_v \leq m + 10$ and $0 \leq f_v - f_u \leq 15$. Therefore we may choose $k = m - 2$, $z_1 = b_{2v-1}$, $z_2 = b_{2v}$, $z_3 = b_{2u}$, $z_4 = b_{2u-1}$. Constructing x_i, y_i $(1 \leq i \leq m-4)$ from the remaining elements of the sequence (b_i) in the obvious way we find that $|F| \leq 15 < m - 2 = k$, each e_i satisfies $e_i \leq m - 6 = k - 4$, and once again (6.10) gives

$$\sum_{i=1}^{k} e_i \leq \sum_{i=1}^{m-2} f_i - 2(m-5) + 3 \leq m + 14 < 4m - 20 = 4k - 12.$$

Case 3) There exists exactly one index u with $m - 5 \leq f_u$. From (6.10) it follows that $f_u \leq (3m+1) - (m-3) = 2m + 4$. We claim that there exist indices v, w different from u such that

$$|b_{2w} + b_{2w-1} - b_{2v} - b_{2v-1} - f_u| \leq m - 2. \tag{6.11}$$

In that case we can choose $k = m - 3$ and $z_1 = b_{2u}$, $z_2 = b_{2u-1}$, $z_3 = b_{2v}$, $z_4 = b_{2w}$, $z_5 = b_{2w-1}$, $z_6 = b_{2u-1}$ to have $|F| \leq m - 2 = k + 1$. Constructing x_i, y_i $(1 \leq i \leq m - 4)$ from the remaining elements of the sequence (b_i) in the obvious way this time we find that each e_i satisfies $e_i \leq m - 6 = k - 3$, and

$$\sum_{i=1}^{k} e_i \leq \sum_{i=1}^{m-2} f_i - (m-5) - 2 + 3 \leq 2m + 7 \leq 3m - 15 = 3k - 6.$$

It only remains to prove the above claim. The idea is to find v, w such a way that f_v, f_w are small and at the same time $b_{2w} - b_{2v}$ lies in a prescribed interval that depends on the size of f_u. It turns out that the optimum strategy for such an approach is the following. First, for any positive integer $\kappa \geq 2$, introduce

$$I_\kappa = \{i \mid 1 \leq i \leq m-2,\ i \neq u,\ f_i \leq \kappa\}.$$

Denote by x the number of indices $i \neq u$ for which $f_i > \kappa$. Then

$$(m - 3 - x) + (\kappa + 1)x \leq \sum_{i=1}^{m-2} f_i - f_u \leq (3m + 1) - (m - 5) = 2m + 6.$$

Thus, $\kappa x \leq m + 9$, and $m - 3 - x \geq (1 - 1/\kappa)m - 3 - 9/\kappa$. We have proved

Claim 6.4. $|I_\kappa| \geq \dfrac{\kappa - 1}{\kappa} m - \dfrac{9}{\kappa} - 3$. *In particular* $t = |I_7| \geq \dfrac{6m - 30}{7}$.

Write $c_0 = 0$ and let

$$\bigcup_{i \in I_7} \{b_{2i-1}, b_{2i}\} = \{c_1 < c_2 < \ldots < c_{2t-1} < c_{2t}\}.$$

Now we separate two subcases as follows.

Case 3A) $m - 5 \leq f_u \leq 2m - 14$. We will prove that there exist $1 \leq i < j \leq t$ such that

$$\frac{m}{2} - 3 \leq \Delta_{i,j} = c_{2j} - c_{2i} \leq m - 7. \tag{6.12}$$

Since we have

$$1 \leq c_{2i} - c_{2i-1}, c_{2j} - c_{2j-1} \leq 7, \tag{6.13}$$

we can argue that

$$m - 12 \leq 2\Delta_{i,j} - 6 \leq c_{2j} + c_{2j-1} - c_{2i} - c_{2i-1} \leq 2\Delta_{i,j} + 6 < 2m - 7,$$

and that implies (6.11). If there exists $1 \leq i \leq t - 1$ such that

$$\frac{m}{2} - 3 \leq c_{2i+2} - c_{2i} \leq m - 7,$$

then (6.12) is immediate. Otherwise we have

$$c_{2i+2} - c_{2i} \le \frac{m}{2} - \frac{7}{2} \quad \text{or} \quad c_{2i+2} - c_{2i} \ge m - 6$$

for every integer $1 \le i \le t-1$. This way we distinguish between 'small gaps' and 'large gaps' in the sequence $c_2, c_4, \ldots, c_{2t}$. The large gaps partition this sequence into 'blocks', where the gap between two consecutive elements within a block is always small. For such a block $B = (c_{2i}, c_{2i+2}, \ldots, c_{2i'})$, the quantity $\ell(B) = 2(i' - i)$ we call the length of the block. Since

$$2 \cdot \Big(\frac{m}{2} - \frac{7}{2}\Big) < m - 6,$$

in order to have a pair i, j with (6.12), it is enough to prove that at least one block has a length $\ge m/2 - 3$. Then the smallest integer j satisfying $c_{2j} - c_{2i} \ge m/2 - 3$ will do the job.

We claim that there cannot be more than 3 blocks. Indeed, since every gap is at least 2, were there 3 or more large gaps, we would find that

$$\begin{aligned} 4m - 1 &\ge \sum_{i=0}^{t-1} (c_{2i+2} - c_{2i}) \ge 3(m-6) + (t-3)2 \\ &\ge 3m - 18 + 2\Big(\frac{6m-30}{7} - 3\Big), \end{aligned}$$

implying $m \le 221/5 < 45$, a contradiction.

Since there are at most 3 blocks, one must contain at least $t/3$ different c_{2i}'s, and thus its length

$$\ell(B) \ge 2\Big(\frac{t}{3} - 1\Big) \ge \frac{4m-20}{7} - 2.$$

Given that $m \ge 26$ we conclude that indeed $\ell(B) \ge m/2 - 3$.

Case 3B) $2m - 13 \le f_u \le 2m + 4$. This time we prove that

$$\frac{m}{2} + 6 \le \Delta_{i,j} \le \frac{3}{2}m - \frac{21}{2} \tag{6.14}$$

holds with suitable $1 \le i < j \le t$. In view of (6.13) this implies

$$m + 6 \le 2\Delta_{i,j} - 6 \le c_{2j} + c_{2j-1} - c_{2i} - c_{2i-1} \le 2\Delta_{i,j} + 6 \le 3m - 15,$$

and from that (6.11) follows. Similarly to the previous case, we may assume that there are only small and large gaps, which in this case means that

$$c_{2i+2} - c_{2i} \le \frac{m}{2} + \frac{11}{2} \quad \text{or} \quad c_{2i+2} - c_{2i} \ge \frac{3}{2}m - 10$$

holds for every integer $1 \le i \le t-1$. Given that (here we use $m \ge 44$)

$$2 \cdot \Big(\frac{m}{2} + \frac{11}{2}\Big) < \frac{3}{2}m - 10,$$

it suffices to prove that there is a block B with $\ell(B) \ge m/2 + 6$.

Were there 2 or more large gaps, we would find that

$$\begin{aligned} 4m-1 &\geq \sum_{i=0}^{t-1}(c_{2i+2}-c_{2i}) \geq 2\Big(\frac{3}{2}m-10\Big)+(t-2)2 \\ &\geq 3m-20+2\Big(\frac{6m-30}{7}-2\Big), \end{aligned}$$

implying $m \leq 221/5 < 45$, a contradiction. Therefore there are at most 2 blocks, one of which containing at least $t/2$ different c_{2i}'s. The length of that block thus satisfies

$$\ell(B) \geq 2\Big(\frac{t}{2}-1\Big) \geq \frac{6m-30}{7}-2.$$

Since $m \geq 172/5$, we find that $\ell(B) \geq m/2+6$, and the proof is complete.

An Extension: Proof of Theorem 6.1

Obviously we may assume that $\varepsilon > 0$ is small enough so that all the below arguments work. We fix such an ε and assume that n is large enough. As in the proof of Theorem 2.20, we may assume that $n = 2m$ is an even number. Put $c = 1/5 - 2\varepsilon$. We will prove that there exists an integer $k \geq (1-c)m-7$ and a partition in the form (6.5) such that for $e_i = x_i - y_i$ $(1 \leq i \leq k)$ and $F = N + \sum_{i=1}^{n-2k}(-1)^i z_i$ the following conditions hold:

$$\sum_{i=1}^{k} e_i \leq 4k-12, \tag{6.15}$$

$$e_i \leq (1-c)m-11 \leq k-4 \quad \text{for} \quad i = 1,2,\ldots,k\ , \tag{6.16}$$

$$|F| \leq (1-c)m-6 \leq k+1\ , \quad \text{and} \tag{6.17}$$

$$\sum_{e_i \leq s} e_i \geq s \quad \text{if} \quad s=1 \quad \text{or} \quad s=2\ . \tag{6.18}$$

As in the proof of Theorem 2.20, we can apply Lemma 6.2 with $\beta = 4$, and then Lemma 6.3 gives the result.

Clearly there exist $1 \leq \mu, \nu \leq n-1$, $\mu \notin \{\nu-1, \nu, \nu+1\}$ such that $a_{\nu+1} - a_\nu = 1$ and $a_{\mu+1} - a_\mu \leq 2$. Putting $x_1 = a_{\nu+1}$, $y_1 = a_\nu$, $x_2 = a_{\mu+1}$, $y_2 = a_\mu$ then takes care of (6.18). Enumerate the remaining $n-4$ elements of the sequence (a_i) as

$$1 \leq b_1 < b_2 < \ldots < b_{2m-4} \leq 4m-2.$$

Take $q = \lceil cm \rceil$. Since

$$\begin{aligned} \sum_{i=1}^{q}(b_{2m-3-i}-b_i) &\geq \sum_{i=1}^{q}(2m-2i-3) = 2qm-q(q+4) \\ &> 2cm^2-(cm+1)(cm+5) = (2c-c^2)m^2-(6cm+5) \\ &> \Big(\frac{9}{25}-\frac{16}{5}\varepsilon-4\varepsilon^2\Big)m^2-2m > \Big(\frac{9}{25}-4\varepsilon\Big)m^2 \geq |N| \end{aligned}$$

and $b_{2m-3-i} - b_i \le 4m-3$ for every i, there exists an integer $0 \le r < cm+1$ such that

$$\Big|N - \mathrm{sgn}(N)\sum_{i=1}^{r}(b_{2m-3-i} - b_i)\Big| \le 2m-2,$$

where $\mathrm{sgn}(N) = +1$, if $N \ge 0$ and $\mathrm{sgn}(N) = -1$, if $N < 0$. Consider

$$r+1 \le b_{r+1} < b_{r+2} < \ldots < b_{2m-4-r} \le 4m-2-r,$$

and let $f_i = b_{r+2i} - b_{r+2i-1}$ for $1 \le i \le m-2-r$, then

$$\sum_{i=1}^{m-r-2} f_i \le \big((4m-2-r) - (r+1)\big) - (m-r-3) \le 3m. \tag{6.19}$$

Were there 3 or more indices i with $f_i > (1-c)m - 11$, it would imply

$$\sum_{i=1}^{m-r-2} f_i > 3\big((1-c)m - 11\big) + (m-r-5) > (4-4c)m - 39 > 3m,$$

a contradiction if m is large enough. Thus there exist an integer $s \in \{0,1,2\}$ and indices $i_1, \ldots, i_s$ such that $f_i > (1-c)m - 11$ if and only if $i \in \{i_1, \ldots, i_s\}$. Moreover, if $s \ge 1$, then for each $j \in \{1, \ldots, s\}$ we have

$$f_{i_j} \le 3m - (m-r-3) < (2+c)m + 4.$$

Consequently, there exist $\delta_1, \ldots, \delta_s \in \{-1, +1\}$ such that

$$\Big|N - \mathrm{sgn}(N)\sum_{i=1}^{r}(b_{2m-3-i} - b_i) - \sum_{j=1}^{s}\delta_j f_{i_j}\Big| < (2+c)m + 4. \tag{6.20}$$

Put $\kappa = \lceil 3/\varepsilon \rceil \le (1-c)m - 11$ and introduce

$$I_\kappa = \{i \mid 1 \le i \le m-r-2,\ f_i \le \kappa\}.$$

Denoting by x the number of indices i with $f_i > \kappa$ we have

$$(m-r-2-x) + (\kappa+1)x \le \sum_{i=1}^{m-r-2} f_i \le 3m,$$

implying $\kappa x < (2+c)m + 3$, and thus

$$t = |I_\kappa| = m-r-2-x > \Big(1 - c - \frac{2+c}{\kappa}\Big)m - 3 - \frac{3}{\kappa} > \Big(\frac{4}{5} + \varepsilon\Big)m.$$

Write $c_0 = 0$ and let

$$\bigcup_{i \in I_\kappa} \{b_{r+2i-1}, b_{r+2i}\} = \{c_1 < c_2 < \ldots < c_{2t-1} < c_{2t}\}.$$

We prove that there exist $1 \le i_1 < j_1 \le t$ such that

$$\frac{2}{5}m \le \Delta_1 = c_{2j_1} - c_{2i_1} \le \frac{4}{5}m. \tag{6.21}$$

This is immediate if there exists $1 \le i \le t-1$ such that

$$\frac{2}{5}m \le c_{2i+2} - c_{2i} \le \frac{4}{5}m,$$

otherwise we have

$$c_{2i+2} - c_{2i} < \frac{2}{5}m \quad \text{or} \quad c_{2i+2} - c_{2i} > \frac{4}{5}m$$

for every integer $1 \le i \le t-1$. Gaps in the sequence $c_2, c_4, \ldots, c_{2t}$, which are larger than $4m/5$, partition this sequence into blocks, where the gap between two consecutive elements within a block is always smaller than $2m/5$. We claim that there cannot be more than 3 such blocks. Were there on the contrary at least 3 large gaps, we would find that

$$4m - 2 \ge \sum_{i=0}^{t-1}(c_{2i+2} - c_{2i}) > 3 \cdot \frac{4}{5}m + (t-3)\cdot 2 > (4+2\varepsilon)m - 6,$$

a contradiction. Now one of the blocks must contain at least $t/3$ different c_{2i}'s, and thus its length satisfies

$$\ell(B) \ge 2\Big(\frac{t}{3} - 1\Big) > \frac{2}{5}m.$$

Consequently, (6.21) holds with suitable elements c_{2i_1}, c_{2j_1} of B. Removing i_1, j_1 from I_κ and repeating the argument we find $1 \le i_2 < j_2 \le t$ such that $\{i_2, j_2\} \cap \{i_1, j_1\} = \emptyset$ and $2m/5 \le \Delta_2 = c_{2j_2} - c_{2i_2} \le 4m/5$. Since for $\alpha = 1, 2$ we have

$$1 \le c_{2i_\alpha} - c_{2i_\alpha - 1}, c_{2j_\alpha} - c_{2j_\alpha - 1} \le \kappa, \tag{6.22}$$

we can argue that

$$2\Delta_\alpha - \kappa + 1 \le \Gamma_\alpha = c_{2j_\alpha} + c_{2j_\alpha - 1} - c_{2i_\alpha} - c_{2i_\alpha - 1} \le 2\Delta_\alpha + \kappa - 1,$$

that is,

$$\frac{4}{5}m - \frac{3}{\varepsilon} < \Gamma_\alpha < \frac{8}{5}m + \frac{3}{\varepsilon}. \tag{6.23}$$

In view of (6.20) and (6.23), there exist an integer $p \in \{0, 1, 2\}$ and $\eta_1, \ldots, \eta_p \in \{-1, +1\}$ such that

$$\Big|N - \operatorname{sgn}(N)\sum_{i=1}^{r}(b_{2m-3-i} - b_i) - \sum_{j=1}^{s}\delta_j f_{i_j} - \sum_{\alpha=1}^{p}\eta_\alpha \Gamma_\alpha\Big| < \frac{4}{5}m + \frac{3}{2\varepsilon} \le (1-c)m - 6.$$

Consequently, we can choose $k = m - r - s - 2p > (1-c)m - 7$, and the elements of the set

$$\bigcup_{i=1}^{r}\{b_i, b_{2m-3-i}\} \cup \bigcup_{j=1}^{s}\{b_{r+2i_j}, b_{r+2i_j-1}\} \cup \bigcup_{\alpha=1}^{p}\{c_{2i_\alpha}, c_{2i_\alpha-1}, c_{2j_\alpha}, c_{2j_\alpha-1}\}$$

can be enumerated as $z_1, \ldots, z_{n-2k}$ so that $F = N + \sum_{i=1}^{n-2k}(-1)^i z_i$ satisfies (6.17). Since $f_i \le (1-c)m - 11$ holds for every $1 \le i \le m-r-2$, $i \notin \{i_1, \ldots, i_s\}$, removing $z_1, \ldots, z_{n-2k}$ form the sequence $b_1, \ldots, b_{2m-4}$, the rest can be rearranged as $x_3, y_3, \ldots, x_k, y_k$ such that $1 \le e_i = x_i - y_i$ satisfies (6.16). Finally, it follows from (6.19) that

$$\sum_{i=1}^{k} e_i \le \sum_{i=1}^{m-r-2} f_i + 3 \le 3m + 3 \le (4-4c)m - 40 \le 4k - 12,$$

therefore condition (6.15) is also fulfilled. This completes the proof of Theorem 6.1.

6.2 Arithmetic Progressions and a Conjecture of Alon

In this section we first collect a few simple consequences of the Dias da Silva–Hamidoune theorem. Based on these we prove Theorem 2.22 in the second subsection. Finally we derive Theorems 2.24 and 2.25 based on Lev's result (Theorem 2.23) and briefly sketch how the ideas of the first two subsections can be applied to prove Alon's conjecture without depending on Lev's theorem.

Preliminaries

Throughout this subsection we work with and integer $d \ge 2$ and a prime p which is usually large enough compared to d. By a d-set we mean a set of cardinality d. To simplify notation, we introduce

$$n_d(p) = \left\lfloor \frac{p+d-2}{d} \right\rfloor < \frac{p}{d} + 1.$$

From now on A will always denote a subset of $[1, p]$. An immediate consequence of the Dias da Silva–Hamidoune theorem (Theorem 2.10) is that if $|A| \ge n_d(p) + d$, then $\Sigma_d(A)$ intersects every residue class modulo p, see [44]. By routine induction one obtains the following generalization (see [44], Corollary 2.3).

Lemma 6.5. *Let j be a positive integer, and assume that $|A| \ge n_d(p) + jd$. Then for every sequence $x_1, x_2, \ldots, x_j$ of integers there exists a sequence $A_1, A_2, \ldots, A_j$ of pairwise disjoint d-sets of A such that $\sigma(A_i) \equiv x_i \pmod p$ for every $1 \le i \le j$.*

To prove Theorem 2.22 we use a method similar to the one developed by Hamidoune in [44]. From this point on, however, we proceed somewhat differently.

Lemma 6.6. *Let j be a positive integer, and assume that $|A| \ge n_d(p)+jd$. Then there exists a sequence $A_1, A_2, \ldots, A_j$ of pairwise disjoint d-sets of A such that $\sigma(A_i) \in \{p, 2p, \ldots, (d-1)p\}$ for every $1 \le i \le j$. In particular, with $K = \lceil j/(d-1) \rceil$, there exists an integer $t \in [1, d-1]$ and a sequence $B_1, B_2, \ldots, B_K$ of pairwise disjoint d-sets of A such that $\sigma(B_i) = tp$ for every $1 \le i \le K$.*

Proof. Apply the previous lemma with $x_1, x_2, \ldots, x_j = 0$. Then $\sigma(A_i) < dp$ is divisible by p for every $1 \le i \le K$. The second statement follows from the pigeonhole principle. □

This inspires the following definition. We denote by $t(A) = t_d(A)$ any integer $1 \le t \le d-1$ for which the number of pairwise disjoint d-subsets B of A with the property $\sigma(B) = tp$ is maximum.

Lemma 6.7. *Assume that $|A| \ge n_d(p) + 2d^4$. Then for every integer x which is divisible by $t(A)$ and satisfies $t(A)dp \le x < d^2p$, there is a subset $X \subset A$ such that $\sigma(X) = x$.*

Proof. Consider the integer $y = x/t(A)$. In view of Lemma 6.5, there exist pairwise disjoint sets $A_1, A_2, \ldots, A_{(t(A)-1)d+1} \subset A$ of cardinality d such that $\sigma(A_i) \equiv y \pmod p$ for each $1 \le i \le (t(A)-1)d+1$. Since $\sigma(A_i) < dp$, there is a subsequence $A_{i_1}, A_{i_2}, \ldots, A_{i_{t(A)}}$ and an integer $0 \le s \le d-1$ such that $\sigma(A_{i_j}) = sp + y_0$, where $0 \le y_0 < p$ and $y \equiv y_0 \pmod p$. Then $B = A_{i_1} \cup \ldots \cup A_{i_{t(A)}}$ satisfies $|B| = t(A)d \le d(d-1)$, $\sigma(B) = t(A)(sp+y_0) < t(A)dp \le x$ and $x - \sigma(B) \equiv x - t(A)y_0 \equiv x - t(A)y \equiv 0 \pmod p$. Moreover, it is also divisible by $t(A)$, hence also by $t(A)p$. Due to the definition of $t(A)$, in view of Lemma 6.6 there exist pairwise disjoint d-sets $B_1, \ldots, B_{2d^2}$ such that $\sigma(B_i) = t(A)p$. Since $|B| < d^2$, wlog. we may assume that $B_1, \ldots, B_{d^2}$ are disjoint from B. Since $x - \sigma(B) < d^2p$ is divisible by $t(A)p$, there is an index $j < d^2$ such that $x - \sigma(B) = \sigma(B_1) \cup \ldots \cup \sigma(B_j)$. Thus, $x = \sigma(B \cup B_1 \cup \ldots \cup B_j)$. □

The Second Conjecture of Lev: Proof of Theorem 2.22

In view of Theorem 2.19 we may assume that

$$\frac{3n-3}{2} \le \ell \le 2n-6.$$

From Corollary 1 in [63] it follows, that

$$[2\ell - 2n + 1, 3\ell + 2] \subseteq \Sigma(A).$$

Moreover, $2\ell - 2n + 1 \le \ell - 5$, thus the above interval contains at least $2\ell + 8$ consecutive integers.

Our strategy is the following. First we choose a prime $(1-\varepsilon)\ell \le p \le \ell$. Note that we need only a few terms to represent each number in the above interval I as an element of $\Sigma(A)$. Thus if ε is small enough, then the density of the remaining elements of A in $[1, p]$ is considerably larger than $1/3$, and then we can use Lemma 6.6 to extend the length of the interval I by $2p$ in each of several iterations, until it gets long enough to continue with the second phase.

Since $\Sigma(A)$ is symmetric about $\sigma(A)/2$, it is enough to extend the interval until it contains $\lfloor \sigma(A)/2 \rfloor$. In the second phase we choose a prime q between ℓ and $(1+\delta)\ell$ and consider A as a subset of $[1, q]$. As the length of I grows, the density of the remaining elements that we

can use for the extension of I is getting smaller. The point is, it stays above a certain bound, and thus in each iteration we can extend the length of I by the same *universal* multiple of q, which allows us to complete the phase.

To see what exactly is needed in the first phase, we start with the second one. If n is large enough, then there is a prime number q such that $\ell \le q \le 17\ell/16$. Denote by L the least common multiple of the numbers $2, 3, \ldots, 17$. Assume that $x \le \sigma(A)/2$, and $y = x - Lq$ belongs to $\Sigma(A)$, that is, there is a subset B of A such that $y = \sigma(B)$. Then $\sigma(A \setminus B) = \sigma(A) - y > \sigma(A)/2 > n^2/4$. Consequently,

$$|A \setminus B| > \frac{n^2}{4\ell} > \frac{n}{8} > \frac{\ell}{16} \ge \frac{p}{17} > N_{18}(p) + \frac{p}{17 \cdot 18} - 1 > N_{18}(p) + 18^2 L,$$

if n and hence p is large enough. According to Lemma 6.6, there exists an integer $t \le 17$ and a sequence $B_1, B_2, \ldots, B_L$ of pairwise disjoint subsets of $A \setminus B$ such that $\sigma(B_i) = tq$ for every $1 \le i \le L$. It follows that

$$x = \sigma(B \cup B_1 \cup \ldots \cup B_{L/t}),$$

proving that $x \in \Sigma(A)$. Accordingly, we only have to prove that

$$[2\ell - 2n + 1, 2\ell - 2n + Lq] \subseteq \Sigma(A).$$

Then it follows from the above argument that $[2\ell - 2n + 1, \lfloor \sigma(A)/2 \rfloor] \subseteq \Sigma(A)$. By symmetry we find that $[\lceil \sigma(A)/2 \rceil, \sigma(A) - (2\ell - 2n + 1)] \subseteq \Sigma(A)$, implying Theorem 2.22.

Turning thus to the first phase, we choose a prime p such that $(15/16)\ell \le p \le \ell$, and put $M = 17L/30$, so that $Lq \le 2Mp$. Let $A' = A \cap [1, p]$, then $|A'| \ge n - (\ell - p)$. Choose any integer x satisfying $3\ell + 2 < x \le 3\ell + 2 + 2Mp$. Then there is an integer $y \in [3\ell + 3 - 2p, 3\ell + 2]$ such that $y \equiv x \pmod{2p}$. Since $2p \le 2\ell$, we have $y \in [2\ell - 2n + 1, 3\ell + 2]$, and thus there is $C \subseteq A$ such that $y = \sigma(C)$. Note that $|C| < \sqrt{6\ell + 4}$ and we have $x - y = 2mp$ with some integer $m \in [1, M]$. Now

$$|A' \setminus C| > n - (\ell - p) - \sqrt{6\ell + 4} > \frac{\ell}{2} - \frac{\ell}{16} - \sqrt{6\ell + 4} > N_3(p) + 12M,$$

provided that n is large enough. In view of Lemma 6.6, there exists a sequence $C_1, C_2, \ldots, C_{2M}$ of pairwise disjoint subsets of $A' \setminus C$ such that either $\sigma(C_i) = p$ holds for every $1 \le i \le 2M$, or $\sigma(C_i) = 2p$ is true for every $1 \le i \le 2M$. In the first case we find that $x = \sigma(C \cup C_1 \cup \ldots \cup C_{2m})$, whereas $x = \sigma(C \cup C_1 \cup \ldots \cup C_m)$ in the second case. Consequently $x \in \Sigma(A)$, hence $[3\ell + 3, 3\ell + 2 + 2Mp] \subseteq \Sigma(A)$. Since $[2\ell - 2n + 1, 3\ell + 2] \subseteq \Sigma(A)$, it follows that

$$[2\ell - 2n + 1, 2\ell - 2n + Lq] \subseteq [2\ell - 2n + 1, 3\ell + 2 + 2Mp] \subseteq \Sigma(A).$$

This completes the proof of Theorem 2.22.

We note that applying the method of Section 6.1 and developing Lemmas 6.5 and 6.6 in a different direction the idea of the previous proof leads to a more effective version of Theorem 2.22. We do not elaborate on this here.

The Conjecture of Alon

Proof of Theorem 2.24. We prove the statement by induction on d with $n_1 = n_1(d) = n_0 + d\log_{3/2} d$, where n_0 is as in Theorem 2.23. If $d \le n/400\ln n$, then

$$d\log_{3/2} d \le \frac{n}{400\ln n}\cdot\frac{\ln n}{\ln(3/2)} < \frac{n}{101},$$

hence the theorem follows with $n_1 = 1.01n_0$. For the inductional step we may clearly assume that

$$n \le \frac{\ell}{d-1} + (d-1) - 2. \tag{6.24}$$

It may also be assumed for the initial step $d = 2$, since otherwise we have $\ell = n$, in which case obviously $\Sigma(A) = [0, \sigma(A)]$.

Assume that $N_q(A) \ge q-1$ holds for every positive integer $q < d$; this is a priori true if $d = 2$. Since $n \ge \lfloor \ell/d \rfloor + d - 1$, it is also true for $q = d$. Moreover, if $q = d + r$ with $1 \le r < d$, then in view of $n > 10d \ge 5d^2/(d-1)$, condition (6.24) implies

$$\ell \ge (d-1)n - (d-1)(d-3) > 4d^2 > \Big(1 + \frac{1}{r}\Big)d(d+r),$$

and consequently

$$N_q(A) \ge n - \Big\lfloor \frac{\ell}{q} \Big\rfloor > \frac{\ell}{d} - \frac{\ell}{d+r} + d - 2 \ge (r+1) + (d-2) = q - 1.$$

Therefore in this case $n_q(A) \ge q-1$ holds for every $q < 2d \le 2\ell/n$. Since $n \ge \ell/d \ge 400\ell\ln n/n$, we have $n \ge 20(\ell\ln n)^{1/2}$. Thus it follows from Theorem 2.23 that $\Sigma(A)$ contains every integer in the interval $[\lambda\sigma(A), (1-\lambda)\sigma(A)]$ with $\lambda = 280\ell/n^2$. Given that

$$\lambda\sigma(A) < \lambda(n\ell) = \frac{280\ell^2}{n} \le 280d\ell,$$

our statement follows with $t = 1$ in this case. This includes the initial step $d = 2$.

It remains to study the case when $N_q(A) \le q-2$ for some integer $2 \le q < d$. Collect in A_0 those elements of A that are not divisible by q, and define the a set of integers A' such that

$$A = A_0 \cup \{qa \mid a \in A'\}.$$

Introduce the integers $\ell' = \lfloor \ell/q \rfloor$ and $d' = \lceil d/q \rceil$, then $2 \le d' \le (2/3)d < d$ and A' is a set of integers in the interval $[1, \ell']$ whose cardinality satisfies

$$n' = |A'| = |A| - |A_0| > \frac{\ell}{d} + d - q > \frac{\ell'}{d'} + d' - 2$$

and

$$n' \ge n - (q-2) > n_0 + d\log_{3/2} d - d \ge n_0 + d\log_{3/2} d' > n_1(d').$$

Since $n > 3d$, we also have

$$\frac{n'}{400\ln n'} > \frac{n-d}{400\ln n} > \frac{2}{3}\cdot\frac{n}{400\ln n} \ge \frac{2d}{3} \ge d'.$$

It follows from the inductional hypothesis that there is an integer $t' \in [1, d'-1]$ such that $\Sigma(A')$ contains all multiples of t' that belong to the interval

$$[280d'\ell', \sigma(A') - 280d'\ell'].$$

Accordingly, $q\Sigma(A') \subseteq \Sigma(A)$ contains all multiples of $t = qt' < d$ lying in the interval

$$[280qd'\ell', q\sigma(A') - 280qd'\ell'].$$

Given that $qd'\ell' \le d'\ell \le (2/3)d\ell < d\ell$, and that in view of $\sigma(A_0) \le (q-2)\ell < d\ell$ we also have

$$q\sigma(A') - 280qd'\ell' > \sigma(A) - \sigma(A_0) - 200d\ell > \sigma(A) - 280d\ell,$$

this completes the inductional step. □

Now it is easy to prove Theorem 2.25. For any integer $s \ge 2$, denote by $\psi(s)$ the least common multiple of the numbers $2, 3, \dots, s$. A routine application of the prime number theorem gives $\ln \psi(s) = (1+o(1))s$. Thus, if $m \le \ell^2$, then $d = \mathrm{snd}(m) < (2+o(1))\ln \ell$. If ℓ is sufficiently large, then $d \le 3\ln \ell$, and a set $A \subseteq [1, \ell]$ of cardinality $|A| = \lfloor \ell/d \rfloor + d - 2$ with the property that $m \notin \Sigma(A)$ can be constructed as follows. Suppose that $m \equiv i \pmod d$, where $1 \le i \le d-1$. Let A consist of all $\lfloor \ell/d \rfloor$ multiples of d in $I = [1, \ell]$, $i-1$ different elements of I that are congruent to 1 modulo d, and $d-i-1$ additional members of I that are congruent to -1 modulo d. It is easy to check that A has the claimed property, see [5] for details.

It remains to check that if ℓ is large enough and $(280+\varepsilon)\ell \ln \ell < m < \ell^2/(8+\varepsilon)\ln^2 \ell$, then with $d = \mathrm{snd}(m)$ it is true that $A \subseteq [1, \ell]$, $|A| \ge \lfloor \ell/d \rfloor + d - 1$ implies that $m \in \Sigma(A)$. We may assume that $\varepsilon < 1$. If ℓ is large enough and $(280+\varepsilon)\ell \ln \ell < m < 574\ell \ln \ell$, then $d < (1+\varepsilon/280)\ln \ell$, the conditions of Theorem 2.24 are satisfied, and m belongs to the interval $[280d\ell, \sigma(A) - 280d\ell]$. Since m is divisible by every integer $t \in [1, d-1]$, it follows that $m \in \Sigma(A)$. If ℓ is large enough and $574\ell \ln \ell \le m < \ell^2/(8+\varepsilon)\ln^2 \ell$, then $d < (2+\varepsilon/20)\ln \ell$, the conditions of Theorem 2.24 are once again satisfied, and m belongs to the interval $[280d\ell, \sigma(A) - 280d\ell]$, since $280d\ell < 574\ell \ln \ell$ and

$$\sigma(A) - 280d\ell > \frac{1}{2}\Big(\frac{\ell}{(2+\varepsilon/20)\ln \ell}\Big)^2 - 574\ell \ln \ell > \frac{\ell^2}{(8+\varepsilon)\ln \ell^2}.$$

Once again, it follows that $m \in \Sigma(A)$. □

Our original idea to prove Theorem 2.25 for a slightly shorter range of m was to follow the method we described in the second subsection for the proof of Theorem 2.22. Here Lemma 6.7 seems to be a good starting point to build up a long arithmetic progression in $\Sigma(A)$ for A dense enough. To control the difference when extending this arithmetic progression the way the large block is extended in the proof of Theorem 2.22 is, however, a nontrivial task which

leads to several technical difficulties that we do not discuss here. Basically we prove a mixture of Theorems 2.23 and 2.24 when ℓ is a prime and d does not exceed a small power of n. The transfer to the general case depends on the following simple lemma:

Lemma 6.8. *Let A be a set of integers and q a positive integer such that $N_q(A) \geq q-1$. Then there exists a proper divisor d of q such that $\Sigma(A)$ intersects each residue class modulo q that is divisible by d.*

We note that this lemma is tight when q is a power of a prime p: If A consist of $q/p-1$ elements that are congruent to -1 modulo q and $(p-1)q/p-1$ additional elements that are congruent to 1 modulo q, then the conclusion fails.

We find it stylish to conclude this dissertation indicating a proof of Lemma 6.8 in the case when q is a prime. Assume for that end that $a_1, \ldots, a_{q-1}$ are nonzero elements of the Galois field $\mathbb{F}_q$. All we have to prove is that

$$\Sigma(a_1, \ldots, a_{q-1}) = \big\{\sum_{i=1}^{q-1} \varepsilon_i a_i \mid \varepsilon_i \in \{0,1\}\big\} = \mathbb{F}_q.$$

Assume on the contrary that $\Sigma(a_1, \ldots, a_{q-1})$ is contained in a set $C \subset \mathbb{F}_q$ of cardinality $q-1$. Put $A_i = \{0, a_i\}$, then $|A_1| = \ldots = |A_{q-1}| = 2$. The polynomial

$$f(x_1, \ldots, x_{q-1}) = \prod_{c \in C} (x_1 + \ldots + x_{q-1} - c) \in \mathbb{F}_q[x_1, \ldots, x_{q-1}]$$

has a leading term $\alpha x_1 \ldots x_{q-1}$, where $\alpha = (q-1)!$ is a nonzero element of $\mathbb{F}_q$. It follows from the polynomial lemma (Lemma 2.27) that f cannot vanish on $A_1 \times \ldots \times A_{q-1}$, a contradiction.

Epilogue

It may be interesting to see how the work contained in this dissertation has developed. My first encounter with combinatorial number theory occurred at the 1996 DIMACS Workshop. During that workshop I solved the first problem of Lev and started studying his work which led to some versions of the results discussed in Chapter 6. The original proofs were fairly complicated, and I never since had the time and the energy to write them up. This summer, however, I received a letter from Stefanie Gerke in London, who wanted to apply Theorem 2.22 for a problem in graph theory. That inquiry finally triggered a follow-up work I could carry out in the peaceful environment of the CWI. It led to a lot of simplifications, stronger results and the papers [57, 58].

At the end of 1999 I visited Oriol Serra in Barcelona, who showed me Alon's paper [3]. The multiplicative analogue I invented during that visit resulted in the paper [20]. During the spring of 2003, enjoying the privileged life of a research associate at the Rényi Institute without any teaching duties, I first started further pursuing that idea, which led to the paper [52]. That was when I convinced myself of the possibility that the Combinatorial Nullstellensatz may be applied to get structural results. Having no stress and time pressure at the time — a rare occasion —, I could concentrate on just one difficult problem for months, and the result was Theorem 2.14. It was also during that period and a month spent at the UPC in Barcelona, when I could finally write the expository paper [53], invited by Shalom Eliahou, Isidoro Gitler and Jarik Nešetřil for a special volume of DM. In order to extend Theorem 2.14 to general abelian groups I had to first invent a new proof of Theorem 2.11, which appeared in [51].

The submission of the paper [54] had yet to wait for another year when I once again had a chance working on it at the I.H.É.S. in France. Extending the ideas of [51] to noncommutative groups I was able to carry out during a month's visit at the ETH Zürich in 2005. A second application of the Combinatorial Nullstellensatz [56] occurred to me during my last visit at the I.H.É.S. the following summer. In retrospect, that paper should have preceded [54], but at that time it seemed very complicated to handle the restricted sumset of two different sets this way. Here we presented them in the more logical order. Once again, I am indebted to all these great institutions where I was given the chance to carry out my research.

Bibliography

[1] N. Alon, Subset sums, *J. Number Th.* **27** (1987) 196–205

[2] N. Alon, Combinatorial Nullstellensatz, *Combin. Prob. Comput.* **8** (1999) 7–29

[3] N. Alon, Additive Latin transversals, *Israel J. Math.* **117** (2000) 125–130

[4] N. Alon, Discrete mathematics: methods and challenges, in: *Proc. Internat. Congr. Math. Vol. I. (Beijing, 2002)*, Higher Educational Press, Beijing, 2002, pp. 119–135

[5] N. Alon and G. Freiman, On sums of subsets of a set of integers, *Combinatorica* **8** (1988) 297–306

[6] N. Alon and Z. Füredi, Covering the cube by affine hyperplanes, *European J. Combin.* **14** (1993) 79–83

[7] N. Alon, M.B. Nathanson, and I.Z. Ruzsa, Adding distinct congruence classes modulo a prime, *Amer. Math. Monthly* **102** (1995) 250–255

[8] N. Alon, M.B. Nathanson, and I.Z. Ruzsa, The polynomial method and restricted sums of congruence classes, *J. Number Th.* **56** (1996) 404–417

[9] N. Alon and M. Tarsi, A nowhere-zero point in linear mappings, *Combinatorica* **9** (1989) 393–395

[10] N. Alon and M. Tarsi, Colorings and orientations of graphs, *Combinatorica* **12** (1992) 125–134

[11] P. Balister and J.P. Wheeler, The Erdős–Heilbronn conjecture for finite groups, to appear in *Acta Arithm.* (2007)

[12] Y.F. Bilu, Structure of sets with small sumsets, *Astérisque* **258** (1999) 77–108

[13] Y.F. Bilu, V.F. Lev, and I.Z. Ruzsa, Rectification principles in additive number theory, *Discrete Comput. Geom.* **19** (1998) 343–353

[14] L.V. Brailovsky and G.A. Freiman, On a product of finite subsets in a torsion free group, *J. Algebra* **130** (1990) 462–476

[15] A.E. Brouwer and A. Schrijver, The blocking number of an affine space, *J. Combin Th. A* **24** (1978) 251–253

[16] A.L. Cauchy, Recherches sur les nombres, *J. École Polytech.* **9** (1813) 99–116

[17] C.C. Chang and H.J. Keisler, Model Theory, North-Holland, 1973

[18] M-C. Chang, A polynomial bound in Freiman's theorem, *Duke Math. J.* **113** (2002) 399–419

[19] I. Chowla, A theorem on the addition of residue classes: Application to the number $\Gamma(k)$ in Waring's problem, *Proc. Indian Acad. Sci. A* **1** (1935) 242–243

[20] S. Dasgupta, Gy. Károlyi, O. Serra, and B. Szegedy, Transversals of additive latin squares, *Israel J. Math.* **126** (2001) 17–28

[21] H. Davenport, On the addition of residue classes, *J. London Math. Soc.* **10** (1935) 30–32

[22] J.A. Dias da Silva and Y.O. Hamidoune, Cyclic spaces for Grassmann derivatives and additive theory, *Bull. London Math. Soc.* **26** (1994) 140–146

[23] D. Eisenbud, M. Green, and J. Harris, Cayley–Bacharach theorems and conjectures, *Bull. Amer. Math. Soc.* **33** (1996) 295–324

[24] S. Eliahou and M. Kervaire, Sumsets in vector spaces over finite fields, *J. Number Th.* **71** (1998) 12–39

[25] S. Eliahou and M. Kervaire, Restricted sums of sets of cardinality $1 + p$ in a vector space over F_p, *Discrete Math.* **235** (2001) 199–213

[26] S. Eliahou and M. Kervaire, Restricted sumsets in finite vector spaces: the case $p = 3$, *Integers* **1** (2001), Research paper A2, 19 pages (electronic)

[27] S. Eliahou and M. Kervaire, Sumsets in dihedral groups, *European J. Combin.* **27** (2007) 617–628

[28] S. Eliahou and M. Kervaire, The small sumsets property for solvable finite groups, *European J. Combin.* **27** (2007) 1102–1110

[29] S. Eliahou, M. Kervaire, and A. Plagne, Optimally small sumsets in finite abelian groups, *J. Number Th.* **101** (2003) 338–348

[30] P. Erdős, Some problems in number theory, in: Computers in Number Theory (A.O.L. Atkin and B.J. Birch, eds.), Acedemic Press, 1971, pp. 405–414

[31] P. Erdős, Some problems and results on combinatorial number theory, *Ann. New York Acad. Sci.* **576** (1989) 132–145

[32] P. Erdős and R.L. Graham, Old and New Problems and Results in Combinatorial Number Theory, *L'Enseignement Mathématique*, Geneva, 1980

[33] W. Feit and J.G. Thompson, Solvability of groups of odd order, *Pacific J. Math.* **13** (1963) 775–1029

[34] G.A. Freiman, On the addition of finite sets. I, *Izv. Vysh. Ucheb. Zaved. Mat.* **13** (1959) 202–213

[35] G.A. Freiman, Inverse problems of additive number theory. On the addition of sets of residues with respect to a prime modulus, *Doklady Akad. Nauk SSSR* **141** (1961) 571–573

[36] G.A. Freiman, Addition of finite sets, *Doklady Akad. Nauk SSSR* **158** (1964) 1038–1041

[37] G.A. Freiman, Foundations of a Structural Theory of Set Addition, *Translations of Mathematical Monographs* **37** AMS, 1973

[38] G.A. Freiman, New analytical results in subset sum problem, *Discrete Math.* **114** (1993) 205–217

[39] G.A. Freiman, L. Low, and J. Pitman, Sumsets with distinct summands and the conjecture of Erdős–Heilbronn on sums of residues, *Astèrisque* **258** (1999) 163–172

[40] M. Goldstern and H. Judah, The Incompleteness Phenomenon, A.K. Peters, Wellesley, 1995

[41] W.D. Gao and D.J. Wang, Additive Latin transversals and group rings, *Israel J. Math.* **140** (2004) 375–380

[42] B. Green and I.Z. Ruzsa, Freiman's theorem in an arbitrary abelian group, *J. London Math. Soc. (2)* **75** (2007) 163–175

[43] Y.O. Hamidoune, A generalization of an addition theorem of Shatrowsky, *European J. Combin.* **13** (1992) 249–255

[44] Y.O. Hamidoune, The representation of some integers as a subset sum, *Bull. London Math. Soc.* **26** (1994) 557–563

[45] Y.O. Hamidoune, On small subset product in a group, *Astérisque* **258** (1999) 281–308

[46] Y.O. Hamidoune, Personal communication (2005)

[47] Y.O. Hamidoune, A.S. Lladó, and O. Serra, On restricted sums, *Combin. Prob. Comput.* **9** (2000) 513–518

[48] Y.O. Hamidoune and Ø.J. Rødseth, An inverse theorem mod p, *Acta Arith.* **92** (2000) 251–262

[49] G. Harcos and I.Z. Ruzsa, A problem on zero subsums in abelian groups, *Period. Math. Hungar.* **35** (1997) 31–34

[50] J.F. Humphreys, A Course in Group Theory, Oxford University Press, 1996

[51] Gy. Károlyi, On restricted set addition in abelian groups, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **46** (2003) 47–54

[52] Gy. Károlyi, The Erdős–Heilbronn problem in abelian groups, *Israel J. Math.* **139** (2004) 349–359

[53] Gy. Károlyi, A compactness argument in the additive theory and the polynomial method, *Discrete Math.* **302** (2005) 124–144

[54] Gy. Károlyi, An inverse theorem for the restricted set addition in abelian groups, *J. Algebra* **290** (2005) 557–593

[55] Gy. Károlyi, Cauchy–Davenport theorem in group extensions, *Enseign. Math.* **51** (2005) 239–254

[56] Gy. Károlyi, Restricted set addition: The exceptional case of the Erdős–Heilbronn conjecture, submitted to *J. Combin. Th. A* (9 pages)

[57] Gy. Károlyi, Balanced subset sums in dense sets of integers, submitted to *Integers. Electron. J. Combin. Number Th.* (13 pages)

[58] Gy. Károlyi, Long arithmetic progressions in subset sums and a conjecture of Alon, in preparation

[59] J.H.B. Kemperman, On complexes in a semigroup, *Nederl. Akad. Wetensch. Indag. Math.* **18** (1956) 247–254

[60] J.H.B. Kemperman, On small sumsets in an abelian group, *Acta Math.* **103** (1960) 63–88

[61] M. Kneser, Abschätzungen der asymptotischen Dichte von Summenmengen, *Math. Z.* **58** (1953) 459–484

[62] S. Lang, Algebraic Number Theory (2nd Edition), *GTM* **110**, Springer, 1994

[63] V.F. Lev, On consecutive subset sums, *Discrete Math.* **187** (1998) 151–160

[64] V.F. Lev, On small subsets in abelian groups, *Astérisque* **258** (1999) 317–321

[65] V.F. Lev, Restricted set addition in groups. I: The classical setting, *J. London. Math. Soc. (2)* **62** (2000) 27–40

[66] V.F. Lev, Restricted set addition in groups. II: A generalization of the Erdős-Heilbronn conjecture, *Electron. J. Combin.* **7** (2000), Research paper R4, 10 pages (electronic)

[67] V.F. Lev, Blocks and progressions in subset sum sets, *Acta Arithm.* **106** (2003) 123–142

[68] E. Lipkin, On representation of r-th powers by subset sums, *Acta Arithm.* **52** (1989) 353–365

[69] M.B. Nathanson, Additive Number Theory. Inverse Problems and the Geometry of Sumsets, *GTM* **165**, Springer, 1996

[70] J.E. Olson, On the symmetric difference of twos sets in a group, *European J. Combin.* **7** (1986) 43–54

[71] A. Pál, Personal communication

[72] S.S. Pillai, Generalization of a theorem of Davenport on the addition of residue classes, *Proc. Indian Acad. Sci. A* **6** (1938) 179–180

[73] J.M. Pollard, A generalization of a theorem of Cauchy and Davenport, *J. London Math. Soc.* **8** (1974) 460–462

[74] L. Pyber, Personal communication

[75] C. Reiher, On Kemnitz' conjecture concerning lattice points in the plane, *Ramanujan J.* **13** (2007) 333–337

[76] D.J.S. Robinson, A Course in the Theory of Groups (2nd ed.), *GTM* **80**, Springer, 1996

[77] L. Rónyai, On a conjecture of Kemnitz, *Combinatorica* **20** (2000) 569–573

[78] I.Z. Ruzsa, Arithmetic progressions and the number of sums, *Period. Math. Hungar.* **25** (1992) 105–111

[79] I.Z. Ruzsa, Generalized arithmetic progressions and sumsets, *Acta Math. Hungar.* **65** (1994) 379–388

[80] I.Z. Ruzsa, Personal communication (2006)

[81] A. Sárközy, Finite addition theorems. II, *J. Number Th.* **48** (1994) 197–218

[82] L. Shatrowsky, A new generalization of the Davenport's–Pillai's theorem on the addition of residue classes, *C. R. (Doklady) Akad. Sci. USSR, N. S.* **45** (1944) 315–317

[83] H. Snevily, The Cayley addition table of Z_n, *Amer. Math. Monthly* **106** (1999) 584–585

[84] Z.W. Sun, On Snevily's conjecture and restricted sumsets, *J. Combin. Th. A* **103** (2003) 291–304

[85] Z.W. Sun and Y.N. Yeh, On various restricted sumsets, *J. Number Th.* **114** (2005) 209–220

[86] T. Tao and V.H. Vu, Additive Combinatorics, Cambridge University Press, 2006

[87] A.G. Vosper, The critical pairs of subsets of a group of prime order, *J. London Math. Soc.* **31** (1956) 200–205, Addendum 280–282

[88] S. Yuzvinsky, Orthogonal pairings of Euclidean spaces, *Michigan Math. J.* **28** (1981) 109–119

[89] G. Zémor, A generalization to noncommutative groups of a theorem of Mann, *Discrete Math.* **126** (1994) 365–372