Additive Kombinatorik und Zahlentheorie
Christian Bernert, Stand: 12. Mai 2026

0 Motivation

0.1 Inverse Probleme und die Geometrie von Summenmengen

Definition 0.1. *Für eine abelsche Gruppe $(G,+)$ und Teilmengen $A,B\subset G$ sei*

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

die Summenmenge von A und B. Wir schreiben auch $2A=A+A$.

Beachte, dass $2A$ im Allgemeinen nicht identisch mit der Menge $\{2a : a\in A\}$ ist (sondern größer).

Wichtige Beispiele sind $G=\mathbb{Z}$ oder $G=\mathbb{R}$, auch $G=\mathbb{F}_p=\mathbb{Z}/p\mathbb{Z}$ oder $G=\mathbb{F}_p^n$.

Wir wollen uns zunächst davon überzeugen, dass einige der wichtigsten zahlentheoretischen Probleme von solchen Summenmengen handeln. Dazu betrachten wir $G=\mathbb{Z}$ sowie die Mengen

$$\begin{aligned}\mathcal{P}&=\{3,5,7,11,13,17,19,23,\dots\}\quad\text{(ungerade Primzahlen)}\\ \mathcal{Q}&=\{0,1,4,9,16,25,\dots\}\quad\text{Quadratzahlen}\\ \mathcal{C}&=\{\dots,-8,-1,0,1,8,27,64,\dots\}\quad\text{Kubikzahlen}\end{aligned}$$

Dann können wir die folgenden Summenmengen betrachten:

$4\mathcal{Q}=\mathcal{Q}+\mathcal{Q}+\mathcal{Q}+\mathcal{Q}=\mathbb{Z}_{\geq 0}$ (Vierquadratesatz, Lagrange 1770)
$2\mathcal{Q}$ (explizite Beschreibung, Zweiquadratesatz von Fermat)
$3\mathcal{P}=\{9,11,13,\dots\}$ (ungerade Zahlen ≥ 9, Helfgott 2013, schwache Goldbachsche Vermutung)
$2\mathcal{P}=\{6,8,10,12,\dots\}$ (alle geraden Zahlen ≥ 6? Goldbachsche Vermutung)
$3\mathcal{C}$ (alle $n\not\equiv 4,5 \pmod 9$?)

Beispielsweise wissen wir nicht, ob $114\in 3\mathcal{C}$ ist, aber 42 ist es wegen

$$42=(-80538738812075974)^3+(80435758145817515)^3+(12602123297335631)^3.$$

(Diese kleinste bekannte Lösung wurde erst 2019 von Andrew Booker und Andrew Sutherland gefunden.)

Diese Liste zeigt, dass es sehr schwer sein kann, Summenmengen explizit zu beschreiben. Ziel der additiven Kombinatorik ist es, Aussagen über Summenmengen zu zeigen, die möglichst wenig Informationen über die Struktur der zugrundeliegenden Menge nutzen. Dies allein wird zwar vermutlich nicht reichen, um die oben genannten Probleme zu lösen, bei anderen wird es aber durchaus weiterhelfen, wie wir gleich sehen werden.
Zunächst wollen wir untersuchen, was sich überhaupt über die Größe der Summenmenge aussagen lässt:

Satz 0.2. *Für endliche nichtleere Teilmengen $A,B\subset\mathbb{R}$ gilt*

$$|A+B|\geq |A|+|B|-1.$$

Beweis. Wir schreiben $A = \{a_1, \dots, a_k\}$ mit $a_1 < a_2 < \cdots < a_k$ und analog $B = \{b_1, \dots, b_m\}$ mit $b_1 < b_2 < \cdots < b_m$. Dann ist

$$a_1 + b_1 < a_1 + b_2 < \cdots < a_1 + b_m < a_2 + b_m < \cdots < a_k + b_m$$

eine streng monoton steigende Folge von $k + m - 1 = |A| + |B| - 1$ Elementen von $A + B$, die somit paarweise verschieden sind. □

In den Übungen werden wir sehen, dass die Gleichheit $|A + B| = |A| + |B| - 1$ genau dann gilt, wenn A und B *arithmetische Progressionen* mit der gleichen Schrittweite d sind. Dabei ist eine arithmetische Progression mit Schrittweite d und Länge k (auch kurz k-AP) eine Menge der Form

$$\{a_0, a_0 + d, a_0 + 2d, \dots, a_0 + (k-1)d\}.$$

Dies ist unser erstes Beispiel eines **inversen Theorems**: Wenn $|2A| = 2|A| - 1$ gilt, d.h. die Summenmenge ist so klein wie möglich, dann muss A eine arithmetische Progression sein.
Der **Satz von Freiman**, unser erstes größeres Ziel in dieser Vorlesung, ist eine weitreichende Verallgemeinerung davon. Grob gesagt ist die Aussage: Ist $|2A| \leq K \cdot |A|$ (wobei wir uns K als Konstante und $|A|$ als sehr groß vorstellen), dann ist A „fast“ eine „verallgemeinerte arithmetische Progression“ (in einem quantitativen Sinn, der von K abhängt).
Die Idee ist also, dass wir für eine zufällig gewählte Menge A eher $|A + A| \approx |A|^2$ erwarten. Ist die Summenmenge dagegen deutlich kleiner, z.B. $|A + A| \leq 100|A|$, dann sollte dies eine starke Aussage über die Struktur von A zur Folge haben.

0.2 Das Summen-Produkt-Phänomen

Für $A \subset \mathbb{N}$ können wir auch die *Produktmenge*

$$A^2 = \{a_1 \cdot a_2 : a_1, a_2 \in A\}$$

untersuchen. Ähnlich wie bei der Summenmenge folgt auch hier leicht $|A^2| \geq 2|A| - 1$ mit Gleichheit genau dann, wenn A eine *geometrische Progression* ist, also eine Menge der Form

$$\{a_0, a_0 \cdot q, a_0 \cdot q^2, \dots, a_0 \cdot q^{k-1}\}.$$

Man sieht nun aber leicht, dass eine Menge mit mindestens drei Elementen nicht gleichzeitig eine arithmetische und eine geometrische Progression sein kann, folglich kann nicht in beiden Abschätzungen für $|2A|$ und $|A^2|$ Gleichheit gelten.
Im Jahr 1983 haben die ungarischen Mathematiker **Paul Erdős** (1913–1996) und **Endre Szemerédi** (geb. 1940) dieses „Summen-Produkt-Phänomen“ genauer untersucht und konnten

$$\max(|A + A|, |A \cdot A|) \geq |A|^{1+\delta}$$

für eine explizite Konstante $\delta > 0$ zeigen, falls A hinreichend groß ist. Mit anderen Worten muss eine der beiden Mengen $A + A$ und $A \cdot A$ deutlich schneller als linear wachsen. Sie vermuteten, dass die Aussage sogar für jede Konstante $\delta < 1$ gilt, also etwa

$$\max(|A + A|, |A \cdot A|) \geq |A|^{1.999999}$$

für hinreichend große Mengen A. Dies bleibt weiterhin ein offenes Problem, aber wir werden in der Vorlesung

$$\max(|A + A|, |A \cdot A|) \geq |A|^{4/3}$$

beweisen, was im Wesentlichen das beste bekannte Resultat ist und 2009 von József Solymosi (geb. 1959), einem weiteren ungarischen Mathematiker, bewiesen wurde.

0.3 Arithmetische Progressionen

Wir haben bereits gesehen, dass arithmetische Progressionen eine zentrale Rolle in inversen Theoremen über Summenmengen spielen. Ein weiterer Kreis von Problemen beschäftigt sich mit der Frage, unter welchen Voraussetzungen eine Menge überhaupt arithmetische Progressionen enthält. Zur Einstimmung betrachten wir zunächst die folgenden drei Sätze:

- (van der Waerden 1927) Wenn wir die natürlichen Zahlen in endlich vielen Farben färben, gibt es eine Farbe mit beliebig langen arithmetischen Progressionen.
- (van der Corput 1930) Die Menge der Primzahlen enthält unendlich viele 3-AP.
- (Green–Tao 2006) Das gleiche gilt für k-AP für beliebige k.

Wir werden weder den Satz von van der Corput noch den Satz von Green–Tao in dieser Vorlesung beweisen, aber wollen dennoch anmerken, dass van der Corput harte analytische Methoden verwendet, während Green und Tao entscheidend auf Ideen der additiven Kombinatorik zurückgreifen.

Im Lichte dieser Sätze stellten **Erdős** und **Paul Turán** (1910–1976) im Jahr 1936 die folgenden Fragen:

1) Enthält jede Menge $A \subset \mathbb{N}$ von „positiver Dichte“ beliebig lange arithmetische Progressionen? (Dies verallgemeinert den Satz von van der Waerden, reicht aber nicht ganz aus, um auch die Primzahlen einzubeziehen, da diese Dichte Null haben.)

2) Reicht es bereits aus, wenn $\sum_{a\in A} \frac{1}{a} = \infty$ gilt? (Das würde sowohl den Fall von „positiver Dichte“ als auch die Menge der Primzahlen umfassen!)

1) wurde für 3-AP von **Klaus Roth** (1925–2015) im Jahr 1952 bewiesen und 1975 für beliebige Längen von **Endre Szemerédi**. Wir werden den Satz von Roth in dieser Vorlesung beweisen.
2) wurde für 3-AP von **Thomas Bloom** und **Olof Sisask** im Jahr 2020 bewiesen (durch eine sehr raffinierte Verallgemeinerung des Arguments von Roth) und ist für längere Progressionen weiterhin offen.

Immerhin lässt sich damit aber sagen: Die Primzahlen enthalten allein deshalb unendlich viele 3-AP, weil es „genügend“ Primzahlen gibt! Dies ist exakt das Leitmotiv der additiven Kombinatorik!

Zum Abschluss dieses Überblickskapitels wollen wir uns noch mit dem Kartenspiel SET beschäftigen.

Das Spiel SET enthält $3^4 = 81$ Karten, die jeweils Symbole in einer von drei möglichen Anzahlen, einer von drei möglichen Formen, Texturen und Farben zeigen, wobei jede Kombination genau einmal vorkommt.
Ein SET ist nun eine Kombination von drei Karten, bei denen in jeder der vier Kategorien entweder alle drei Karten die gleiche Eigenschaft haben oder alle drei verschiedenen Ausprägungen vorkommen.
Zu Beginn des Spiels werden 12 Karten auf den Tisch gelegt. Wenn sich unter diesen kein SET befindet, legt man drei dazu usw. Es ist nun eine natürliche Frage, wie viele Karten man mindestens hinlegen muss, um sicher ein SET zu entdecken. Dieses Problem wurde bereits 1971 gelöst: Unter 21 Karten gibt es sicher ein SET, unter 20 noch nicht unbedingt.

Was hat dies nun mit arithmetischen Progressionen zu tun?
Wir können die Karten mathematisch als Elemente von $\mathbb{F}_3^4 = \{0,1,2\}^4$ modellieren, wobei jede der vier Koordinaten einer der vier Kategorien entspricht und die drei möglichen Einträge den drei möglichen Ausprägungen.
Die erste Beobachtung ist nun, dass $\{x,y,z\}$ genau dann ein SET ist, wenn $x+y+z=0$ ist. Weil wir uns in Charakteristik 3 befinden, lässt sich dies noch weiter zu $x+z=2y$ bzw. $x-y=y-z$ umformen, d.h. ein SET ist exakt eine arithmetische Progression in $\mathbb{F}_3^4$!

Die oben beschriebene Frage ist also äquivalent dazu, wie groß eine Teilmenge von $\mathbb{F}_3^4$ mindestens sein muss, um sicher eine arithmetische Progression zu enthalten.

Verallgemeinert man diese Fragestellung auf n Ausprägungen, so würde ein Analogon des Satzes von Roth (den man komplett analog beweisen kann) sagen, dass es unter 1% der 3^n Karten sicher ein SET gibt, falls n hinreichend groß ist.

Vollkommen überraschend wurde aber 2016 von **Jordan Ellenberg** und **Dion Gijswijt** bewiesen, dass es bereits genügt, ca. 2.8^n der 3^n Karten zu wählen.

Es gilt also eine viel stärkere Aussage, als wir sie für Teilmengen von $\mathbb{N}$ auch nur erwarten können! Noch besser: Während der Beweis von Bloom und Sisask sehr lang und technisch ist (ihre Arbeit umfasst 95 Seiten), ist der Beweis von Ellenberg-Gijswijt sehr elegant und kurz (4 Seiten). Wir werden ihn auch in der Vorlesung sehen.

1 Elementare Resultate über Summenmengen

1.1 Summen von Restklassen

Wir erinnern uns daran, dass wir für abelsche Gruppen $(G, +)$ und Teilmengen $A, B \subset G$ die *Summenmenge*

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

definiert haben. Für $G = \mathbb{R}$ haben wir

$$|A + B| \geq |A| + |B| - 1$$

für nichtleere endliche Mengen A, B gezeigt. In den Übungen haben wir den Fall $G = \mathbb{R}^2$ betrachtet.

Diese Argumente haben essentiell die Ordnung bzw. die Geometrie dieser Gruppen benutzt. Ist andererseits etwa G endlich, so könnten wir $A = B = G$ wählen und es wäre $A + B = G$ überhaupt nicht größer als A bzw. B.

Etwas allgemeiner: Gilt $|A| + |B| - 1 > |G|$, dann kann die obige Ungleichung sicherlich nicht gelten (einfach, weil G nicht genügend Elemente hat). Das folgende Resultat zeigt, dass dies für $G = \mathbb{F}_p$ auch das einzige Problem ist, welches auftreten kann:

Satz 1.1 (Cauchy-Davenport)**.** *Sei p prim und $A, B \subset \mathbb{F}_p$ nichtleer. Dann gilt*

$$|A + B| \geq \min(p, |A| + |B| - 1).$$

Insbesondere gilt also $A + B = \mathbb{F}_p$, sobald $|A| + |B| > p$ ist (Übung: Warum folgt das auch leicht direkt aus dem Schubfachprinzip?).

Als Anwendung betrachten wir die Menge $\mathcal{Q} = \{x^2 : x \in \mathbb{F}_p\} \subset \mathbb{F}_p$ der quadratischen Reste modulo p. Ist $p > 2$ ungerade, so verrät uns die elementare Zahlentheorie, dass $|\mathcal{Q}| = \frac{p+1}{2}$ ist. Cauchy-Davenport zeigt dann sofort $\mathcal{Q} + \mathcal{Q} = \mathbb{F}_p$, mithin lässt sich jede Restklasse modulo p als Summe von zwei quadratischen Resten schreiben.

Der Satz ist benannt nach dem französischen Mathematiker **Augustin-Louis Cauchy** (1789–1857) und dem britischen Mathematiker **Harold Davenport** (1907–1969). Wie man sieht, ist Davenport erst 50 Jahre nach Cauchys Tod geboren. Warum ist der Satz dennoch nach beiden benannt? Nun, Davenport hat den Satz 1935 bewiesen und erst später festgestellt, dass Cauchy ihn bereits 1813 entdeckt hatte.

Beweis. Wir verwenden (starke) Induktion über $|B|$. Ist $|B| = 1$, so ist die Aussage klar, denn dann ist $|A + B| = |A|$.
Sei nun $|B| \geq 2$ und die Aussage bereits für alle kleineren Mengen (und jeweils alle möglichen Mengen A) bekannt.
Wir beginnen mit einigen vorbereitenden Manövern. Zunächst können wir o.B.d.A. annehmen, dass $0 \in B$ ist. Denn verschieben wir alle Elemente von B um eine Konstante, ändern wir nichts an den Kardinalitäten von B oder der Summenmenge.
Nach Annahme gibt es nun noch ein weiteres Element $b_0 \in B\backslash\{0\}$. Ist $A + b_0 = A$, so ist A invariant unter Verschiebung um b_0, dann aber auch unter Verschiebung um $2b_0$, um $3b_0$ etc. Da $b_0, 2b_0, 3b_0, \ldots$ alle Elemente von $\mathbb{F}_p$ durchläuft, müsste dann allerdings $A = \mathbb{F}_p$ gelten und in diesem Fall gilt unsere Behauptung offensichtlicherweise.
Wir können nun annehmen, dass $A + b_0 \neq A$ gilt. Das bedeutet, dass es ein Element $a \in A$ gibt mit $a + b_0 \notin A$. Wieder können wir o.B.d.A. annehmen, dass $a = 0$ gilt, mithin $0 \in A \cap B$, aber $b_0 \notin A$.
Nun definieren wir die neuen Mengen $A' = A \cup B$ und $B' = A \cap B$. Wir notieren die folgenden Eigenschaften:

- $|A'| + |B'| = |A| + |B|$ (denn wir haben die Elemente von $B\backslash A$ aus B herausgenommen und in A' eingefügt)
- $|B'| < |B|$ (denn $b_0 \notin B'$)
- $A' + B' \subset A + B$ (denn eine Summe $a' + b' \in A' + B'$ hat $a' \in A, b' \in B$ oder $a' \in B, b' \in A$)

Wir haben also Elemente aus B nach A geschoben und die Summenmenge dabei höchstens kleiner gemacht. Aus der Induktionsvoraussetzung (angewandt auf A und B') ergibt sich nun

$$\begin{aligned} |A + B| &\geq |A' + B'| \\ &\geq \min(p, |A'| + |B'| - 1) \\ &= \min(p, |A| + |B| - 1) \end{aligned}$$

wie behauptet. □

Die Transformation $(A, B) \mapsto (A \cup B, A \cap B)$ ist häufig in ähnlichen Induktionsbeweisen nützlich und wird auch *Dyson-Transformation* genannt.

Der Satz von Cauchy-Davenport hat auch ein inverses Theorem, den wir hier ohne Beweis erwähnen (dieser ist nicht wirklich schwer, lediglich etwas technisch).

Satz 1.2 (Vosper)**.** *Wenn $A, B \subset \mathbb{F}_p$ die Bedingung*

$$|A + B| = |A| + |B| - 1 < p$$

sowie $|A|, |B| \geq 2$ erfüllen, dann sind A und B arithmetische Progressionen mit der gleichen Schrittweite.

Als Beispiel (welches in den Übungen nützlich sein wird), stellen wir ohne Beweis fest, dass die oben betrachtete Menge $\mathcal{Q}$ für $p \geq 7$ keine arithmetische Progression ist.

1.2 Das Ruzsa-Kalkül

Nachdem wir nun an einigen Beispielen gesehen haben, wie Resultate über Summenmengen in verschiedenen Gruppen aussehen können und welche Probleme es geben kann, wollen wir nun einige nützliche Resultate sammeln, die für beliebige abelsche Gruppen gültig sind.

Diese lassen sich unter dem Stichwort *Ruzsa-Kalkül* zusammenfassen, benannt nach dem ungarischen Mathematiker **Imre Ruzsa** (geb. 1953).
Aus technischen Gründen müssen wir dafür allerdings zunächst noch ein der Summenmenge verwandtes Konzept einführen.

Definition 1.3. *Für $A, B \subset G$ sei*

$$A - B := A + (-B) = \{a - b : a \in A, b \in B\}$$

die Differenzenmenge und (für A, B endlich und nichtleer)

$$d(A, B) := \frac{|A - B|}{\sqrt{|A| \cdot |B|}}$$

die Ruzsa-Distanz von A und B.

Die Ruzsa-Distanz ist also lediglich eine normalisierte Version der Kardinalität von $|A - B|$. Diese Normalisierung mag zunächst merkwürdig aussehen, ihre Rechtfertigung ergibt sich aber aus dem folgenden Satz:

Satz 1.4 (Dreiecksungleichung von Ruzsa)**.** *Für $A, B, C \subset G$ endlich und nichtleer gilt*

$$d(A, C) \leq d(A, B) \cdot d(B, C)$$

bzw. äquivalent

$$|B| \cdot |A - C| \leq |A - B| \cdot |B - C|.$$

Beachtet man $d(A, B) = d(B, A)$, so verhält sich $\log d(A, B)$ folglich fast wie eine Metrik, lediglich gilt im Allgemeinen nicht $d(A, A) = 1$ bzw. $\log d(A, A) = 0$. Wir bemerken auch, dass $d(A, -A) = \frac{|A+A|}{|A|}$ ist, die Größe der Summenmenge lässt sich also ebenfalls mit der Ruzsa-Distanz messen.

Beweis. Wir zeigen die zweite Version der Ungleichung, indem wir eine injektive Abbildung $\Phi : B \times (A - C) \to (A - B) \times (B - C)$ konstruieren. Zu jedem $d \in A - C$ wählen wir dazu Repräsentanten $a_d \in A, c_d \in C$ mit $d = a_d - c_d$. Diese sind natürlich im Allgemeinen nicht eindeutig, aber unser Argument wird uns erlauben, hier eine vollkommen willkürliche Wahl zu treffen. Nun definieren wir die Abbildung Φ durch $\Phi(b, d) := (a_d - b, b - c_d)$. Diese ist nun sicherlich wohldefiniert.
Um die Injektivität zu beweisen, zeigen wir, dass wir aus einem Paar $(d_1, d_2) \in \text{Bild}(\Phi)$ das Urbild eindeutig rekonstruieren können. Zunächst können wir d als $d_1 + d_2$ rekonstruieren. Kennen wir d, so aber auch a_d und c_d (denn diese Repräsentanten haben wir ja zu Beginn festgelegt) und dann auch $b = a_d - d_1 (= d_2 + c_d)$. □

Dieser Trick mit den Repräsentanten fühlt sich vielleicht beim ersten Lesen etwas „geschummelt“ an, aber es hat tatsächlich alles seine Richtigkeit!

Erinnern wir uns daran, dass unser Ziel der Beweis des Satzes von Freiman ist. Wir wollen also zeigen, dass eine Menge A mit $|A + A| \leq K|A|$ (wobei wir uns K als Konstante und $|A|$ als groß vorstellen) in irgendeiner Weise „wie eine arithmetische Progression“ aussieht.
Als ersten Schritt wollen wir dazu zeigen, dass aus einer Abschätzung $|A + A| \leq K|A|$ der Summenmenge auch Ungleichungen

$$|A + A + A| \ll_K |A|, \quad |A + A + A + A| \ll_K |A|, \quad \ldots \quad |A - A| \ll_K |A|$$

über iterierte Summenmengen und Differenzenmengen folgen.

Hier haben wir die asymptotische Notation (auch Vinogradov- bzw. Landau-Notation genannt)

$$f \ll g \quad \Leftrightarrow \quad f = O(g) \quad \Leftrightarrow \quad \exists C > 0 : |f| \leq C \cdot g.$$

Die Notation $\ll_K$ bedeutet dann, dass die implizite Konstante C auch von K abhängen darf (aber z.B. nicht von A, sonst wäre die Aussage trivialerweise wahr).
Tatsächlich werden wir explizite Werte für die impliziten Konstanten angeben. Die Notation verwenden wir vor allem dann, wenn wir uns nicht für den konkreten Wert dieser Konstanten interessieren.

Satz 1.5 (Plünnecke 1970, Ruzsa 1989, Petridis 2011)**.** *Gegeben seien $A, B \subset G$ endlich und nichtleer mit $|A+B| \leq K \cdot |A|$. Dann gibt es eine nichtleere Teilmenge $X \subset A$ mit*

$$|X + nB| \leq K^n |X|.$$

Hier ist $nB = B + B + \cdots + B$ die n-fache Summenmenge von B. Zunächst ist die Menge X natürlich etwas mysteriös, aber wir können leicht Aussagen folgern, die von X unabhängig sind:

Korollar 1.6. *(i) Unter den gleichen Bedingungen wie oben gilt $|nB| \leq K^n \cdot |A|$. Ist insbesondere $|A+A| \leq K \cdot |A|$, so folgt $|nA| \leq K^n |A|$.*

(ii) Nach Ruzsas Dreiecksungleichung gilt

$$|mB - nB| \leq \frac{|X + mB| \cdot |X + nB|}{|X|} \leq K^{m+n} |X|.$$

Ist insbesondere $|A+A| \leq K \cdot |A|$, so gilt

$$|mA - nA| \leq K^{m+n} |A|.$$

Der ursprüngliche Beweis von **Helmut Plünnecke** war lang und kompliziert und benutzte viel Graphentheorie. Auch der von Ruzsa gefundene etwas vereinfachte Zugang war noch nicht wirklich geeignet, ihn in dieser Vorlesung zu verwenden. Diesen Beweis findet man etwa im Buch von Nathanson oder auch (in etwas abgewandelter Form) bei Tao und Vu. Zum Glück hat **Giorgis Petridis** 2011 einen sehr einfachen und kurzen Beweis gefunden, den wir hier benutzen wollen. Der entscheidende Schritt ist der folgende Hilfssatz.

Lemma 1.7 (Petridis)**.** *Gegeben seien $A, B \subset G$ endlich und nichtleer mit $|A+B| \leq K \cdot |A|$. Dann gibt es eine nichtleere Teilmenge $X \subset A$ mit*

$$|C + X + B| \leq K \cdot |C + X|$$

für alle $C \subset G$.

Beweis von Satz 1.5 mit Lemma 1.7: Wir wählen das X aus dem Lemma.
Mit $C = \{0\}$ folgt $|X + B| \leq K \cdot |X|$.
Mit $C = B$ folgt $|X + 2B| \leq K \cdot |X + B| \leq K^2 \cdot |X|$.
Mit $C = 2B$ folgt $|X + 3B| \leq K \cdot |X + 2B| \leq K^3 \cdot |X|$.
Die Behauptung folgt nun induktiv durch Wiederholen dieses Arguments. □

Beweis von Lemma 1.7: Woher soll das X kommen? Sicherlich muss $|X+B| \leq K \cdot |X|$ gelten. Wir wählen daher $X \subset A$ als diejenige Menge, für die $K' := \frac{|X+B|}{|X|}$ minimal ist. Wegen $\frac{|A+B|}{|A|} \leq K$ gilt sicherlich $K' \leq K$.
Tatsächlich zeigen wir sogar $|C + X + B| \leq K' \cdot |C + X|$.
Dies tun wir per Induktion über $|C|$.
Für $|C| = 1$ ist dies einfach $|X + B| \leq K' \cdot |X|$, was nach Definition von X bzw. K' gilt.
Sei nun $|C| \geq 2$ und $c_0 \in C, C' = C \backslash \{c_0\}$.
Um die Summen zu untersuchen, die durch das Hinzufügen von c_0 hinzukommen, definieren wir X_{c_0} als die maximale Teilmenge von X, für die $C' + X$ und $c_0 + X_{c_0}$ disjunkt sind. Dann ist

$$C + X = (C' + X) \cup (c_0 + X) = (C' + X) \sqcup (c_0 + X_{c_0})$$

sowie

$$C + X + B \subset (C' + X + B) \cup (c_0 + X + B) \backslash (c_0 + (X \backslash X_{c_0}) + B)$$

und damit

$$\begin{aligned} |C + X + B| &\leq |C' + X + B| + |X + B| - |X \backslash X_{c_0} + B| \\ &\leq K' \cdot |C' + X| + K' \cdot |X| - K' \cdot |X \backslash X_{c_0}| \\ &= K' \cdot |C + X|, \end{aligned}$$

wobei wir bei der Abschätzung des ersten Summanden die Induktionsvoraussetzung für C', für den zweiten die Definition von X bzw. K' und für den dritten die Minimalität von K' benutzt haben. □

Wir haben gezeigt, dass für beliebige abelsche Gruppen aus $|A + A| \leq K \cdot |A|$ schon $|A + A + A| \ll_K |A|$ folgt. Dies ist für nichtabelsche Gruppen falsch:
Dazu betrachten wir eine Gruppe G, eine Untergruppe $H \subset G$ und ein Element $g \in G$, welches H nicht zentralisiert, also mit $gHg^{-1} \neq H$. Für $A = H \cup \{g\}$ ist dann $A \cdot A \subset H \cup Hg \cup gH \cup \{g^2\}$, also $|A^2| \leq 3|A|$.
Andererseits enthält $A \cdot A \cdot A$ die Menge HgH, die so groß wie $|H|^2 \approx |A|^2$ sein kann.
Tatsächlich lässt sich aber zeigen, dass für nichtkommutative Gruppen immer noch gilt: Ist A^3 klein, dann auch $A^4, A^5, \ldots$.

1.3 Freiman-Homomorphismen und das Einbettungslemma

Wir haben bereits in der Einführung gesehen, dass es manchmal einfacher ist, in $\mathbb{F}_p$ statt in $\mathbb{Z}$ zu arbeiten, auch wenn man auf den ersten Blick vielleicht exakt das Gegenteil erwarten würde.
Ein wesentlicher Grund dafür ist, dass einige der späteren Argumente Fourier-Analysis benutzen werden und diese für endliche Gruppen eine wesentlich einfachere Struktur besitzt.
Da der Satz von Freiman allerdings von Teilmengen von $\mathbb{Z}$ handelt, müssen wir uns zunächst mit der Frage beschäftigen, ob sich Aussagen über Teilmengen von $\mathbb{Z}$ in geeigneter Weise in Aussagen über Teilmengen von $\mathbb{F}_p$ übersetzen lassen.

Allgemeiner wollen wir uns fragen, wann sich zwei Teilmengen $A \subset G$ und $B \subset H$ verschiedener Gruppen im wesentlichen ähnlich bzgl. ihrer Summenmenge verhalten.
Die einfachste Situation wäre natürlich, wenn wir einen Gruppenhomomorphismus $\varphi : G \to H$ hätten, der A auf B abbildet. Allerdings ist dies im Allgemeinen eine zu starke Voraussetzung, denn z.B. gibt es keine nichttrivialen Gruppenhomomorphismen zwischen $\mathbb{Z}$ und $\mathbb{F}_p$.

Das für uns richtige Konzept ist dagegen das eines *Freiman-Homomorphismus*, welches wir nun definieren:

Definition 1.8. *Gegeben seien zwei abelsche Gruppen G und H sowie Teilmengen $A \subset G$ und $B \subset H$. Sei $h \geq 2$ eine natürliche Zahl. Ein* ***Freiman-Homomorphismus der Ordnung*** *h ist eine Abbildung $\varphi : A \to B$ mit der Eigenschaft, dass für $a_1, \dots, a_h \in A$ und $a'_1, \dots, a'_h \in A$ aus*

$$a_1 + \cdots + a_h = a'_1 + \cdots + a'_h$$

auch

$$\varphi(a_1) + \cdots + \varphi(a_h) = \varphi(a'_1) + \cdots + \varphi(a'_h)$$

folgt. Analog definieren wir einen ***Freiman-Isomorphismus der Ordnung*** *h. Gibt es einen solchen Isomorphismus, so heißen A und B* ***Freiman-isomorph der Ordnung*** *h.*

Beispiel:

(i) Jeder Gruppenhomomorphismus ist ein Freiman-Homomorphismus von beliebiger Ordnung. Ein Freiman-Homomorphismus von Ordnung h ist auch von Ordnung h' für $h' \leq h$.

(ii) Jede arithmetische Progression $\{a, a+d, \dots, a+(k-1)d\} \subset \mathbb{Z}$ der Länge k ist Freiman-isomorph zu $\{0, 1, 2, \dots, k-1\} \subset \mathbb{Z}$.

(iii) Es sei $A = \{0, 1, 2\} \subset \mathbb{Z}$ und $B = \{0, 1, 2\} \subset \mathbb{F}_3$. Dann ist die Identität $\varphi : A \to B$ ein Freiman-Homomorphismus von beliebiger Ordnung, aber kein Freiman-Isomorphismus, denn es gilt $\varphi(1) + \varphi(2) = \varphi(0) + \varphi(0)$, obwohl nicht $1 + 2 = 0 + 0$ (in $\mathbb{Z}$) gilt.

(iv) Analog ist für $A = \{0, 1, 2\} \subset \mathbb{F}_5$ und $B = \{0, 1, 2\} \subset \mathbb{Z}$ die Identität $\varphi : A \to B$ ein Freiman-Homomorphismus von Ordnung 2, aber nicht von Ordnung 3.

Freiman-Isomorphismen der Ordnung h erhalten also die Struktur der h-fachen Summenmenge von A, insbesondere ist dann $|hA| = |hB|$.
Tatsächlich können wir aber auch iterierte Summen- und Differenzmengen kontrollieren:

Lemma 1.9. *Seien G, H abelsche Gruppen und $A \subset G, B \subset H$ endlich und Freiman-isomorph zur Ordnung $h(k+l)$. Dann sind die Mengen $kA - lA$ und $kB - lB$ Freiman-isomorph zur Ordnung h.*

Beweis. Es sei $\varphi : A \to B$ die entsprechende Abbildung. Dann können wir zunächst $\widetilde{\varphi} : kA - lA \to kB - lB$ definieren als

$$\widetilde{\varphi}(a_1 + \cdots + a_k - a_{k+1} - \cdots - a_{k+l}) := \varphi(a_1) + \cdots + \varphi(a_k) - \varphi(a_{k+1}) - \cdots - \varphi(a_{k+l}).$$

Dies ist wohldefiniert, denn ist

$$a_1 + \cdots + a_k - a_{k+1} - \cdots - a_{k+l} = a'_1 + \cdots + a'_k - a'_{k+1} - \cdots - a'_{k+l},$$

so gilt

$$a_1 + \cdots + a_k + a'_{k+1} + \cdots + a'_{k+l} = a'_1 + \cdots + a'_k + a_{k+1} + \cdots + a_{k+l}$$

und damit nach Annahme auch

$$\varphi(a_1) + \cdots + \varphi(a_k) + \varphi(a'_{k+1}) + \cdots + \varphi(a'_{k+l}) = \varphi(a'_1) + \cdots + \varphi(a'_k) + \varphi(a_{k+1}) + \cdots + \varphi(a_{k+l}),$$

was wiederum nach Umformung äquivalent zur Wohldefiniertheit von $\widetilde{\varphi}$ ist.
Vollkommen analog sieht man, dass $\widetilde{\varphi}$ sogar ein Freiman-Homomorphismus von Ordnung h ist. Argumentiert man analog mit der Umkehrabbildung φ^{-1}, erhält man auch die Isomorphie. □

Als nächstes zeigen wir, dass es einen guten Weg gibt, Teilmengen von $\mathbb{Z}$ Freiman-isomorph als Teilmenge von Restklassen zu realisieren.

Satz 1.10 (Einbettungslemma von Ruzsa)**.** *Sei $A \subset \mathbb{Z}$ endlich und $h \geq 2$ sowie $N > 2|hA-hA|$. Dann gibt es eine Menge $A' \subset A$ mit $|A'| \geq \frac{|A|}{h}$, die Freiman-isomorph von Ordnung h zu einer Teilmenge von $\mathbb{Z}/N\mathbb{Z}$ ist.*

Insbesondere kann im Fall $|A + A| \leq K|A|$ wegen der Plünnecke-Ungleichung $N \ll_K |A|$ gewählt werden, wir können also A' „dicht“ in $\mathbb{Z}/N\mathbb{Z}$ einbetten!
Zum Aufwärmen zeigen wir zunächst einen einfacheren Satz von Erdős, der ein ähnliches Argument benutzt. Eine Menge A heißt *summenfrei*, falls es keine $a_1, a_2, a_3 \in A$ gibt mit $a_1 + a_2 = a_3$.

Satz 1.11 (Erdős)**.** *Sei $A \subset \mathbb{Z}$, dann gibt es eine summenfreie Menge $A' \subset A$ mit $|A'| > |A|/3$.*

Beweis. Wähle eine große Primzahl $p \equiv 2 \mod 3$. Zu jedem $1 \leq q \leq p-1$ sei A_q die Menge der $a \in A$, für die qa kongruent zu einer Zahl in $\left(\frac{p}{3}, \frac{2p}{3}\right)$ modulo p ist. Jedes $a \in A$ ist für mehr als $\frac{1}{3}$ der Werte von q in A_q enthalten. Damit gibt es ein q mit $|A_q| > \frac{|A|}{3}$ und diese Menge ist summenfrei, da aus $a + b = c$ schon $qa + qb = qc$ folgt, was modulo p unmöglich ist. □

Durch Betrachten des größten Elements von A sieht man leicht, dass $A = \{1, 2, \ldots, N\}$ keine summenfreie Menge mit mehr als $(N + 1)/2$ Elementen enthält. Tatsächlich ist die Konstante $1/3$ im Satz von Erdős aber sogar (asymptotisch) optimal, wie 2014 bewiesen wurde.

Beweis von Satz 1.10: Sei zunächst p eine hinreichend große Primzahl, sodass die Einbettung von A in $\mathbb{F}_p$ ein Freiman-Isomorphismus ist. Zu einem zunächst beliebigen $1 \leq q \leq p-1$ betrachten wir nun die Abbildung $\varphi : A \to \mathbb{Z}, a \mapsto (aq)_p \in \{0, 1, 2, \ldots, p-1\}$, die a auf den Repräsentanten von aq modulo p abbildet.
Für jedes q gibt es nun nach dem Schubfachprinzip sicherlich eine Menge $A_q \subset A$ mit $|A_q| \geq \frac{|A|}{h}$, für die alle $(qa)_p$ im gleichen Intervall $\left(\frac{jp}{h}, \frac{(j+1)p}{h}\right)$ liegen.
Dies erzwingt, dass φ eingeschränkt auf A_q ein Freiman-Isomorphismus ist, denn aus $a_1 + \cdots + a_h = a'_1 + \cdots + a'_h$ folgt sicherlich

$$p \mid (qa_1)_p + \cdots + (qa_h)_p - (qa'_1)_p - \cdots - (qa'_h)_p$$

und nach Konstruktion ist die rechte Seite im Betrag kleiner als p, muss also gleich Null sein. Die Umkehrung ist klar.
Wir wollen diese Abbildung nun noch mit der Einbettung $\{0, 1, \ldots, p-1\} \to \mathbb{Z}/N\mathbb{Z}$ verknüpfen. Diese Verkettung ist sicherlich immer noch ein Freiman-Homomorphismus, aber wann ist es ein Isomorphismus? Wir zeigen, dass dies für mindestens eine Wahl von q klappt.
Angenommen, es gibt $a_1, \ldots, a_h, a'_1, \ldots, a'_h \in A_q$ mit

$$N \mid (qa_1)_p + \cdots + (qa_h)_p - (qa'_1)_p - \cdots - (qa'_h)_p.$$

Sicherlich ist die rechte Seite weniger als p im Betrag, also muss sie von der Form kN mit $|k| < \frac{p}{N}$ sein. Andererseits ist sie kongruent zu $q(a_1 + \cdots + a_h - a'_1 - \cdots - a'_h)$ modulo p, also zu qD für ein $D \in hA - hA$.
Für jede Wahl von k und D ist damit q eindeutig bestimmt. Die Anzahl der Werte von q, bei denen wir keinen Freiman-Isomorphismus erhalten, ist also maximal $\frac{2p}{N} \cdot |hA - hA|$ und damit für große p sicherlich weniger als $p - 1$.

2 Der Satz von Freiman

Wir wollen zunächst den Satz von Freiman ordentlich formulieren. Grob gesagt soll er aussagen, dass aus der Annahme $|A+A| \leq K \cdot |A|$ schon folgt, dass A „wie eine arithmetische Progression“ aussieht. Eine konkrete Illustration dieses Prinzips liefert der sogenannte „$3k-4$-Satz von Freiman“: Ist $|A+A| \leq 3|A|-4$, so gibt es eine arithmetische Progression P mit $|P| \leq 2|A|-3$, die A enthält.

Ein instruktives Beispiel ist die Menge $A = \{1, 2, \ldots, N-1, M\}$ für natürliche Zahlen $N \leq M$. Es ist $|A| = N$ und für $M \geq 2N-2$ gilt $|A+A| = 3N-3$. Eine arithmetische Progression, die A enthält, müsste allerdings mindestens Länge M haben, was viel größer als $|A| = N$ sein kann.

Dieses Beispiel zeigt, dass wir für eine Verallgemeinerung des $3k-4$-Satzes auch eine Verallgemeinerung unseres Begriffs arithmetischer Progressionen benötigen:

Definition 2.1. *Sei G eine abelsche Gruppe. Eine **verallgemeinerte arithmetische Progression (VAP)** von Rang d in G ist eine Menge der Form*

$$\{a + n_1 v_1 + \cdots + n_d v_d : 0 \leq n_i < N_i (1 \leq i \leq d)\}$$

*für $N_1, \ldots, N_d \geq 1$ und $a, v_1, \ldots, v_d \in G$. Die VAP heißt **echt**, falls $|G| = N_1 \cdot \ldots \cdot N_d$ gilt.*

Satz 2.2 (Freiman)**.** *Ist $A \subset \mathbb{Z}$ endlich mit $|A+A| \leq K \cdot |A|$, dann gibt es eine verallgemeinerte arithmetische Progression P von Rang $d \ll_K 1$ mit $|P| \ll_K |A|$ und $A \subset P$.*

In unserem Beispiel oben könnten wir einfach $P = \{n_1 + n_2 \cdot M : 0 \leq n_1 < N-1, 0 \leq n_2 < 2\}$ als VAP von Rang 2 mit $|P| = 2(N-1) \ll |A|$ wählen.
Die letzte entscheidende Zutat des Beweises werden wir im nächsten Abschnitt beweisen:

Lemma 2.3 (Bogolyubov-Ruzsa)**.** *Sei p prim und $R \subset \mathbb{Z}/p\mathbb{Z}$ nichtleer mit $|R| = \lambda p$. Dann gibt es eine VAP $Q \subset 2R - 2R$ von Rang $n \leq \lambda^{-2}$ mit $|Q| \gg_\lambda p$.*

Zunächst zeigen wir, wie sich daraus der Satz von Freiman ergibt:

Beweis des Satzes von Freiman. Sei $A \subset \mathbb{Z}$ mit $|A+A| \leq K|A|$. Nach dem Einbettungslemma zusammen mit dem Bertrandschen Postulat finden wir eine Primzahl $p \ll_K |8A - 8A| \ll_K |A|$ (wegen der Plünnecke-Ungleichung) und eine Menge $A' \subset A$ mit $|A'| \gg_K |A|$, die Freiman-isomorph von Ordnung 8 zu einer Teilmenge $R \subset \mathbb{Z}/p\mathbb{Z}$ ist. Dann sind $2A' - 2A'$ und $2R - 2R$ Freiman-isomorph von Ordnung 2. Nach dem Lemma von Bogolyubov-Ruzsa enthält $2R - 2R$ und damit auch $2A' - 2A'$ eine VAP Q_1 mit $|Q_1| \gg_K |A|$ und Rang $\ll_K 1$.
Diese wollen wir nun benutzen, um eine kleine VAP zu konstruieren, die A enthält. Dazu sei $X \subset A$ maximal, sodass die Mengen $Q_1 + x$ mit $x \in X$ paarweise disjunkt sind. Wegen $Q_1 + X \subset 3A - 2A$ ist dann

$$|X| = \frac{|Q_1 + X|}{|Q_1|} \leq \frac{|3A - 2A|}{|Q_1|} \ll_K 1$$

nach der Plünnecke-Ungleichung. Damit finden wir sicherlich eine VAP Q_2 mit $|Q_2| \ll_K 1$ und Rang $\ll_K 1$, die X enthält. Wir setzen nun $Q = Q_1 - Q_1 + Q_2$. Dies ist sicherlich eine VAP von Rang $\ll_K 1$ mit $|Q| \ll |A|$. Nach Voraussetzung gibt es für jedes $a \in A$ ein $x \in X \subset Q_2$ und $q, q' \in Q_1$ mit $a + q' = x + q$, also $a = x + q - q' \in Q$, folglich gilt $A \subset Q$ und wir haben die gewünschte VAP gefunden. □

Der Beweis des Bogolyubov-Ruzsa-Lemmas erfolgt in zwei Schritten:

(1) Mithilfe von **Fourier-Analysis** zeigen wir, dass für eine solche große Menge A von Restklassen die Menge $2A - 2A$ eine große *Bohrmenge* enthält.

(2) Dann zeigen wir mithilfe der **Geometrie der Zahlen**, dass eine solche Bohrmenge immer eine große (verallgemeinerte) arithmetische Progression enthält.

2.1 Fourier-Analysis auf $\mathbb{Z}/m\mathbb{Z}$

Wir führen zunächst die Fourier-Transformation auf $G = \mathbb{Z}/m\mathbb{Z}$ ein. Dazu sei $e(x) = e^{2\pi i x}$.

Definition 2.4. *Die **Charaktere** von G sind gegeben durch*

$$\chi_r : G \to \mathbb{C}, g \mapsto e(rg/m)$$

*für jedes $r \in G$. Die konstante Abbildung χ_0 heißt der **triviale Charakter**.*
*Für eine Funktion $f : G \to \mathbb{C}$ definieren wir ihre **Fourier-Transformierte** $\widehat{f} : G \to \mathbb{C}$ als*

$$\widehat{f}(r) = \sum_{g \in G} \chi_r(g) f(g).$$

Für eine Teilmenge $A \subset G$ schreiben wir $\widehat{A}$ für die Fourier-Transformierte der Indikatorfunktion von A, also

$$\widehat{A}(r) = \sum_{a \in A} \chi_r(a).$$

Offensichtlich gilt $|\widehat{A}(r)| \leq |A| = |\widehat{A}(0)|$. Wir sammeln nun einige wichtige Eigenschaften:

Lemma 2.5 (Orthogonalitätsrelationen)**.** *Es gilt*

$$\sum_{g \in G} \chi_r(g) = \begin{cases} m & r = 0 \\ 0 & r \neq 0 \end{cases}$$

und analog

$$\sum_{r \in G} \chi_r(g) = \begin{cases} m & g = 0 \\ 0 & g \neq 0 \end{cases}.$$

Beweis. Wegen der Symmetrie $\chi_r(g) = \chi_g(r)$ genügt es, die erste Aussage zu zeigen. Für $r = 0$ ist die Aussage klar. Für $r \neq 0$ ist die Summe links

$$\sum_{g=0}^{m-1} \left(e^{2\pi i r/m}\right)^g = \frac{e^{2\pi i m r/m} - 1}{e^{2\pi i r/m} - 1} = 0$$

nach der geometrischen Summenformel. □

Damit lässt sich nun leicht die folgende Identität zeigen:

Lemma 2.6 (Parseval)**.** *Für $f : G \to \mathbb{C}$ gilt*

$$\sum_{r \in G} \left|\widehat{f}(r)\right|^2 = m \sum_{g \in G} |f(g)|^2.$$

Insbesondere gilt also $\sum_{r\in G}\left|\widehat{A}(r)\right|^2 = m\cdot|A|$.

Beweis. Es ist

$$\begin{aligned}\sum_{r\in G}\left|\widehat{f}(r)\right|^2 &= \sum_{r\in G}\sum_{g_1,g_2\in G}\chi_r(g_1)f(g_1)\overline{\chi_r(g_2)f(g_2)}\\ &= \sum_{g_1,g_2\in G} f(g_1)\overline{f(g_2)}\sum_{r\in G}\chi_r(g_1-g_2)\\ &= m\sum_{g\in G}|f(g)|^2,\end{aligned}$$

wobei wir im letzten Schritt die Orthogonalität benutzt haben, um zu sehen, dass die innere Summe für $g_1\neq g_2$ verschwindet und für $g_1=g_2$ genau m ist. □

Es gilt auch die Inversionsformel

$$f(g) = \frac{1}{m}\sum_{r\in G}\widehat{f}(-r)\chi_r(g),$$

die wir allerdings nicht weiter verwenden werden (sie lässt sich ebenfalls direkt nachrechnen). Die Fourier-Transformation ist also fast zu sich selbst invers (bis auf Vorzeichen und Normalisierung, bei der man ohnehin immer genau hinschauen muss, weil sie sich von Quelle zu Quelle unterscheidet).
Man kann die Werte von $\widehat{f}(r)$ also als *Fourierkoeffizienten* interpretieren: Die Charaktere χ_r bilden eine besonders einfache Basis der Funktionen $f: G\to\mathbb{C}$. Jede Funktion f lässt sich somit als eine Linearkombination dieser Charaktere schreiben und die Werte von $\widehat{f}(r)$ geben exakt an, welche Koeffizienten in dieser Linearkombination auftreten.

Wir können nun den ersten Schritt der oben beschriebenen Strategie umsetzen. Für eine reelle Zahl x sei dazu $\|x\|$ der Abstand von x zur nächsten ganzen Zahl.

Definition 2.7. *Für Restklassen $r_1,\dots,r_n\in\mathbb{Z}/m\mathbb{Z}$ und eine reelle Zahl $\varepsilon>0$ ist*

$$B(r_1,\dots,r_n;\varepsilon) := \{g\in\mathbb{Z}/m\mathbb{Z} : \|gr_i/m\|\leq\varepsilon, i=1,2,\dots,n\}$$

*die **Bohrmenge** mit Frequenzen $r_1,\dots,r_n$ und Breite ε.*

Insbesondere ist also $B(0;\varepsilon)=B(r_1,\dots,r_n;1/2)=\mathbb{Z}/m\mathbb{Z}$.
Man kann sich leicht davon überzeugen, dass im Fall $n=1$ die Bohrmenge $B(r;\varepsilon)$ immer eine arithmetische Progression ist. Allgemeinere Bohrmengen sind also auch eine Art Verallgemeinerung von arithmetischen Progressionen. Allerdings können sie (im Gegensatz zu arithmetischen Progressionen) sehr große Summenmengen haben.

Satz 2.8 (Bogolyubov)**.** *Sei $m\geq 2$ und $A\subset\mathbb{Z}/m\mathbb{Z}$ nichtleer sowie $\lambda\in(0,1]$ mit $|A|=\lambda m$. Dann gibt es ein $n\leq\lambda^{-2}$ und paarweise verschiedene Restklassen $r_1,\dots,r_n\in\mathbb{Z}/m\mathbb{Z}$ mit*

$$B(r_1,\dots,r_n;1/4)\subset 2A-2A$$

Wir können nicht erwarten, dass A selbst eine Bohrmenge (und damit eine große VAP) enthält: Ein Beispiel ist eine AP, in der einige Elemente fehlen. In der Menge $2A-2A$ verschwinden diese Lücken.

Beweis. Es ist

$$\frac{1}{m}\sum_{r\in G}|\widehat{A}(r)|^4\chi_r(g) = \frac{1}{m}\sum_{r\in G, a_1,a_2,a_3,a_4\in A} e(r(g-a_1-a_2+a_3+a_4)/m)$$

gleich der Anzahl an Lösungen von $g = a_1+a_2-a_3-a_4$, insbesondere also genau dann positiv, wenn $g \in 2A-2A$. Da die Lösungsanzahl reell ist, ist auch die linke Seite reell und damit gleich ihrem Realteil, mithin ist genau dann $g \in 2A-2A$, wenn

$$\sum_{r\in G}|\widehat{A}(r)|^4\cos(2\pi gr/m)$$

positiv ist. Für ein noch zu bestimmendes $R = \{r_1,\dots,r_n\} \subset G$ mit $r_1 = 0$ und $g \in B(r_1,\dots,r_n;1/4)$ ist dann $\|gr_i/m\| \le 1/4$ für alle i und damit $\cos(2\pi r_i g/m) \ge 0$. Alle Terme in der obigen Gleichung mit $r \in R$ haben also einen nicht-negativen Beitrag, der Term $r = 0$ trägt $|A|^4$ bei. Damit genügt es R so zu wählen, dass

$$\sum_{r\notin R}|\widehat{A}(r)|^4 < |A|^4$$

gilt um $B(r_1,\dots,r_n;1/4) \subset 2A-2A$ zu folgern. Wählen wir

$$R = \{r \in G : |\widehat{A}(r)| \ge M\}$$

für einen noch zu bestimmenden Parameter M, so folgt wegen der Parseval-Identität

$$\sum_{r\in G}|\widehat{A}(r)|^2 = m|A| = \lambda^{-1}|A|^2$$

schon

$$\sum_{r\notin R}|\widehat{A}(r)|^4 \le M^2\sum_{r\notin R}|\widehat{A}(r)|^2 < M^2\sum_{r\in G}|\widehat{A}(r)|^2 = M^2\lambda^{-1}|A|^2$$

Wir können also $M = \sqrt{\lambda}|A|$ wählen. Wegen der Parseval-Identität kann es gleichzeitig nicht zu viele große Fourier-Koeffizienten geben. Genauer gilt

$$\lambda^{-1}|A|^2 = \sum_{r\in G}|\widehat{A}(r)|^2 \ge \sum_{r\in R}|\widehat{A}(r)|^2 \ge n\lambda|A|^2$$

und damit $|R| = n \le \lambda^{-2}$ wie behauptet. □

2.2 Geometrie der Zahlen

Um den Beweis des Bogolyubov-Ruzsa-Lemmas und damit des Satzes von Freiman abzuschließen, benötigen wir noch einen letzten Schritt, nämlich die folgende Aussage:

Lemma 2.9. *Sei $p \ge 2$ prim und $r_1,\dots,r_n \in \mathbb{Z}/p\mathbb{Z}$. Dann gibt es eine n-dimensionale VAP $Q \subset \mathbb{Z}/p\mathbb{Z}$ mit $Q \subset B(r_1,\dots,r_n;1/4)$ sowie $|Q| \gg_n p$.*

(Die Aussage stimmt auch, wenn wir $1/4$ durch einen beliebigen Wert $\delta \in (0,1/2)$ ersetzen, allerdings wird dann die implizite Konstante in der Abschätzung für $|Q|$ auch von δ abhängen.)
Für den Beweis dieses letzten Lemmas müssen wir zunächst etwas ausholen und in die *Geometrie der Zahlen* eintauchen, die von Hermann Minkowski um 1900 ausgearbeitet wurde. Grob gesagt geht es dabei um die Untersuchung von Punkten, die einerseits eine arithmetische Bedingung (*Gitter*) und andererseits eine Ungleichungsbedingung erfüllen. Zunächst führen wir einige wichtige Begriffe ein:

Definition 2.10. *Ein* ***Gitter*** $\Lambda \subset \mathbb{R}^m$ *von* ***Rang*** n *ist eine diskrete Untergruppe*

$$\Lambda = \{a_1 v_1 + \cdots + a_n v_n : a_i \in \mathbb{Z}\} \subset \mathbb{R}^m$$

für linear unabhängige Vektoren $v_1, \ldots, v_n \in \mathbb{R}^m$.
Eine solche Menge $\{v_1, \ldots, v_n\}$ *von Vektoren heißt* ***Basis*** *des Gitters. Der* ***Fundamentalbereich*** *von* Λ *bzgl. dieser Basis ist das Parallelepiped*

$$F = \{a_1 v_1 + \cdots + a_n v_n : 0 \leq a_i < 1\} \subset \mathbb{R}^m.$$

Ist $n = m$, *so nennen wir* Λ *ein* ***volles Gitter*** *und nennen*

$$\det(\Lambda) := |\det(v_1, \ldots, v_n)| = \mathrm{vol}(F)$$

die ***Determinante*** *oder das* ***Kovolumen*** *des Gitters* Λ.

Man beachte, dass der Fundamentalbereich von der Wahl der Basis abhängt, die Determinante des Gitters aber nicht (denn die Basiswechselmatrix muss in $\mathrm{SL}_k(\mathbb{Z})$ liegen, also Determinante ± 1 haben).

Beispiel:

(i) Das Gitter $\Lambda = \mathbb{Z}^n \subset \mathbb{R}^n$ hat die Standardbasis $(e_1, \ldots, e_n)$ und Determinante 1. Der Fundamentalbereich bzgl. dieser Basis ist ein Einheitswürfel.

(ii) Das Gitter $\Lambda = \mathbb{Z}^2 \subset \mathbb{R}^2$ hat neben der Basis $\{(1,0),(0,1)\}$ auch die Basis $\{(1,0),(1,1)\}$. Hier ist der Fundamentalbereich ein Parallelogramm, aber immer noch mit Fläche 1. Tatsächlich liefert jede Menge $\{(a,b),(c,d)\}$ von Vektoren in $\mathbb{Z}^2$ mit $ad - bc = \pm 1$ eine Basis.

(iii) Sind zwei Gitter $\Lambda_1 \subset \Lambda_2 \subset V$ gegeben, so gilt $\det(\Lambda_2) \mid \det(\Lambda_1)$, denn eine Basis von Λ_1 hat eine ganzzahlige Koordinatendarstellung bzgl. einer Basis von Λ_2, die Determinante des Basiswechsels ist also ganzzahlig (aber nicht unbedingt ± 1, da die Inverse nicht ganzzahlig sein muss).

(iv) Für gegebene $(a,b,c) \in \mathbb{Z}^3 \backslash \{(0,0,0)\}$ ist die Menge $\{(x,y,z) \in \mathbb{Z}^3 : ax+by+cz = 0\} \subset \mathbb{R}^3$ ein Gitter von Rang 2.

(v) Für gegebene $(a,b,c) \in \mathbb{Z}^3$ mit $c \neq 0$ ist $\{(x,y) \in \mathbb{Z}^2 : c \mid ax + by\} \subset \mathbb{R}^2$ ein volles Gitter. Ist $\mathrm{ggT}(a,b,c) = 1$, so ist die Determinante $|c|$.

Das erste wichtige Resultat sagt uns, unter welchen Bedingungen wir in bestimmten Mengen Gitterpunkte erwarten können:

Definition 2.11. *Eine Menge* $V \subset \mathbb{R}^m$ *ist* ***(zentral-)symmetrisch****, falls zu* $v \in V$ *stets auch* $-v \in V$ *gilt.*
Die Menge $V \subset \mathbb{R}^m$ *heißt* ***konvex****, falls zu* $v_1, v_2 \in V$ *und* $t \in [0,1]$ *auch* $tv_1 + (1-t)v_2 \in V$ *gilt.*

Satz 2.12 (Minkowski I)**.** *Sei* $\Lambda \subset \mathbb{R}^n$ *ein volles Gitter und* $V \subset \mathbb{R}^n$ *konvex und zentralsymmetrisch mit* $\mathrm{vol}(V) > 2^n \det(\Lambda)$. *Dann enthält* V *einen Punkt* $v \in \Lambda \backslash \{0\}$.

Beweis. Wir betrachten die reskalierte Menge $\frac{1}{2}V = \{\frac{v}{2} : v \in V\}$ mit Volumen $\frac{\mathrm{vol}(V)}{2^n}$. Wir zeigen, dass die Verschiebungen $\frac{1}{2}V + \gamma$ für verschiedene $\gamma \in \Lambda$ nicht paarweise disjunkt sein können. Dann folgt, dass es $v_1, v_2 \in V$ gibt mit $\frac{v_1 - v_2}{2} = \gamma_2 - \gamma_1 \in \Lambda\backslash\{0\}$ und wegen der Konvexität und der Zentralsymmetrie ist auch $\frac{v_1 - v_2}{2} = \frac{v_1 + (-v_2)}{2} \in V$.
Nehmen wir nun also an, die Verschiebungen wären paarweise disjunkt. Für einen großen Parameter R betrachten wir nun einen Würfel mit Seitenlänge R. Dieser enthält asymptotisch etwa $\frac{R^n}{\det(\Lambda)}$ viele Gitterpunkte von Λ. Wenn wir für jeden dieser Gitterpunkte die Menge $\frac{1}{2}V + \lambda$ betrachten, liegen diese alle in einem Würfel mit Seitenlänge $R + c$, wobei c nur von V abhängt. Wären sie disjunkt, hätten sie allein schon ein Volumen von $\frac{R^n}{\det(\Lambda)} \cdot \frac{\mathrm{vol}(V)}{2^n}$, was kleiner als $(R+c)^n$ sein muss. Damit folgt

$$\frac{\mathrm{vol}(V)}{2^n \det(\Lambda)} < \frac{(R+c)^n}{R^n}$$

und damit $\mathrm{vol}(V) \leq 2^n \det(\Lambda)$, da die rechte Seite für große R beliebig nah an 1 kommt. □

Für unsere Anwendung wird es allerdings entscheidend sein, nicht nur die Existenz eines Gitterpunkts in einer bestimmten Menge zu zeigen, sondern ihre Anzahl und Struktur relativ präzise zu bestimmen. Dazu benötigen wir noch ein weiteres Konzept:

Definition 2.13. *Sei $\Lambda \subset \mathbb{R}^n$ ein volles Gitter. Die **sukzessiven Minima** $\lambda_1, \ldots, \lambda_n$ von Λ werden durch die folgende Eigenschaft charakterisiert: Es ist λ_k die kleinste reelle Zahl, für die es k linear unabhängige Vektoren $v_1, \ldots, v_k \in \Lambda$ gibt mit $|v_1|, \ldots, |v_k| \leq \lambda_k$.*

Äquivalent können wir sie iterativ konstruieren: Wir wählen einen Vektor $v_1 \in \Lambda\backslash\{0\}$ minimaler Länge $|v_1| = \lambda_1$. Dann wählen wir einen Vektor v_2 im Komplement von $\mathrm{span}(v_1)$ mit minimaler Länge $|v_2| = \lambda_2$. Dann wählen wir einen Vektor v_3 im Komplement von $\mathrm{span}(v_1, v_2)$ mit minimaler Länge $|v_3| = \lambda_3$ etc.

Auf diese Weise erhalten wir eine Basis $\{v_1, \ldots, v_n\}$ des Vektorraums mit $|v_i| = \lambda_i$. Diese ist nicht unbedingt eindeutig, die äquivalente Charakterisierung in der Definition zeigt aber, dass die Werte von λ_i unabhängig von dieser Wahl der Basis sind.

Obacht: Im Allgemeinen ist die so konstruierte Basis keine Basis des *Gitters* Λ! Das kleinste Gegenbeispiel tritt allerdings für $n = 5$ auf... Man kann allgemein zeigen, dass es immer eine Basis $\{w_1, \ldots, w_n\}$ von Λ gibt mit $\lambda_i \ll_n |w_i| \ll_n \lambda_i$, was für Anwendungen typischerweise gut genug ist.

Sicherlich gilt $0 < \lambda_1 \leq \cdots \leq \lambda_n$. Aus dem ersten Satz von Minkowski folgt $\lambda_1^n \ll \det(\Lambda)$.

Der zweite Satz von Minkowski macht eine wesentlich stärkere Aussage:

Satz 2.14 (Minkowski II)**.** *Sei $\Lambda \subset \mathbb{R}^n$ ein volles Gitter. Dann gilt*

$$\det(\Lambda) \leq \lambda_1 \ldots \lambda_n \ll_n \det(\Lambda).$$

Beweis. Zunächst ist die untere Schranke klar, denn die zu den sukzessiven Minima gehörenden minimalen Vektoren $v_i \in \Lambda$ mit $|v_i| = \lambda_i$ bilden die Basis eines Untergitters von Λ, ihre Determinante ist also ein ganzzahliges Vielfaches von $\det(\Lambda)$, mithin mindestens $\det(\Lambda)$.

Für die obere Schranke betrachten wir zunächst eine orthogonale Transformation (Gram-Schmidt), die die Basis $(v_1, \dots, v_n)$ in eine untere Dreiecksform bringt, also $v_1 = (x_{11}, 0, 0, \dots, 0)$, $v_2 = (x_{12}, x_{22}, 0, 0, \dots, 0)$ etc. Wir betrachten nun das Ellipsoid

$$E = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \frac{x_1^2}{\lambda_1^2} + \dots + \frac{x_n^2}{\lambda_n^2} < 1 \right\},$$

welches offensichtlich konvex und zentralsymmetrisch ist. Wir behaupten, dass E keinen Gitterpunkt $x \in \Lambda \backslash \{0\}$ enthält. Ist nämlich $y \in \Lambda$ ein solcher Punkt, bei dem die k-te Koordinate als letzte von Null verschieden ist, so muss $|y| \geq \lambda_k$ gelten (nach Minimalität von λ_k) und damit

$$\frac{y_1^2}{\lambda_1^2} + \dots + \frac{y_n^2}{\lambda_n^2} = \frac{y_1^2}{\lambda_1^2} + \dots + \frac{y_k^2}{\lambda_k^2} \geq \frac{y_1^2 + \dots + y_k^2}{\lambda_k^2} = \frac{|y|^2}{\lambda_k^2} \geq 1$$

und damit $y \notin E$. Damit enthält E keine von Null verschiedenen Punkte aus Λ. Nach dem Minkowskischen Gitterpunktsatz folgt dann aber wie gewünscht

$$\lambda_1 \dots \lambda_n \ll \mathrm{vol}(E) \ll \det(\Lambda). \quad \square$$

Damit können wir nun endlich auch unser Ausgangsproblem lösen:

Beweis von Lemma 2.9: Ist $n = 1$ und $r_1 = 0$, so ist nichts zu tun. Andernfalls können wir annehmen, dass alle r_i von Null verschieden sind. Wir betrachten die Summe

$$\Lambda = \{(r_1 x, \dots, r_n x) : x \in \mathbb{Z}\} + (p\mathbb{Z})^n \subset \mathbb{R}^n$$

Dies ist sicherlich eine additive Untergruppe von $\mathbb{Z}^n$, mithin ein Gitter. Konkret durchläuft der Vektor $(r_1 x, \dots, r_n x)$ nach Konstruktion genau p verschiedene Vektoren modulo p, somit besteht Λ aus p disjunkten verschobenen Kopien des Untergitters $(p\mathbb{Z})^n$ und es folgt $\det(\Lambda) = p^{n-1}$.
Wir betrachten nun die sukzessiven Minima $0 < \lambda_1 \leq \dots \leq \lambda_n$ und zugehörige minimale Vektoren $v_1, \dots, v_n \in \Lambda$. Zu jedem v_i gibt es nach Konstruktion ein $x_i \in \mathbb{Z}/p\mathbb{Z}$ mit

$$v_i \equiv (r_1 x_i, \dots, r_n x_i) \pmod{p}.$$

Wir definieren unsere VAP nun wie folgt:

$$Q := \{a_1 x_1 + \dots + a_n x_n : -\ell_i \leq a_i \leq \ell_i\} \subset \mathbb{Z}/p\mathbb{Z}$$

für $\ell_i := \frac{p}{4n\lambda_i}$. Wir behaupten nun, dass Q die gewünschten Eigenschaften besitzt.
Einerseits ist Q in der Bohrmenge enthalten: Für $x \in Q$ und $r \in \{r_1, \dots, r_n\}$ ist nämlich

$$\left\| \frac{rx}{p} \right\| \leq \sum_{i=1}^n |a_i| \left\| \frac{r x_i}{p} \right\| \leq \sum_{i=1}^n \ell_i \frac{|v_i|}{p} = \frac{1}{4}.$$

Andererseits müssen wir die Größe von Q abschätzen. Dazu überprüfen wir zunächst, dass Q eine echte VAP ist, d.h. alle Summen $\sum_i a_i x_i$ sind paarweise verschieden. Sonst gäbe es a_i, a_i' mit $\sum_i (a_i - a_i') x_i = 0$. Dann folgt aber $\sum_i (a_i - a_i') v_i \equiv 0 \pmod{p}$. Allerdings sind die Einträge des Vektors links nach Konstruktion im Betrag maximal $2\sum_i \ell_i |v_i| = \frac{p}{2}$, mithin müsste sogar $\sum_i (a_i - a_i') v_i = 0$ gelten, was der linearen Unabhängigkeit der v_i widerspricht. Damit ist Q echt und es folgt

$$|Q| \geq \prod_{i=1}^n (1 + 2\lfloor \ell_i \rfloor) \geq \prod_{i=1}^n \ell_i = \frac{p^n}{(4n)^n \lambda_1 \dots \lambda_n} \gg_n p,$$

wobei wir im letzten Schritt den zweiten Satz von Minkowski benutzt haben. $\square$

3 Inzidenzgeometrie und das Summen-Produkt-Problem

In den letzten Abschnitten haben wir gesehen, dass eine Menge mit kleiner Summenmenge notwendig eine Struktur ähnlich der einer arithmetischen Progression hat. Analog kann man sich fragen, was sich über die Struktur von Mengen mit kleiner Produktmenge $A \cdot A$ aussagen lässt. Diese sollten in gewisser Weise eine Struktur ähnlich der einer *geometrischen Progression* haben. Da arithmetische und geometrische Progressionen sehr unterschiedliche Strukturen haben, sollten nicht beide Fälle gleichzeitig eintreten, mit anderen Worten sollte stets mindestens eine der Mengen $A + A$ oder $A \cdot A$ „groß“ sein. Mit der Technologie, die wir im Beweis des Satzes von Freiman benutzt haben, lässt sich aber nur eine sehr schwache quantitative Version dieser Heuristik beweisen.
Stattdessen wollen wir in diesem Abschnitt sehen, wie sich mit viel einfacheren Methoden recht starke Ergebnisse in dieser Richtung zeigen lassen. Unser erstes Hauptresultat lautet wie folgt:

Satz 3.1 (György Elekes 1997). *Sei $A \subset \mathbb{R}\backslash\{0\}$ endlich und nichtleer. Dann gilt*

$$|A + A| \cdot |A \cdot A| \gg |A|^{5/2}.$$

Insbesondere gilt

$$\max(|A + A|, |A \cdot A|) \gg |A|^{5/4}.$$

Wie bereits in der Einleitung diskutiert haben Erdös und Szemerédi zuerst eine Version der zweiten Aussage mit $|A|^{1+\delta}$ auf der rechten Seite für ein gewisses $\delta > 0$ gezeigt und vermutet, dass die Aussage für jedes $\delta \in (0, 1)$ stimmt (wobei die implizite Konstante von δ abhängt). Mit anderen Worten sollte eine der beiden Mengen stets fast so groß wie möglich sein. Die beste bekannte Aussage ist derzeit $\delta = \frac{1}{3} + \frac{2}{951}$ (Bloom, 2025).

Für $A \subset \mathbb{F}_p$ sieht die Situation anders aus, denn für $A = \mathbb{F}_p$ ist $|A + A| = |A \cdot A| = |A|$. Allerdings kann man zeigen, dass es für kleinere Mengen $A \subset \mathbb{F}_p$ auch ein Summen-Produkt-Phänomen gibt, auch wenn wir dies hier nicht beweisen werden:

Satz 3.2 (Mohammadi, Stevens 2021). *Für $A \subset \mathbb{F}_p$ mit $|A| \leq p^{1/2}$ gilt*

$$\max(|A + A|, |A \cdot A|) \gg |A|^{5/4}.$$

Die entscheidende Idee von Elekes war es, die Aussage geometrisch zu interpretieren: Bezeichnen wir mit $\mathcal{P}$ die Punktmenge $(A + A) \times (A \cdot A) \subset \mathbb{R}^2$ und mit $\mathcal{L}$ die Menge der $|A|^2$ Geraden der Form $y = a_1(x - a_2)$ für $a_1, a_2 \in A$, so geht jede Gerade aus $\mathcal{L}$ durch mindestens $|A|$ Punkte aus $\mathcal{P}$, nämlich durch die Punkte $(a_2 + a_3, a_1 \cdot a_3)$ mit $a_3 \in A$.
Damit sind wir in der folgenden allgemeinen Situation: Gegeben eine endliche Menge $\mathcal{P}$ von p Punkten und eine endliche Menge $\mathcal{L}$ von ℓ Geraden in der Ebene, was lässt sich über die Anzahl $I = I(\mathcal{P}, \mathcal{L})$ an *Inzidenzen*, also an Paaren (P, L) mit $P \in \mathcal{P}, L \in \mathcal{L}$ und $P \in L$ aussagen?
Sicherlich ist $I \leq p \cdot \ell$. Mit wenig Aufwand kann man eine deutlich bessere Schranke erhalten:
Bezeichnen wir für jede Gerade $L \in \mathcal{L}$ mit $r(L)$ die Anzahl an Punkten auf L, so erhalten wir mit Cauchy-Schwarz

$$I^2 = \left(\sum_L r(L)\right)^2 \leq \ell \cdot \sum_L r(L)^2.$$

Es ist aber $\sum_L r(L)^2$ einfach die Anzahl an Tripeln (P_1, P_2, L) mit $P_1, P_2 \in L$. Sicherlich gibt es höchstens I solche Tripel mit $P_1 = P_2$. Für $P_1 = P_2$ ist dagegen L durch P_1, P_2 eindeutig bestimmt, also gibt es höchstens p^2 solche Tripel. Insgesamt folgt $I^2 \leq \ell \cdot (I + p^2)$ und damit $I \ll p\sqrt{\ell} + \ell$. Vertauscht man in diesem Argument die Rollen von p und ℓ, erhält man analog auch $I \ll \ell\sqrt{p} + p$.

In unserer Konfiguration oben mit $p = |A+A| \cdot |A \cdot A|$ und $\ell = |A|^2$ gilt $I \gg |A|^3$. Aus den obigen Abschätzungen ergibt sich in diesem Fall allerdings nur die triviale Schranke $p \gg |A|^2$. Dies ist auch nicht weiter überraschend, denn wir haben nichts über den Grundkörper ausgenutzt und in endlichen Körpern ist sicherlich $A \cdot A = A + A = A = K$ möglich, sodass wir keine bessere Abschätzung erwarten können. Umgekehrt können wir allerdings die Geometrie der euklidischen Ebene benutzen, um unsere Inzidenzschranke zu verbessern:

Satz 3.3 (Szemerédi-Trotter 1983)**.** *In der reellen Ebene gilt*

$$I(\mathcal{P}, \mathcal{L}) \ll p^{2/3}\ell^{2/3} + p + \ell.$$

Die Schranke ist scharf, wie das folgende Beispiel zeigt: Für $k \in \mathbb{N}$ sei $\mathcal{P} = \{(x, y) : 1 \leq x \leq k, 1 \leq y \leq 2k^2\}$ und $\mathcal{L} = \{y = mx + b : 1 \leq m \leq k, 1 \leq b \leq k^2\}$. Dann ist $p = 2k^3$ und $\ell = k^3$. Jede Gerade aus $\mathcal{L}$ geht durch genau k Punkte in $\mathcal{P}$, folglich ist $I(\mathcal{P}, \mathcal{L}) = k^4$.

Beweis von Satz 3.1. Wir wenden den Satz von Szemerédi-Trotter auf die Menge der $p = |A+A| \cdot |A \cdot A|$ Punkte und $|A|^2$ Geraden wie oben an. Dann ist $I \gg |A|^3$ und es folgt

$$|A|^3 \ll p^{2/3}|A|^{4/3} + p + |A|^2,$$

woraus sich durch Umstellen sofort $p \gg |A|^{5/2}$ ergibt. □

Für den Satz von Szemerédi-Trotter geben wir einen eleganten Beweis nach László Székely (1997). Dazu benötigen wir zunächst eine Aussage über die *Kreuzungszahl* eines Graphen, also die minimale Anzahl von Schnittpunkten zweier Kanten eines Graphen, wenn dieser in die Ebene eingebettet wird. Die Kreuzungszahl ist also genau dann 0, wenn der Graph planar ist.

Satz 3.4 (Kreuzungszahl-Ungleichung)**.** *Es sei G ein Graph mit E Kanten und V Knoten. Gilt $E \geq 4V$, dann ist die Kreuzungszahl von G mindestens $\frac{1}{64}\frac{E^3}{V^2}$.*

Beispielsweise hat der vollständige Graph K_n für große n Kreuzungszahl $\gg n^4$.

Beweis. Sei zunächst G planar, zusammenhängend und mit mindestens einem Kreis. Ist F die Anzahl an Flächen, in die der Graph die Ebene teilt (inklusive der äußeren), so gilt nach der Eulerschen Formel $V - E + F = 2$. Andererseits gilt $3F \leq 2E$, denn jede Fläche grenzt an mindestens drei Kanten, aber jede Kante an höchstens zwei Flächen.
Insgesamt folgt $E \leq 3V - 6 \leq 3V$. Die Ungleichung $E \leq 3V$ gilt damit für jeden planaren Graphen, unabhängig davon, ob dieser zusammenhängend ist oder einen Kreis besitzt.
Aus einem Graphen mit k Kreuzungen können wir sicherlich immer k Kanten so entfernen, dass der Graph danach planar ist. Folglich ist die Kreuzungszahl K eines beliebigen Graphen mindestens $E - 3V$. Dies ist natürlich deutlich schwächer als die Schranke, die wir zeigen wollen. Der Trick ist nun, die Ungleichung $K \geq E - 3V$ für einen zufälligen Teilgraphen von G zu betrachten. Dazu sei $p \in [0, 1]$ ein Parameter, den wir später wählen. Wir betrachten einen zufälligen Teilgraphen von G, bei dem wir jeden Knoten mit Wahrscheinlichkeit p unabhängig voneinander wählen (und dann alle Kanten zwischen den verbliebenen Knoten).
Nun gilt $\mathbb{E}(K(G')) \leq p^4 K$, $\mathbb{E}(E(G')) = p^2 E$ sowie $\mathbb{E}(V(G')) = pV$. Es folgt

$$p^4 K \geq p^2 E - 3pV$$

und damit $K \geq E/p^2 - 3V/p^3$. Nun wählen wir $p = \frac{4V}{E} \in [0, 1]$ und erhalten die Behauptung. □

Beweis von Satz 3.3. Wir entfernen zunächst alle Geraden aus $\mathcal{L}$, die höchstens einen Punkt aus $\mathcal{P}$ enthalten. Das sind maximal ℓ Inzidenzen, ist also kein Problem. Aus den übrigen Geraden konstruieren wir einen Graphen, dessen Knoten die Punkte aus $\mathcal{P}$ sind und dessen Kanten die Segmente von Geraden aus $\mathcal{L}$ zwischen zwei benachbarten Punkten aus $\mathcal{P}$ sind. Aus jeder Gerade mit $k \geq 2$ Inzidenzen werden dann $k-1 \geq \frac{k}{2}$ Segmente, also gilt $E \geq I/2$ und natürlich $V = p$. Ist $E < 4V$, so folgt $I \leq 8p$, was in Ordnung ist. Andernfalls ist die Bedingung der Kreuzungszahl-Ungleichung erfüllt und es gibt mindestens $\frac{1}{64}\frac{E^3}{V^2} \geq \frac{1}{512}\frac{I^3}{p^2}$ Kreuzungen. Aber jede Kreuzung ist ein Schnitt von zwei Geraden, davon gibt es folglich maximal ℓ^2. Somit gilt $\frac{1}{512}\frac{I^3}{p^2} \leq \ell^2$ und es folgt $I \leq 8p^{2/3}m^{2/3}$. □

Zum Abschluss wollen wir noch eine elegante Idee von József Solymosi (2009) kennenlernen, mit der wir den Exponenten noch einmal verbessern können:

Satz 3.5 (Solymosi 2009)**.** *Sei $A \subset \mathbb{R}\backslash\{0\}$ endlich mit $|A| \geq 2$. Dann gilt*

$$|A+A|^2 \cdot |A \cdot A| \gg \frac{|A|^4}{\log|A|}.$$

Insbesondere gilt

$$\max(|A+A|, |A \cdot A|) \gg_\varepsilon |A|^{4/3-\varepsilon}.$$

Beweis. O.B.d.A. enthalte A nur positive reelle Zahlen. Wir bezeichnen mit

$$E_m(A) = \#\{(a_1,a_2,a_3,a_4) \in A^4 : a_1a_2 = a_3a_4\}$$

die *multiplikative Energie* von A. Ähnlich wie bei der additiven Energie zeigt eine Anwendung der Cauchy-Schwarz-Ungleichung leicht $E_m(A) \geq \frac{|A|^4}{|A \cdot A|}$. Wir sind also fertig, wenn wir $E_m(A) \ll (\log|A|) \cdot |A+A|^2$ zeigen können. Wir können die Bedingung in der multiplikativen Energie umschreiben zu $a_1/a_3 = a_4/a_2$. Für $x \in A/A$ ist die Lösungszahl von $a_1/a_3 = x$ exakt $|xA \cap A|$. Wir können also

$$E_m(A) = \sum_{x \in A/A} |xA \cap A|^2$$

schreiben. Nun verwenden wir einen nützlichen Trick, *dyadische Zerlegung*: Die Größe von $|xA \cap A|$ variiert zwischen 1 und $|A|$, lässt sich also in $\ll \log|A|$ viele Intervalle der Form $[R, 2R)$ zerlegen. Damit erhalten wir

$$E_m(A) \ll (\log|A|) \cdot \max_{1 \leq R \leq |A|} R^2 \cdot \#\{x \in A/A : R \leq |xA \cap A| < 2R\}.$$

Dabei haben wir den Faktor $\log|A|$ verloren, aber die Information dazugewonnen, dass alle betrachteten Mengen $xA \cap A$ in etwa gleich groß sind. Diese können wir nun geometrisch interpretieren als Punkte aus $A \times A$, die auf der gleichen Ursprungsgerade mit Steigung x liegen. Angeommen, es gibt m solche Werte $x_1 < x_2 < \cdots < x_m$. Entscheidend ist nun, dass die $\gg R^2$ Summen der Form $a+b$ mit $a \in x_iA \cap A$ und $b \in x_{i+1}A \cap A$ alle verschieden sind (denn aus $a+b = a'+b'$ folgt $a-a' = b-b'$, aber $a-a'$ liegt auf der Geraden mit Steigung x_i und $b-b'$ auf der Geraden mit Steigung x_{i+1}) und auch für verschiedene i verschiedene Summen liefern (denn diese Summen liegen immer im Kegel zwischen den beiden Geraden). Schließlich können wir auch zwischen der Geraden mit Steigung x_m sowie der senkrechten Geraden durch $x = \min A$ noch einmal $\gg R^2$ verschiedene Summen konstruieren. Insgesamt erhalten wir damit $\gg mR^2$ verschiedene Elemente von $(A+A) \times (A+A)$ und es folgt $mR^2 \ll |A+A|^2$ und damit die Behauptung. □

4 Die Polynommethode

Als Motivation erinnern wir uns noch einmal an den Satz von Cauchy-Davenport: Für nichtleere Mengen $A, B \subset \mathbb{F}_p$ gilt

$$|A+B| \geq \min(|A|+|B|-1, p).$$

Unser kombinatorischer Beweis war nicht kompliziert, aber die Beweistechnik ist nicht besonders robust. Dies zeigt zum Beispiel die folgende leichte Variation, die 1964 von Erdös und Heilbronn vermutet wurde, aber erst 1994 von da Silva und Hamidoune bewiesen werden konnte:

Satz 4.1 (Erdös-Heilbronn-Vermutung)**.** *Ist $A \subset \mathbb{F}_p$ endlich und nichtleer, so gilt*

$$\#\{a_1+a_2 : a_1, a_2 \in A, a_1 \neq a_2\} \geq \min(2|A|-3, p).$$

Wie das Beispiel $A = \{0, 1, \ldots, k-1\}$ leicht zeigt, ist die Schranke scharf.
Unsere vorige Beweistechnik für den Satz von Cauchy-Davenport bricht hier komplett zusammen. In diesem Abschnitt wollen wir die Polynommethode entwickeln und mit ihr einen robusteren Beweis für Cauchy-Davenport geben, mit dem sich dann auch die Erdös-Heilbronn-Vermutung zeigen lässt.

4.1 Der kombinatorische Nullstellensatz

Die für uns in diesem Abschnitt grundlegende Eigenschaft von Polynomen ist, dass diese nicht an allzu vielen Stellen gleichzeitig verschwinden können. Eine sehr leichte, aber oft bereits nützliche Version für Polynome in mehreren Variablen ist die folgende Aussage:

Lemma 4.2 (Nullstellen-Schranke)**.** *Sei K ein Körper und $P \in K[x_1, \ldots, x_n]$ von Grad maximal d_i in x_i sowie $S_1, \ldots, S_n \subset K$ Mengen mit $|S_i| > d_i$. Angenommen, $P(x_1, \ldots, x_n) = 0$ für alle $(x_1, \ldots, x_n) \in S_1 \times \cdots \times S_n$. Dann ist P das Nullpolynom.*

Beweis. Wir verwenden Induktion über n. Für $n = 1$ reduziert sich das Problem auf die bekannte Tatsache, dass ein Polynom $P \neq 0$ von Grad d über einem beliebigen Körper maximal d Nullstellen haben kann. Sei nun $n > 1$ und die Aussage für $n-1$ bereits bekannt. Wir schreiben

$$P(x_1, \ldots, x_n) = \sum_{j=0}^{d_n-1} P_j(x_1, \ldots, x_{n-1}) x_n^j.$$

Für feste $x_1, \ldots, x_{n-1} \in S_1 \times \cdots \times S_{n-1}$ ist die rechte Seite ein Polynom in x_n von Grad maximal d_n, was für alle $x_n \in S_n$ verschwindet, mithin das Nullpolynom ist. Also ist $P_j(x_1, \ldots, x_{n-1}) = 0$. Da dies für alle $x_1, \ldots, x_{n-1} \in S_1 \times \cdots \times S_{n-1}$ der Fall ist und die P_j ebenfalls Grad maximal d_i in x_i haben, folgt aus der Induktionsvoraussetzung, dass P_j für alle j das Nullpolynom ist und dann auch P. □

Für viele Anwendungen ist die Bedingung aber zu stark, den Grad in jeder Variablen einzeln kontrollieren zu müssen. Eine raffinierte Weiterentwicklung der gleichen Idee liefert der folgende Satz:

Satz 4.3 (Kombinatorischer Nullstellensatz)**.** *Sei K ein Körper und $P \in K[x_1, \ldots, x_n]$ von Grad $d_1 + \cdots + d_n$ sowie $S_1, \ldots, S_n \subset K$ Mengen mit $|S_i| > d_i$. Angenommen, $P(x_1, \ldots, x_n) = 0$ für alle $(x_1, \ldots, x_n) \in S_1 \times \cdots \times S_n$. Dann ist der Koeffizient von $x_1^{d_1} \cdot \ldots \cdot x_n^{d_n}$ in P gleich Null.*

Beweis. Wir verwenden Induktion über n, wobei der Fall $n = 1$ sofort aus dem vorigen Lemma folgt. Im allgemeinen Fall können wir sicherlich $|S_i| = d_i + 1$ annehmen. Wir teilen nun P mit Rest durch $\prod_{s\in S_n}(x_n - s)$ und erhalten eine Darstellung

$$P(x_1, \ldots, x_n) = Q(x_1, \ldots, x_n) \prod_{s\in S_n} (x_n - s) + R(x_1, \ldots, x_n)$$

mit

$$R(x_1, \ldots, x_n) = x_n^{d_n} R_{d_n}(x_1, \ldots, x_{n-1}) + x_n^{d_n-1} R_{d_n-1}(x_1, \ldots, x_{n-1}) + \cdots + R_0(x_1, \ldots, x_{n-1}).$$

Nach Konstruktion verschwindet R auf ganz $S_1 \times \cdots \times S_n$. Für feste $(x_1, \ldots, x_{n-1}) \in S_1 \times \cdots \times S_{n-1}$ ist dies aber wieder ein Polynom in x_n von Grad d_n, was auf ganz S_n verschwindet, muss also das Nullpolynom sein. Damit verschwinden alle R_i auf $S_1 \times \ldots S_{n-1}$. Man sieht nun aber, dass der Koeffizient von $x_1^{d_1} \cdot \ldots \cdot x_n^{d_n}$ in P gleich dem Koeffizienten von $x_1^{d_1} \cdot \ldots \cdot x_{n-1}^{d_{n-1}}$ in R_{d_n} ist, welcher nach Induktionsvoraussetzung verschwindet. □

Als Anwendung wollen wir zunächst noch einmal den Satz von Cauchy-Davenport beweisen:

Beweis von Cauchy-Davenport. O.B.d.A. sei $|A| + |B| \leq p + 1$, dann wollen wir $|A + B| \geq |A| + |B| - 1$ zeigen. Angenommen, es ist $|A + B| \leq |A| + |B| - 2$. Für eine Menge $C \subset \mathbb{F}_p$ mit $A + B \subset C$ und $|C| = |A| + |B| - 2$ betrachten wir das Polynom

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

von Grad $|A| + |B| - 2$. Nach Konstruktion verschwindet P auf $A \times B$. Nach dem kombinatorischen Nullstellensatz muss dann der Koeffizient von $x^{|A|-1}y^{|B|-1}$ verschwinden, aber dieser ist $\binom{|A|+|B|-2}{|A|-1}$, was nicht Null in $\mathbb{F}_p$ (also nicht durch p teilbar) ist, da $|A|+|B|-2 < p$ vorausgesetzt war. □

Mit dieser Technik lässt sich nun auch leicht die Erdös-Heilbronn-Vermutung lösen:

Beweis von Satz 4.1. Für $p = 2$ ist die Aussage trivial. Sei nun $p \geq 3$ ungerade, dann können wir o.B.d.A. annehmen, dass $2|A| - 3 \leq p$ gilt. Angenommen, die linke Seite wäre höchstens $2|A| - 4$. Dann betrachten wir wieder eine Menge C, die alle diese Summen und insgesamt exakt $2|A| - 4$ Elemente enthält sowie das Polynom

$$P(x, y) = (x - y)^2 \prod_{c\in C}(x + y - c).$$

Dann hat P Grad $2|A| - 2$ und verschwindet nach Konstruktion auf ganz $A \times A$, mithin muss nach dem Kombinatorischen Nullstellensatz der Koeffizient von $x^{|A|-1}y^{|A|-1}$ in P verschwinden. Dieser ist aber

$$2\binom{2|A| - 4}{|A| - 1} - 2\binom{2|A| - 4}{|A| - 2} = -\frac{2 \cdot (2|A| - 4)!}{(|A| - 1)!(|A| - 2)!},$$

was nach Konstruktion nicht durch p teilbar und damit nicht 0 in $\mathbb{F}_p$ ist. □

Bemerkung: Die interessierte Leserin möge sich überlegen, warum das gleiche Argument mit dem eigentlich naheliegenderen Polynom $P(x, y) = (x - y) \prod_{c\in C}(x + y - c)$ nicht funktioniert (und eine vermeintlich bessere Schranke gezeigt) hätte.

4.2 Das Kakeya-Problem über endlichen Körpern

Das Kakeyaproblem (1917) fragt ursprünglich nach der kleinsten Menge in der Ebene, in der man eine Nadel der Länge 1 stetig einmal um 360° um sich selbst drehen kann. Zum Beispiel geht das offensichtlich in einer Kreisscheibe mit Radius $\frac{1}{2}$ (Fläche $\pi/4$), aber auch in einem gleichseitigen Dreieck mit Höhe 1 (Fläche $1/\sqrt{3}$) und es ist nicht schwer, noch etwas bessere Beispiele hinzuschreiben. Überraschenderweise konnte Besikowitsch 1919 solche Mengen konstruieren, die beliebig kleines Lebesgue-Maß haben. Seine Konstruktion reduziert das Problem zunächst auf die Konstruktion von *Besikowitsch-Mengen*, das sind Mengen, die ein Einheitssegment in jede Richtung enthalten.
Auch wenn solche Mengen beliebig kleines Lebesgue-Maß haben können, besagt die *Kakeya-Vermutung*, dass eine Besikowitsch-Menge im $\mathbb{R}^n$ immer noch *Hausdorff-Dimension* n haben muss (also in einem gewissen Sinne nicht allzu klein sein kann). Die Kakeya-Vermutung war bis vor kurzem nur für $n = 2$ bekannt, im Februar 2025 wurde sie für $n = 3$ von Hong Wang und Joshua Zahl bewiesen, für alle $n \geq 4$ ist sie weiterhin offen.

Im Jahr 1999 wurde vom Thomas Wolff ein Analogon der Kakeya-Vermutung über endlichen Körpern aufgestellt. Die Hoffnung bei solchen Problemen ist, dass sie über endlichen Körpern etwas leichter zugänglich sind, sich die Methoden aber im Idealfall später auch auf den reellen Fall übertragen lassen. Lange wurde geglaubt, dass die Kakeya-Vermutung auch über endlichen Körpern ein sehr schweres Problem ist, bis Zeev Dvir im Jahr 2008 einen sehr kurzen und eleganten Beweis gefunden hat, mit dem wir uns nun beschäftigen.

Satz 4.4 (Kakeya-Vermutung über $\mathbb{F}_p$, Dvir 2008)**.** *Sei p prim und $K \subset \mathbb{F}_p^n$ eine Menge, die für jeden Vektor $y \in \mathbb{F}_p^n$ eine Gerade der Form $\{x + ty : x, t \in \mathbb{F}_p\}$ enthält. Dann ist $|K| \gg_n p^n$. Konkret kann die implizite Konstante $\frac{1}{n!}$ gewählt werden.*

Für den Beweis benötigen wir zunächst einige allgemeine Aussagen über Polynome über endlichen Körpern. Sei dazu $\mathrm{Poly}_D(\mathbb{F}_p^n)$ der Vektorraum aller Polynome über $\mathbb{F}_p$ in n Variablen, deren Gesamtgrad maximal D ist. Mithilfe elementarer Kombinatorik sieht man, dass die Dimension dieses Vektorraums $\binom{D+n}{n}$ ist, insbesondere mindestens $\frac{D^n}{n!}$.

Lemma 4.5 (Parameter-Schranke)**.** *Ist $S \subset \mathbb{F}_p^n$ und D so, dass $|S| < \binom{D+n}{n}$ ist, dann gibt es ein Polynom $P \in \mathrm{Poly}_D(\mathbb{F}_p^n)\backslash\{0\}$, welches auf ganz S verschwindet.*

Beweis. Für jeden Punkt $x \in S$ ist die Bedingung $P(x) = 0$ eine lineare Relation, die die Koeffizienten von P erfüllen müssen. Nach Voraussetzung ist die Anzahl dieser Relationen kleiner als die Dimension des Vektorraums, mithin hat der Lösungsraum positive Dimension und enthält damit in jedem Fall ein $P \neq 0$. □

Beweis von Satz 4.4: Angenommen, es gibt eine Besikowitsch-Menge K mit $|K| < \binom{p+n-1}{n}$. Dann gibt es nach der Parameter-Schranke ein Polynom P von Grad $D \leq p-1$, welches auf ganz K verschwindet. Wir schreiben $P = P_D + R$, wobei P_D homogen von Grad D ist und R von Grad maximal $D-1$.
Für ein beliebiges $y \in \mathbb{F}_p^n\backslash\{0\}$ gibt es nach Voraussetzung ein $x \in \mathbb{F}_p^n$ mit $P(x+ty) = 0$ für alle $t \in \mathbb{F}_p$. Als Polynom in t hat $P(x+ty)$ nun Grad maximal $D \leq p-1$ mit Leitkoeffizient $P_D(y)$. Nach der Nullstellenschranke für $n = 1$ muss es damit identisch Null sein, also folgt $P_D(y) = 0$. Damit verschwindet P_D für alle $y \in \mathbb{F}_p^n$. Nach der Nullstellenschranke ist dann aber P_D das Nullpolynom, was ein Widerspruch zu unserer Annahme ist, dass P Grad D besitzt.
Damit folgt, dass jede Besikowitsch-Menge $|K| \geq \binom{p+n-1}{n} \geq \frac{p^n}{n!}$ erfüllt. □

4.3 Arithmetische Progressionen in $\mathbb{F}_3^n$

In diesem Abschnitt wollen wir noch ein anderes Problem untersuchen, welches über endlichen Körpern einen ganz anderen Zugang erlaubt als über $\mathbb{Z}$. Dies betrifft die bereits in der Einleitung diskutierte Frage, wie groß eine Menge sein kann, die keine arithmetische Progression der Länge 3 (3-AP) enthält. (Hier und im Folgenden werden nur nichttriviale 3-AP betrachtet, Folgen der Form (x, x, x) können wir natürlich nicht ausschließen.)

Zum Problem für Teilmengen von $\{1, 2, \ldots, N\} \subset \mathbb{Z}$ werden wir in den nächsten Kapiteln noch zurückkommen. Hier sei nur angedeutet, dass es ein sehr schwieriges Problem ist und alle nichttrivialen Schranken für die maximale Größe $r_3(N)$ einer solchen Menge von der Form $\frac{N}{\omega(N)}$ sind für eine Funktion $\omega(n)$, die relativ langsam gegen unendlich geht.

Untersuchen wir stattdessen Teilmengen von $\mathbb{F}_p^n$ für eine feste Primzahl p und $n \to \infty$, so ist die triviale obere Schranke p^n. Speziell für $p = 3$ handelt es sich wie in der Einleitung diskutiert um das Problem, wie groß eine Teilmenge von $\mathbb{F}_3^n$ sein kann, die kein SET enthält.

Da die Menge $\{0, 1\}^n$ sicherlich keine 3-AP enthält, liegt die richtige Antwort irgendwo zwischen 2^n und 3^n. Mit ähnlichen (technisch aufwändigen) Methoden wie für den Fall von $\mathbb{Z}$ konnte Meshulam 1995 die obere Schranke $\ll \frac{3^n}{n}$ zeigen, was schließlich 2012 von Bateman und Katz zu $\frac{3^n}{n^{1+c}}$ für ein gewisses (sehr kleines) $c > 0$ verbessert wurde.

Es war lange Zeit eine offene Frage und wurde für ein sehr schwieriges Problem gehalten, ob es eine Konstante $c < 3$ gibt, sodass solche Mengen höchstens $\ll c^n$ Elemente enthalten. Umso überraschender war es, als 2016 ein sehr kurzer Beweis dafür gefunden wurde:

Satz 4.6 (Ellenberg-Gijswijt 2016)**.** *Sei $p > 2$ prim. Dann gibt es eine Konstante $c_p < p$ mit der folgenden Eigenschaft: Für jede Menge $S \subset \mathbb{F}_3^n$ ohne nicht-triviale 3-AP gilt $|S| \ll c_p^n$. Es kann $c_3 = 2.756$ gewählt werden. Für große p kann man $c_p \approx 0.85p$ wählen.*

Beweis. Wir betrachten Polynome in $\mathbb{F}_p[x_1, \ldots, x_n]$, die in jeder Variablen von Grad höchstens $p-1$ sind. Diese sind in Bijektion mit den Funktionen $f : \mathbb{F}_p^n \to \mathbb{F}_p$. Für $0 \le d \le (p-1)n$ sei M_d die Menge solcher Monome, die Gesamtgrad höchstens d haben, und $m_d = |M_d|$ ihre Anzahl. Es ist etwa $m_0 = 1$, $m_{(p-1)n} = p^n$ und allgemeiner $m_d = p^n - m_{(p-1)n-d-1}$, da die Abbildung $x_1^{d_1} \ldots x_n^{d_n} \mapsto x_1^{p-1-d_1} \ldots x_n^{p-1-d_n}$ die Monome aus M_d bijektiv auf die Monome abbildet, die Grad mindestens $(p-1)n - d$ haben, wovon es genau $p^n - m_{(p-1)n-d-1}$ gibt.
Wir betrachten nun den Vektorraum V der Polynome in M_d, die auf dem Komplement von $2S = \{2s : s \in S\}$ verschwinden. Sicherlich hat V Dimension mindestens $m_d - (p^n - |S|) = |S| - m_{(p-1)n-d-1}$. Nach Annahme verschwinden alle Polynome aus V dann auch auf $S \overset{\vee}{+} S = \{s_1 + s_2 : s_1, s_2 \in S, s_1 \neq s_2\}$, denn diese liegt im Komplement von $2S$.
Wir zeigen, dass jedes dieser Polynome dann sogar auf allen bis auf höchstens $2m_{d/2}$ Elementen von $2S$ verschwindet: Sei P ein solches Polynom, welches auf $S \overset{\vee}{+} S$ verschwindet. Drücken wir $P(x+y)$ als Linearkombination der Monome in x und y aus, so hat in jedem Summanden einer der Ausdrücke in x oder in y höchstens Grad $d/2$. Wir können also

$$P(x+y) = \sum_{m \in M_{d/2}} m(x) F_m(y) + \sum_{m' \in M_{d/2}} m'(y) G_{m'}(x)$$

mit Monomen m, m' von Grad höchstens $d/2$ schreiben. Die $|S| \times |S|$-Matrix mit Einträgen $P(s_1 + s_2)$ ist damit eine Summe von höchstens $2m_{d/2}$ Matrizen von Rang 1 und hat somit Rang höchstens $2m_{d/2}$. Nach Annahme ist es aber eine Diagonalmatrix, d.h. ihr Rang ist gleich der Anzahl der Diagonaleinträge, die nicht Null sind. Damit sind maximal $2m_{d/2}$ der Werte $P(2s)$ mit $s \in S$ von Null verschieden.

Damit haben wir einen Vektorraum V der Dimension mindestens $|S| - m_{(p-1)n-d-1}$, dessen Elemente jeweils auf allen bis auf maximal $2m_{d/2}$ Elementen von $2S$ verschwinden.
Wir behaupten, dass dies $\dim V \leq 2m_{d/2}$ impliziert. Dazu sei $P \in V$ derart, dass P an maximal vielen Stellen von $2S$ nicht verschwindet. Sei M dieses Maximum. Wäre $\dim V > M$, so könnten wir eine Funktion Q aus V konstruieren, die an allen diesen M Stellen verschwindet, aber nicht auf ganz $2S$. Dann wäre aber $P+Q$ an mehr Stellen als P von Null verschieden, Widerspruch! Somit folgt $\dim V \leq M \leq 2m_{d/2}$ wie behauptet. Es ergibt sich

$$|S| \leq m_{(p-1)n-d-1} + 2m_{d/2}.$$

Wählen wir nun $d = \left\lfloor \frac{2(p-1)n}{3} \right\rfloor$, so erhalten wir

$$|S| \leq 3m_{2(p-1)n/3}.$$

Schließlich schätzen wir noch die Anzahl solcher Monome als $m_{2(p-1)n/3} \ll c_p^n$ ab. Dies gelingt mit einem einfachen Trick (in der Stochastik würde man es nach Chernoff benennen, in der Zahlentheorie nach Rankin): Für ein zunächst beliebiges $t \in (0,1]$ ist

$$m_d = \sum_{0 \leq d_1,\dots,d_n \leq p-1 : d_1+\dots+d_n \leq d} 1 \leq \sum_{0 \leq d_i \leq p-1} t^{d_1+\dots+d_n-d} = \frac{(1+t+t^2+\dots+t^{p-1})^n}{t^d}.$$

Damit folgt $m_{(p-1)n/3} \leq c_p^n$ für

$$c_p = \min_{t \in (0,1]} \frac{1+t+t^2+\dots+t^{p-1}}{t^{(p-1)/3}}.$$

Elementare Analysis zeigt $c_p < p$ für alle p (indem wir $t = 1-\varepsilon$ betrachten), $c_p < 0.85p$ (indem wir $t = 1 - 2.149/p$ betrachten) und $c_3 < 2.756$ (das Minimum liegt bei $t = \frac{\sqrt{33}-1}{8}$). □

5 Der Satz von Roth über arithmetische Progressionen

Nachdem wir im letzten Abschnitt Teilmengen von $\mathbb{F}_p^n$ untersucht haben, die keine 3-AP enthalten, wollen wir uns nun der gleichen Frage für Mengen von ganzen Zahlen zuwenden.

Definition 5.1. *Sei $k \geq 3$. Wir bezeichnen mit $r_k(N)$ die Größe der größten Teilmenge von $\{1,2,\dots,N\}$, die keine k-AP enthält.*

Sicherlich gilt $r_3(N) \leq r_4(N) \leq r_5(N) \leq \dots \leq N$.
Zum Beispiel gilt $r_3(1) = 1, r_3(2) = 2, r_3(3) = 2$ und $r_3(m+n) \leq r_3(m) + r_3(n)$. Daraus folgt leicht induktiv $r_3(n) \leq \frac{2n+2}{3}$ für alle n. Mit etwas mehr Aufwand kann man auf ähnliche Weise leicht Schranken der Art $r_3(n) \leq cn$ für gewisse Konstanten $c \in (0,1)$ und alle hinreichend großen n zeigen. Erdős und Turán haben dies 1936 zuerst systematisch untersucht und die folgende Vermutung aufgestellt, die 1953 von Klaus Friedrich Roth bewiesen wurde.

Satz 5.2 (Roth 1953)**.** *Es gilt $r_3(N) \ll \frac{N}{\log\log N}$. Insbesondere gilt $\lim_{N\to\infty} r_3(N)/N = 0$.*

Der Quotient $\frac{1}{\log\log N}$ geht zwar gegen Null, aber natürlich nur sehr langsam. In den letzten 70 Jahren wurde dieser quantitative Aspekt in vielen Schritten verbessert. Ein Meilenstein war 2020 der Beweis der Schranke $r_3(N) \ll \frac{N}{(\log N)^{1+c}}$ für eine positive Konstante c durch Tom Bloom und Olof Sisask.
Weil es bis N etwa $\frac{N}{\log N}$ Primzahlen gibt, impliziert dies insbesondere, dass es unendlich viele 3-AP von Primzahlen gibt. Das war zwar schon lange vorher bekannt, allerdings folgt es nun zum ersten

Mal nur aus der Tatsache, dass es genügend Primzahlen gibt!
Die aktuell beste bekannte Schranke ist $r_3(N) \ll \frac{N}{(\log N)^{1/9}}$ von 2023, wiederum von Bloom und Sisask, basierend auf einer neuen Idee von Kelley und Meka.
Erdős und Turán untersuchten auch längere Progressionen und fragten, ob hier ähnliche Aussagen gelten. Dies konnte erst 1975 von Endre Szemerédi bewiesen werden:

Satz 5.3 (Szemerédi)**.** *Sei $k \geq 3$. Dann gilt $\lim_{N\to\infty} r_k(N)/N = 0$.*

Äquivalent (Übung) enthält jede Teilmenge $A \subset \mathbb{N}$ postiver Dichte beliebig lange arithmetische Progressionen. In diesem Sinn ist der Satz von Szemerédi eine (weitreichende) Verallgemeinerung des Satzes von van der Waerden, der besagt, dass bei einer Färbung der natürlichen Zahlen in endlich vielen Farben immer eine der Farben beliebig lange arithmetische Progressionen enthält.

Auch der Satz von van der Waerden besitzt eine quantitative Version: Zu jedem r und k gibt es eine Zahl $W(r,k)$, sodasss bei einer Färbung der ersten $W(r,k)$ natürlichen Zahlen in r Farben stets eine einfarbige k-AP entsteht.
Der ursprüngliche Beweis von van der Waerden liefert allerdings nur sehr schlechte Schranken für $W(r,k)$, diese sind etwa von der Art $W(2,k) \ll 2^{2^{2^{\cdots}}}$, wobei der Potenzturm aus etwa k Potenzen besteht. Eine Hauptmotivation von Erdős und Turán war es, bessere Schranken für $W(r,k)$ zu bekommen. Während der ursprüngliche Beweis von Szemerédi keine quantitativen Ergebnisse lieferte, konnte Timothy Gowers 2001 (mit anderen Methoden) die Schranke

$$r_k(N) \ll \frac{N}{(\log\log N)^{2^{-2^{k+9}}}}$$

beweisen. Daraus folgt leicht die Schranke $W(2,k) \ll 2^{2^{2^{2^{2^{k+9}}}}}$. Diese ist zwar immer noch nicht besonders stark (und sicherlich weit von der Wahrheit entfernt), aber immerhin ein großer Fortschritt gegenüber dem vorigen Potenzturm.
Bevor wir den Satz von Roth beweisen, wollen wir zunächst untere Schranken für $r_3(N)$ diskutieren. Hier geht es also darum, möglichst große Mengen zu konstruieren, die keine 3-AP enthalten. Dies ist natürlich einerseits ein an sich interessantes Problem, andererseits verrät uns eine untere Schranke auch immer etwas darüber, welche Methoden zum Beweis der oberen Schranke *nicht* funktionieren werden (weil sie ein zu starkes Ergebnis zeigen würden).
Die erste Konstruktion stammt aus der Arbeit von Erdős und Turán (und wohl von George Szekeres):

Satz 5.4. *Es gilt $r_3(N) \geq \frac{1}{2}N^{\log(2)/\log(3)} \gg N^{0.63}$.*

Beweis. Wir wählen $k = \lfloor \log_3(N) \rfloor$ und betrachten die Menge A der Zahlen von 1 bis $3^k - 1$, die in ihrer Darstellung zur Basis 3 nur die Ziffern 0 und 1 enthalten. Dann enthält A genau $2^k \geq 2^{\log_3(N)-1} = \frac{1}{2}N^{\log(2)/\log(3)}$ Elemente. Wir behaupten, dass A keine 3-AP enthält. Angenommen, es gäbe solche $x, y, z \in A$, nicht alle gleich, mit $x + z = 2y$. Nach Konstruktion können bei der ternären Addition von $x + z$ und $y + y$ keine Überträge entstehen. Es muss also an jeder Stelle die Summe der Ziffern von x und z gleich dem Doppelten der entsprechenden Ziffer von y sein. Hat y also an einer Stelle die Ziffer 0, muss dies auch für x und z gelten, analog für die Ziffer 1. Dann ist aber $x = y = z$, was ausgeschlossen wurde. □

Diese Konstruktion ist recht natürlich und es ist nicht unmittelbar klar, wie sie sich verbessern lässt. Erdős und Turán vermuteten sogar, dass sie optimal ist. Dies stellte sich allerdings bald als falsch heraus, wie 1942 Raphaël Salem und Donald C. Spencer entdeckten:

Satz 5.5 (Salem-Spencer 1942)**.** *Es gilt* $r_3(N) \gg N^{1-2\log\log\log N/\log\log N}$*. Insbesondere gilt* $r_3(N) \gg N^{0.9999999}$*.*

Beweis. Die Idee ist, die Konstruktion von Szekeres auf eine größere Basis als 3 zu verallgemeinern. Dazu wählen wir einen Parameter $d \geq 2$ und betrachten diejenigen Zahlen, die in Basis $2d-1$ nur die Ziffern $0, 1, \ldots, d-1$ enthalten. Damit ist weiterhin sichergestellt, dass bei der Addition zweier dieser Zahlen keine Überträge entstehen können, sodass wir das Problem ziffernweise untersuchen können. Allerdings kann es nun wegen $0+2 = 1+1$ durchaus 3-AP geben. Um dies zu vermeiden, betrachten wir die Menge A aller Zahlen mit genau d Stellen in Basis $2d-1$, die jede der Ziffern $0, 1, \ldots, d-1$ genau einmal enthalten. Dann hat A genau $d!$ Elemente. Wir können hier jedes $d < \frac{\log N}{\log\log N}$ wählen, denn dann sind alle Zahlen in A kleiner als

$$(2d-1)^d < (\log N)^{\frac{\log N}{\log\log N}} = N.$$

Wählen wir $d \approx \frac{\log N}{\log\log N}$, so gilt nach der Stirling-Formel

$$|A| = d! > \left(\frac{d}{e}\right)^d \gg \left(\frac{\log N}{3\log\log N}\right)^{\frac{\log N}{\log\log N}} \gg \frac{N}{N^{2\log\log\log N/\log\log N}}.$$

Wir zeigen nun, dass A keine 3-AP enthält. Angenommen, es gäbe nun $x, y, z \in A$, nicht alle gleich, mit $x+z = 2y$. Dann müssen die stellenweisen Ziffernsummen ebenfalls übereinstimmen. Für die Stelle, an der y die Ziffer $d-1$ hat, bedeutet dies aber, dass auch x und z die Ziffer $d-1$ haben müssen, anders ist die Ziffernsumme $2(d-1)$ nicht realisierbar. Dann muss aber auch an der Stelle, an der y die Ziffer $d-2$ hat, bei x und z ebenfalls die Ziffer $d-2$ stehen, denn $d-1$ kann dort nicht mehr vorkommen und anders ist die Summe $2(d-2)$ nicht erreichbar. Wiederholen dieses Arguments zeigt $x = y = z$, Widerspruch! □

Etwas später konnte die Konstruktion von Felix Behrend noch etwas verbessert werden:

Satz 5.6 (Behrend 1946)**.** *Es gilt*

$$r_3(N) \gg \frac{N}{\exp(4\sqrt{\log N})}.$$

Beweis. Für noch zu wählende Parameter m und d betrachten wir m-stellige Zahlen in Basis $2d-1$, die nur die Ziffern $0, 1, \ldots, d-1$ enthalten. So entstehen weiterhin keine Überträge. Statt nun aber etwas über die Häufigkeit der Ziffern auszusagen, betrachten wir die Summe der Quadrate der Ziffern. Diese ist sicherlich kleiner als md^2. Es gibt also einen Wert s, der bei mindestens $\frac{d^m}{md^2}$ dieser Zahlen als Summe der Quadrate der Ziffern angenommen wird. Wir wählen nun A als die Menge dieser Zahlen. Diese Menge enthält keine 3-AP, denn die Vektoren von Ziffern der Zahlen aus A liegen nach Konstruktion auf einer Sphäre im $\mathbb{R}^n$ und eine Sphäre und eine Gerade schneiden sich in maximal zwei Punkten, somit kann A keine drei kollinearen Vektoren enthalten. Nun müssen wir noch die Größe von A untersuchen: Wählen wir d beliebig und $m = \lfloor \frac{\log N}{\log(2d-1)} \rfloor$, so sind alle Zahlen in A kleiner als N und wir erhalten

$$|A| \geq \frac{d^m}{md^2} \gg \frac{d^{\frac{\log N}{\log(2d)}}}{(\log N)d^3} = \frac{N}{(\log N)N^{\frac{\log 2}{\log(2d)}}d^3} = \frac{N}{\exp\left(\log\log N + \frac{(\log 2)\log N}{\log(2d)} + 3\log(d)\right)}.$$

Wählen wir hier $\log(d) \approx \sqrt{\log N}$, erhalten wir die Behauptung. □

Die Schranke von Behrend sieht nicht besonders natürlich aus, dennoch hat seit 1946 niemand eine substantiell bessere Konstruktion gefunden. Mit der oberen Schranke von Kelley und Meka gibt es inzwischen gute Gründe zu glauben, dass sie sogar im wesentlichen optimal ist!

Die bisher konstruierten unteren Schranken zeigen, dass die Vermutung von Erdős und Turán, wenn überhaupt, nur gerade so stimmen kann. Insbesondere muss jede obere Schranke viel schwächer sein als alles, was wir im Fall $\mathbb{F}_p^n$ herausgefunden haben (dies hätte eine obere Schranke der Form $r_3(N) \ll N^c$ für ein $c < 1$ suggeriert, was aber den unteren Schranken widersprechen würde!). Insofern war es relativ überraschend, als Klaus Roth 1953 einen Beweis der Vermutung für 3-AP finden konnte.
Für den Beweis benötigen wir noch einmal Fourier-Analysis. Obwohl Roth selbst Fourier-Analysis direkt auf $\mathbb{Z}$ verwendete, ist es technisch etwas einfacher, das Problem in $\mathbb{Z}/p\mathbb{Z}$ einzubetten.
Dazu wählen wir einfach eine Primzahl $p \in (2N, 4N)$, die es nach dem Bertrandschen Postulat sicherlich gibt. Wir können eine Menge $A \subset \{1, 2, \dots, N\}$ nun mit der Menge der entsprechenden Restklassen modulo p identifizieren.
Enthält A in den ganzen Zahlen keine 3-AP, so bleibt dies wegen $p > 2N$ auch modulo p gültig, denn wäre $x + z \equiv 2y \pmod p$, so folgt wegen $0 \leq x + z, 2y < p$ sofort $x + z = 2y$.
Der nächste Beweisschritt basiert nun auf der Heuristik, dass eine zufällige Teilmenge $A \subset \mathbb{Z}/p\mathbb{Z}$ von $|A| = \delta p$ Elementen sogar sehr viele 3-AP enthalten sollte, nämlich in etwa $\delta^3 p^2$. Denn die Anzahl der 3-AP in $\mathbb{Z}/p\mathbb{Z}$ selbst ist p^2 und jede davon sollte mit einer Wahrscheinlichkeit von δ^3 komplett in A liegen.
Den vagen Begriff der „Zufälligkeit“ können wir mithilfe von Fourier-Analysis präzise machen:

Lemma 5.7 (Fourier-Koeffizienten kontrollieren 3-APs)**.** *Sei $p > 2$ prim und $A \subset \mathbb{Z}/p\mathbb{Z}$ mit $|A| = \delta p$. Ist $\max_{r \neq 0} |\widehat{A}(r)| \leq \frac{\delta^2}{2} p$, so enthält A mindestens $\frac{1}{2}\delta^3 p^2$ arithmetische Progressionen der Länge 3 (wobei triviale Progressionen (x, x, x) zunächst mitgezählt werden).*

Beweis. Nach den Orthogonalitätsrelationen ist die Anzahl der 3-AP (inklusive der trivialen) gleich

$$\frac{1}{p} \sum_{r \in G} \widehat{A}(r)^2 \widehat{A}(-2r) \geq \frac{\widehat{A}(0)^3}{p} - \frac{1}{p} \max_{r \neq 0} |\widehat{A}(r)| \sum_r |\widehat{A}(r)|^2 \geq \delta^3 p^2 - \frac{\delta^2}{2} \cdot \delta p^2 = \frac{\delta^3}{2} p^2$$

nach der Dreiecksungleichung, unserer Annahme und Parseval. □

Hier haben wir die trivialen 3-AP mitgezählt. Ihre Anzahl ist aber lediglich $|A| = \delta p$. Ist $\delta \gg \frac{1}{p^{1/2}}$, so ist diese Anzahl kleiner als $\frac{1}{2}\delta^3 p^2$, mithin muss es unter der Voraussetzung des Lemmas auch nicht-triviale 3-AP geben.
Die entscheidende Frage ist allerdings, was wir machen, wenn A mindestens einen großen Fourier-Koeffizienten besitzt, also nicht „zufällig“ ist.
Die geniale Idee von Roth war es, in diesem Fall ein „Verdichtungsargument“ zu nutzen: Wie wir gleich sehen werden, lässt sich die Existenz eines großen Fourier-Koeffizienten nutzen um zu zeigen, dass A nicht besonders gleichmäßig über verschiedene Restklassen verteilt ist. Insbesondere muss es dann eine Restklasse geben, in der A eine deutlich höhere Dichte als δ hat. Wir erhalten mithin eine kleinere, aber „dichtere“ Menge, die ebenfalls keine 3-AP enthält. Mit dieser können wir das Argument nun wiederholen und so nach endlich vielen Schritten einen Widerspruch erhalten, da die Dichte nie größer als 1 werden kann.
Wir setzen nun den ersten Schritt dieser Strategie um:

Lemma 5.8. *Sei $A \subset \{1, 2, \dots, N\}$ eine Teilmenge mit $|A| = \delta N$, die keine 3-AP enthält. Gilt $\delta > \frac{512}{N^{1/6}}$, so gibt es eine arithmetische Progression P mit $|P| \geq N^{1/3}$, sodass $A \cap P$ mindestens $(\delta + \delta^2/32)|P|$ Elemente enthält.*

Beweis. Wir wählen $p \in (2N, 4N)$ und betten A in $\mathbb{Z}/p\mathbb{Z}$ ein. Sei δ_p die Dichte von dieser eingebetteten Menge, also $\delta_p p = |A| = \delta N$. Insbesondere ist $\delta > \frac{8}{\sqrt{N}}$, also $\delta_p > \frac{2}{\sqrt{p}}$. Nach Annahme enthält A nur die trivialen 3-AP, also $\delta_p p < \frac{1}{2}\delta_p^3 p^2$ viele. Nach Lemma 5.7 muss es also ein $r \neq 0$ geben mit $|\widehat{A}(r)| > \frac{\delta_p^2}{2} p > \frac{\delta^2}{8} p$.
Nach der Definition der Fourierkoeffizienten bedeutet das aber

$$\left| \sum_{x \in A} e(\alpha x) \right| > \frac{\delta^2}{8} p$$

für $\alpha = r/p$. Wir wollen nun eine bessere rationale Approximation für α bekommen. Dazu verwenden wir den Approximationssatz von Dirichlet (aus den Übungen), der uns für einen Parameter $Q \geq 1$ eine Approximation $\frac{a}{q}$ von α gibt mit $q \leq Q$ und $\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{qQ}$.
Wir wählen $Q = N^{1/2}$. Wir betrachten nun für dieses $q \leq N^{1/2}$ eine Partition der Menge $\{1, 2, \ldots, p\}$ in arithmetische Progressionen P_i mit gemeinsamer Differenz q und Längen $N^{1/3} \leq |P_i| \leq 2N^{1/3}$.
Der Punkt dieser Konstruktion ist, dass $e(\alpha x)$ nun auf jedem dieser P_i ungefähr konstant ist. Ist nämlich $x, y \in P_i$, so gilt $|x - y| = qm$ für ein $m \leq |P_i|$ und damit

$$|e(\alpha x) - e(\alpha y)| = |e(\alpha |x - y|) - 1| = |e(qm\alpha - am) - 1| \leq 2\pi m |q\alpha - a| \leq \frac{4\pi}{N^{1/6}} < \frac{\delta}{32}.$$

Wählen wir ein festes $x_i \in P_i$, so folgt

$$\left| \sum_{x \in P_i \cap A} e(\alpha x) - e(\alpha x_i) |P_i \cap A| \right| \leq |P_i \cap A| \cdot \frac{\delta}{32}$$

und damit

$$\left| \sum_i e(\alpha x_i) |P_i \cap A| \right| \geq \frac{\delta^2}{8} p - \frac{\delta^2}{32} p.$$

Andererseits ist

$$\left| \sum_{x \in P_i} e(\alpha x) - e(\alpha x_i) |P_i| \right| \leq |P_i| \cdot \frac{\delta}{32}$$

und damit wegen

$$\sum_i \sum_{x \in P_i} e(\alpha x) = \sum_{x=1}^{p} e(\alpha x) = 0$$

nach Orthogonalität (hier benutzen wir $r \neq 0$!) sicherlich

$$\left| \sum_i e(\alpha x_i) |P_i| \right| \leq \frac{\delta}{32} p.$$

Schließlich ergibt sich durch Kombination

$$\left| \sum_i e(\alpha x_i) \left(|P_i \cap A| - \delta |P_i| \right) \right| \geq \frac{\delta^2}{16} p$$

und damit nach der Dreiecksungleichung

$$\sum_i \left| |P_i \cap A| - \delta |P_i| \right| \geq \frac{\delta^2}{16} p.$$

Daraus folgt nun, dass A auf einer der P_i eine höhere Dichte haben muss, z.B. mit folgendem Trick: Schreibe $\Delta_i = |P_i \cap A| - \delta|P_i|$. Wegen $\sum_i \Delta_i = 0$ ist dann

$$\sum_i |\Delta_i| = \sum_i (|\Delta_i| + \Delta_i) = 2 \sum_{i:\Delta_i \geq 0} \Delta_i.$$

Wäre $\Delta_i < \frac{\delta^2}{32}|P_i|$ für alle i, wäre diese Summe aber zu klein. Also muss es ein i mit $\Delta_i > \frac{\delta^2}{16}|P_i|$ geben und das ist exakt die Behauptung. □

Nun ist der Beweis des Satzes von Roth nur noch eine Rechenaufgabe:

Beweis von Satz 5.2. Die arithmetische Progression P aus dem Verdichtungslemma können wir mit der Menge $\{1, 2, \ldots, |P|\}$ identifizieren. Bei dieser Identifikation wird $P \cap A$ zu einer Teilmenge von $\{1, 2, \ldots, |P|\}$ ohne 3-AP (hier benutzen wir entscheidend, dass eine 3-AP invariant unter Verschiebung und Skalierung ist!).
Aus einer Teilmenge von $\{1, 2, \ldots, N\}$ mit Dichte δ haben wir also eine Teilmenge von $\{1, 2, \ldots, N'\}$ für ein $N' \geq N^{1/3}$ mit Dichte $\delta' \geq \delta + \delta^2/32$ konstruiert.
Iterieren wir diese Verdichtung, erhalten wir eine Folge $N = N_1 > N_2 > \cdots > N_k$ mit $N_{i+1} \geq N_i^{1/3}$ und Dichten $\delta = \delta_1 < \delta_2 < \cdots < \delta_k$ mit $\delta_{i+1} \geq \delta_i + \delta_i^2/32$.
Nach $\leq \lceil \frac{32}{\delta} \rceil$ Schritten hat sich die Dichte mithin zumindestens 2δ verdoppelt, allgemein nach $\lceil \frac{2^{6-k}}{\delta} \rceil$ Schritten k-mal verdoppelt.
Sicherlich kann sich die Dichte höchstens $\log_2(1/\delta)$ mal verdoppeln. Insgesamt werden damit maximal $k \ll \frac{1}{\delta}$ viele Schritte gemacht.
Endet die Iteration, muss aber die Bedingung $\delta < \delta_k < \frac{512}{N_k^{1/6}}$ gelten. Wegen $N_k \geq N^{1/3^{k-1}}$ bedeutet dies $N^{1/(2 \cdot 3^k)} \ll \frac{1}{\delta}$ und damit $\log N \ll 3^k \cdot \log(1/\delta)$, also $\log\log N \ll \frac{1}{\delta}$ und damit $\delta \ll \frac{1}{\log\log N}$. □

6 Szemerédi: Ein Stein von Rosette für die moderne Mathematik

Nachdem wir den Satz von Roth bewiesen haben, wollen wir nun noch den Satz von Szemerédi diskutieren. Auch wenn es vollkommen unmöglich ist, in diesem Rahmen einen vollständigen Beweis zu geben, ist es interessant, die sehr unterschiedlichen mathematischen Methoden zu vergleichen, mit denen verschiedene Mathematiker in den letzten 50 Jahren den Satz bewiesen haben. Dies gliedert sich im Wesentlichen in die drei folgenden Bereiche:

- der graphentheoretische Zugang über das Regularitätslemma von Szemerédi (und dessen Varianten für Hypergraphen), ca. 1975
- der ergodentheoretische Zugang über das Korrespondenzprinzip von Fürstenberg, ca. 1977
- der Beweis via Fourieranalysis „höherer Ordnung“ von Gowers, ca. 2001

Die Beweise haben unterschiedliche Vor- und Nachteile. Der ergodentheoretische Zugang ist sehr robust und erlaubt es, Varianten zu zeigen, bei denen statt einer 3-AP $\{x, x+y, x+2y\}$ Muster wie $\{x, x+y^2\}$ untersucht werden (in vielen Fällen gibt es für diese Aussagen bisher keine Beweise ohne Ergodentheorie). Dagegen liefert dieser Zugang zunächst keine und der graphentheoretische nur sehr schlechte quantitative Ergebnisse über $r_k(N)$. Hier liegen vor allem die Vorzüge des fourieranalytischen Ansatzes, der die bisher besten bekannten Schranken produziert.
Wie Tao mit seinem Ausspruch vom „Stein von Rosette“ der modernen Mathematik betont hat, liefert aber dieser Satz als Schnittstelle der zunächst sehr unterschiedlichen Gebiete (Kombinatorik, Dynamik, Analysis) interessante Analogien, die auch für unabhängige Resultate in jedem dieser Gebiete bereits entscheidende Einflüsse und neue Ideen gegeben haben.

Wir wollen nun kurz auf die drei unterschiedlichen Ansätze eingehen. Der gemeinsame Nenner wird in jedem Fall eine gewisse Dichotomie zwischen *Struktur* und *Zufälligkeit* sein, die wir bereits im Beweis des Satzes von Roth gesehen haben.

6.1 Graphentheorie

Ruzsa und Szemerédi haben 1978 beobachtet, dass sich der Satz von Roth auf das folgende graphentheoretische Ergebnis zurückführen lässt:

Satz 6.1 (Dreiecks-Entfernungs-Lemma, Ruzsa-Szemerédi 1978)**.** *Für jedes $\varepsilon > 0$ gibt es ein $\delta > 0$ mit der folgenden Eigenschaft: Hat ein Graph auf n Knoten höchstens δn^3 Dreiecke, dann kann man ihn durch Entfernen von höchstens εn^2 Kanten dreiecksfrei machen.*

Dieses Lemma ist wiederum eine Folgerung aus dem *Regularitätslemma von Szemerédi*, dass grob vereinfacht sagt: Die Ecken eines beliebigen Graphen lassen sich so in *wenige* Mengen aufteilen, dass die Kanten zwischen verschiedenen dieser Mengen eine *zufällige* Struktur haben. (Für unsere Anwendung bedeutet diese Zufälligkeit dann: Gibt es drei dieser Mengen, zwischen denen es ein Dreieck gibt, dann gleich $\gg n^3$ viele, was ausgeschlossen war.)

Beweis des Satzes von Roth mit dem Dreiecks-Entfernungs-Lemma. Gegeben sei eine Menge $A \subset \mathbb{Z}/p\mathbb{Z}$ für $p > 2$ ohne (nicht-triviale) 3-AP. Wir betrachten einen Graphen mit Knotenmenge $X \sqcup Y \sqcup Z$, wobei X, Y, Z jeweils Kopien von $\mathbb{Z}/p\mathbb{Z}$ ist. Der Graph hat also $3p$ Knoten.
Weiter zeichnen wir eine Kante $(x, y) \in X \times Y$, falls $x - y \in A$ ist, eine Kante $(y, z) \in Y \times Z$, falls $y - z \in A$ ist, sowie eine Kante $(x, z) \in X \times Z$, falls $\frac{x-z}{2} \in A$ ist (dies ist modulo p wohldefiniert).
Nach Konstruktion bildet nun (x, y, z) genau dann ein Dreieck, wenn die zugehörigen Differenzen $x - y, y - z, \frac{x-z}{2}$ eine $3 - AP$ in A bilden.
Nach Annahme gibt es somit genau $|A|p \leq p^2$ Dreiecke, nämlich genau die, die zu den trivialen 3-AP gehören. Weiterhin ist jede der $3|A|p$ Kanten in genau einem Dreieck enthalten.
Für jedes feste δ und hinreichend großes p gibt es nun maximal $p^2 < \delta(3p)^3$ Dreiecke, also lässt sich nach dem Dreiecks-Entfernungs-Lemma für jedes $\varepsilon > 0$ und hinreichend großes n der Graph durch Entfernen von maximal εp^2 Kanten dreiecksfrei machen. Nach Konstruktion müssen wir aber $|A|p$ Kanten entfernen, um alle Dreiecke zu entfernen. Es folgt $|A|p \leq \varepsilon p^2$ und damit $|A| \leq \varepsilon p$, d.h. A hat beliebig kleine Dichte, falls p hinreichend groß ist.
□

Mit einer (sehr komplizierten) Version des Regularitätslemmas für Hypergraphen kann man dann auch ein „Hypergraph-Entfernungs-Lemma“ und damit als Korollar den Satz von Szemerédi folgern.
Ein Problem des Regularitätslemmas für quantitative Überlegungen ist, dass zumindest in seinen ursprünglichen Beweis der Satz von van der Waerden eingeht, was natürlich ungünstig ist, wenn man als Folgerung die Schranke im Satz von van der Waerden verbessern möchte. Inzwischen gibt es zwar andere Beweise, aber die Schranken sind immer noch sehr schlecht.

6.2 Ergodentheorie

Die Ergodentheorie ist ein komplett eigenständiger Bereich der Mathematik, den wir hier nur knapp zusammenfassen können: Die Idee ist, das Verhalten von iterierten Abbildungen auf bestimmten Räumen (die man sich geometrisch oder rein algebraisch vorstellen kann) zu untersuchen.

Definition 6.2. *Ein **dynamisches System** ist ein Wahrscheinlichkeitsraum $(X, \mathcal{B}, \mu)$ zusammen mit einer maßerhaltenden Abbildung $T : X \to X$, d.h. $\mu(T^{-1}A) = \mu(A)$ für alle $A \in \mathcal{B}$.*

Beispiele:

- $X = S^1 = \mathbb{R}/\mathbb{Z}$ mit dem Lebesguemaß μ und der Abbildung $T(x) = 2x \pmod 1$. Diese Abbildung führt bei iterierter Betrachtung von Urbildern schnell zu chaotischem Verhalten, man nennt sie deswegen auch *mischend*: Beispielsweise gilt $\mu(T^{-n}A \cap B) \to \mu(A)\mu(B)$, d.h. die Wahrscheinlichkeiten, bei B und n Schritte später in A zu landen, sind asymptotisch unabhängig voneinander.
- $X = S^1 = \mathbb{R}/\mathbb{Z}$ mit dem Lebesguemaß μ und der Abbildung $T(x) = x + \alpha \pmod 1$ für eine Konstante α, geometrisch also eine *Drehung*. Diese Abbildung ist sehr rigide und überhaupt nicht chaotisch. Über $\mu(T^{-n}A \cap B)$ für beliebige $A, B \in \mathcal{B}$ lässt sich entsprechend auch wenig aussagen.

Obwohl sich diese Systeme sehr unterschiedlich verhalten, gibt es Aussagen, die in beiden Fällen zutreffen, wie etwa der Poincarésche Wiederkehrsatz, der für beliebige dynamische Systeme gilt: Für beliebiges $A \in \mathcal{B}$ mit $\mu(A) > 0$ ist $\mu(A \cap T^{-n}A) > 0$ unendlich oft. Man kann dies (bzw. eine Folgerung daraus) so interpretieren, dass fast jeder Punkt aus A unendlich oft nach A zurückkehrt, egal wie klein A ist. Während dies sehr elementar zu zeigen ist, hat Hillel Fürstenberg 1977 eine weitreichende Verallgemeinerung aufgestellt:

Satz 6.3 (Fürstenbergs multiple Wiederkehr)**.** *Für jedes dynamische System $(X, \mathcal{B}, \mu, T)$ und jedes $A \in \mathcal{B}$ mit $\mu(A) > 0$ sowie $k \in \mathbb{N}$ gilt*

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-(k-1)n}A) > 0$$

unendlich oft.

Für $k = 2$ ist dies natürlich einfach der Satz von Poincaré. Fürstenbergs Beobachtung war, dass dies leicht für *mischende Systeme* wie die Verdopplungsabbildung $x \mapsto 2x$ auf dem Torus, aber *auch* für *kompakte Systeme* wie die Rotation auf dem Torus ist! Mit einer weitreichenden Strukturtheorie dynamischer Systeme zeigte er dann, dass sich in einem gewissen Sinn jedes dynamische System aus diesen beiden Grundtypen (*Struktur* und *Zufall*!) zusammenbauen lässt und konnte somit den multiplen Wiederkehrsatz zeigen.
Könnten wir den Satz auf das dynamische Sytem $T : \mathbb{Z} \to \mathbb{Z}, x \mapsto x + 1$ mit einem gleichverteilten Wahrscheinlichkeitsmaß anwenden, würde er sofort den Satz von Szemerédi implizieren. Leider gibt es solch ein Wahrscheinlichkeitsmaß auf $\mathbb{Z}$ nicht, dennoch lässt sich ohne großen Aufwand zeigen:

Satz 6.4 (Fürstenbergsches Korrespondenzprinzip)**.** *Multiple Wiederkehr ist äquivalent zum Satz von Szemerédi.*

Dieser Satz wird ein *Prinzip* genannt, weil er sehr robust ist. Beispielsweise lässt sich die Existenz von Mustern wie $\{x, x + y^2\}$ analog übersetzen in ein Wiederkehrresultat für $\mu(A \cap T^{-n^2}A)$.

6.3 Fourieranalysis höherer Ordnung

Wir wollen zunächst unsere Beweisstrategie für den Satz von Roth etwas umformulieren: Für eine Funktion $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ betrachten wir die Fourier-Norm $\|\widehat{f}\|_\infty := \max_r |\widehat{f}(r)|$. Dann lässt sich unser Beweis wie folgt strukturieren. Für $A \subset \mathbb{Z}/p\mathbb{Z}$ mit Dichte δ gilt:

i) (*Zufall*) Hat die normalisierte Indikatorfunktion $1_A - \delta$ kleine Fourier-Norm, so hat A in etwa die erwartete Anzahl $\delta^3 p^2$ an 3-AP (insbesondere mindestens eine).

ii) (*Struktur*) Hat $1_A - \delta$ große Fourier-Norm, so hat sie eine große Korrelation mit einem linearen Charakter χ_r und dies impliziert, dass A auf einer AP erhöhte Dichte besitzt.

Die gleiche Strategie lässt sich nicht ohne weiteres für längere Progressionen umsetzen, da es nicht stimmt, dass die Fourier-Norm auch die Anzahl längerer k-APs kontrolliert.
Tim Gowers konnte aber für jedes k eine Norm $\|f\|_{U^{k-1}}$ konstruieren (die wir heute *Gowers uniformity norm* nennen) und zeigen, dass diese die Anzahl von k-AP kontrolliert:

i) (*Zufall*) Hat $1_A - \delta$ kleine U_{k-1}-Norm, so enthält A in etwa die erwartete Anzahl an k-AP (insbesondere mindestens eine).

Der schwierigere Teil war zu untersuchen, was im anderen Fall passiert. Dies konnte Gowers zunächst nicht vollständig lösen, aber 2012 wurde die *inverse Vermutung für die Gowers-Normen* vollständig bewiesen (von Green, Tao und der israelischen Mathematikerin Tamar Ziegler), sodass wir nun wissen:

ii) (*Struktur*) Hat $1_A - \delta$ große U_{k-1}-Norm, so hat sie eine große Korrelation mit einer *Nilfolge* von Schrittweite k, was grob gesagt eine Verallgemeinerung von Folgen der Art $e(-P(x))$ für ein Polynom P von Grad k ist. Da Nilfolgen gleichverteilt in Restklassen sind, impliziert dies, dass A auf einer AP erhöhte Dichte besitzt.

Aus einer verbesserten quantitativen Version dieses *inversen Theorems* (große Gowers-Norm impliziert Korrelation mit Nilfolge) konnten die drei Doktoranden (!) James Leng, Ashwin Sah und Mehtaab Sawhney dann 2024 in einem großen Durchbruch die Schranke **fehlende Klammer???**

$$r_k(N) \ll \frac{N}{\exp((\log\log N)^{c_k}}$$

für ein gewisses $c_k \in (0,1)$ und alle $k \geq 3$ zeigen, die erste wesentliche Verbesserung von Gowers' Schranke $r_k(N) \ll \frac{N}{(\log\log N)^{c_k}}$ für $k \geq 5$ in über 20 Jahren.

7 Das Lemma von Balog, Szemerédi und Gowers

Zum Abschluss kehren wir in diesem Abschnitt noch einmal zu den Untersuchungen von Summenmengen zurück, und dem Zusammenhang zur **Energie**.

Definition und Satz 7.1. *Sei $A \subset \mathbb{Z}$ endlich und nichtleer. Die **Energie** von A ist definiert als*

$$E(A) = \#\{(a_1, a_2, a_3, a_4) : a_1 + a_2 = a_3 + a_4\}.$$

Es gilt $|A|^2 \leq E(A) \leq |A|^3$. Weiter gilt

$$|A + A| \geq \frac{|A|^4}{E(A)}.$$

Insbesondere muss eine Menge mit kleiner Energie stets eine große Summenmenge haben, was wir mehrmals benutzt haben. Wir haben allerdings auch gesehen, dass die Umkehrung nicht stimmt: Eine Menge kann durchaus große Energie und dennoch große Summenmenge haben. Ein typisches Beispiel dafür ist eine Vereinigung von zwei etwa gleich großen Mengen, von denen eine große Energie (etwa eine AP) und die andere große Summenmenge (etwa eine zufällige Menge) hat, sodass sich beides auf die Vereinigung überträgt.
Der letzte wichtige Meilenstein der additiven Kombinatorik, den wir in dieser Vorlesung kennenlernen werden, sagt, dass dies in gewissem Sinne auch das einzige ist, was passieren kann. Die Aussage wurde 1994 von Balog und Szemerédi bewiesen und 2001 von Gowers quantitativ verbessert.

Satz 7.2 (Balog-Szemerédi-Gowers-Lemma)**.** *Sei $A \subset \mathbb{Z}$ endlich und nichtleer und $K > 0$ mit $E(A) \geq \frac{|A|^3}{K}$. Dann gibt es $A' \subset A$ mit $|A'| \gg_K |A|$ sowie $|A' - A'| \ll_K |A|$.*

Mit anderen Worten muss eine Menge mit großer Energie zwar selbst nicht kleine Summenmenge haben (also additive Struktur besitzen), wohl aber eine große Teilmenge von ihr. Dies ist für viele Anwendungen bereits sehr nützlich, insbesondere in Kombination mit dem Satz von Freiman, der Mengen mit kleiner Summenmenge charakterisiert.
Wir folgen dem Beweis von Tomasz Schoen in der Darstellung von Tom Bloom:

Lemma 7.3. *Ist* $E(A) \geq \frac{|A|^3}{K}$*, dann gibt es eine Menge* $X \subset A$ *mit* $|X| \gg_K |A|$*, sodass*

$$r_{A-A}(a-b) \gg_K |A|$$

für mindestens $0.998|X|^2$ *der Paare* $(a,b) \in X^2$ *gilt.*

Hier lässt sich 0.998 durch eine beliebige Konstante kleiner als 1 ersetzen.

Beweis. Wir konstruieren die Menge X probabilistisch. Dazu wählen wir $X = A \cap (A+s)$, wobei wir s zufällig mit Wahrscheinlichkeit $\frac{r_{A-A}(s)}{|A|^2}$ wählen. Dann ist $|A| = r_{A-A}(s)$, die erwartete Größe von X ist folglich

$$\mathbb{E}|X| = \frac{1}{|A|^2} \sum_s r_{A-A}(s)^2 = \frac{E(A)}{|A|^2}$$

und damit nach Cauchy-Schwarz auch $\mathbb{E}|X|^2 \geq (\mathbb{E}|X|)^2 \geq \frac{E(A)^2}{|A|^4}$. Sei

$$G = \{(a,b) \in A^2 : r_{A-A}(a-b) \leq 0.001 \frac{E(A)^2}{|A|^5}\}$$

die Menge der „schlechten“ Paare (mit wenigen Darstellungen). Dann ist

$$\begin{aligned}
\mathbb{E}|X^2 \cap G| &= \sum_{a,b \in G} \mathbb{P}(a,b \in X) \\
&= \frac{1}{|A|^2} \sum_{(a,b) \in G} \sum_{s : a-s, b-s \in A} r_{A-A}(s) \\
&\leq \frac{1}{|A|} \sum_{(a,b) \in G} \#\{s : a-s, b-s \in A\} \\
&= \frac{1}{|A|} \sum_{(a,b) \in G} r_{A-A}(a-b) \\
&\leq 0.001 \frac{E(A)^2}{|A|^5} \cdot \frac{|G|}{|A|} \\
&\leq 0.001 \frac{E(A)^2}{|A|^4}.
\end{aligned}$$

Damit ist

$$\mathbb{E}(|X|^2 - 500|X^2 \cap G|) \geq \frac{1}{2} \frac{E(A)^2}{|A|^4},$$

folglich gibt es ein X mit

$$|X|^2 - 500|X^2 \cap G| \geq \frac{1}{2} \frac{E(A)^2}{|A|^4}.$$

Insbesondere ist dann $|X| \gg \frac{E(A)}{|A|^2} \gg |A|$ und $|X^2 \cap G| \leq 0.002|X|^2$, für die übrigen mindestens $0.998|X|^2$ Paare $(a,b) \in X^2$ gilt mithin $r_{A-A}(a-b) \gg \frac{E(A)^2}{|A|^5} \gg |A|$. □

Nun müssen wir die Menge X nur noch etwas modifizieren, um die Menge A' aus dem BSG-Lemma zu erhalten:

Beweis von Satz 7.2. Wir betrachten die Menge X aus dem Lemma und konstruieren einen Graphen mit Knotenmenge X und einer Kante zwischen a und b, falls $a \neq b$ ist und $r_{A-A}(a-b) \gg |A|$ gilt (mit der gleichen impliziten Konstante wie in der Konstruktion von X). Dann hat der Graph mindestens $\frac{7}{16}|X|^2$ Kanten. Wir wählen nun A' als die Menge der $x \in X$, die Grad mindestens $\frac{3}{4}|X|$ in diesem Graphen haben. Da die Summe aller Grade mindestens $\frac{7}{8}|X|^2$ ist und die Knoten außerhalb von A' insgesamt höchstens $\frac{3}{4}|X|^2$ dazu beitragen, haben die Knoten in A' eine Summe der Grade von mindestens $\frac{1}{8}|X|^2$, insbesondere ist $|A'| \gg |X|$.
Für alle $a, b \in A'$ gibt es nun mindestens $\frac{|X|}{2} \gg |A|$ Elemente $c \in X$ mit $r_{A-A}(a-c), r_{A-A}(b-c) \gg |A|$. Damit hat $a-b = (a-c)-(b-c)$ mindestens $\gg |A|^3$ Darstellungen als $a_1-a_2+a_3-a_4$ mit $a_i \in A$. Aber insgesamt gibt es nur $|A|^4$ solche Darstellungen, also kann es höchstens $\ll |A|$ Elemente der Form $a-b \in A'-A'$ geben, d.h. $|A'-A'| \ll |A|$. □

Zum Abschluss wollen wir einen Satz beweisen, in dessen Beweis die meisten wichtigen Resultate dieser Vorlesung als Bausteine eingehen:

Satz 7.4. *Ist $A \subset \mathbb{Z}$ endlich und hinreichend groß und besitzt mindestens $0.00001|A|^2$ viele 3-APs, dann enthält A eine 1000-AP.*

Natürlich stehen die Konstanten hier repräsentativ für beliebig kleine bzw. beliebig große Konstanten.

Beweis. Betten wir A in ein geeignetes $\mathbb{Z}/p\mathbb{Z}$ ein, so ist die Anzahl der 3-AP (wobei wir auch triviale mitzählen) wie im Beweis des Satzes von Roth gleich

$$\frac{1}{p}\sum_r \widehat{A}(r)^2\widehat{A}(-2r).$$

Nach der Cauchy-Schwarz-Ungleichung ist dies nach oben beschränkt durch

$$\left(\frac{1}{p}\sum_r \left|\widehat{A}(r)\right|^4\right)^{1/2}\left(\frac{1}{p}\sum_r \left|\widehat{A}(r)\right|^2\right)^{1/2} = E(A)^{1/2}|A|^{1/2}.$$

Aus der Annahme folgt mithin $E(A) \gg |A|^3$. Nach dem Lemma von Balog, Szemerédi und Gowers muss es dann ein $A' \subset A$ mit $|A'| \gg |A|$ sowie $|A'-A'| \ll |A'|$ geben. Nach dem Satz von Freiman gibt es dann eine VAP P von Dimension $d \ll 1$ und mit $|P| \ll |A|$, die A' enthält. Insbesondere ist einer der d „Seitenlängen“ von P mindestens $\gg |A|^{1/d}$. Wir können somit P in arithmetische Progressionen der Länge $\gg |A|^{1/d}$ partitionieren. Nach dem Schubfachprinzip muss eine dieser Progressionen Q mindestens $\frac{|Q|}{|P|} \cdot |A'| \gg |Q|$ viele Elemente von A' enthalten. Mit anderen Worten hat A' in Q eine positive Dichte, enthält also nach dem Satz von Szemerédi eine 1000-AP. □