The Balog–Szemerédi–Gowers Theorem

Angeliki Koutsoukou-Argyraki, Mantas Bakšys, and Chelsea Edmonds
University of Cambridge
{ak2110, mb2412, cle47}@cam.ac.uk

March 17, 2025

Abstract

We formalise the Balog–Szemerédi–Gowers Theorem, a profound result in additive combinatorics which played a central role in Gowers's proof deriving the first effective bounds for Szemerédi's Theorem [2]. The proof is of great mathematical interest given that it involves an interplay between different mathematical areas, namely applications of graph theory and probability theory to additive combinatorics involving algebraic objects. This interplay is what made the process of the formalisation, for which we had to develop formalisations of new background material in the aforementioned areas, more rich and technically challenging. We demonstrate how locales, Isabelle's module system, can be employed to handle such interplays. To treat the graph-theoretic aspects of the proof, we make use of a new, more general undirected graph theory library developed recently by Chelsea Edmonds, which is both flexible and extensible [1]. For the formalisation we followed a proof presented in the 2022 lecture notes by Timothy Gowers "Introduction to Additive Combinatorics" for Part III of the Mathematical Tripos taught at the University of Cambridge [3]. In addition to the main theorem, which, following our source, is formulated for difference sets, we also give an alternative version for sumsets which required a formalisation of an auxiliary triangle inequality following a proof by Yufei Zhao from his book "Graph Theory and Additive Combinatorics" [4]. We moreover formalise a few additional results in additive combinatorics that are not used in the proof of the main theorem. This is the first formalisation of the Balog–Szemerédi–Gowers Theorem in any proof assistant to our knowledge.

Contents

Acknowledgements

Angeliki Koutsoukou-Argyraki is funded by the ERC Advanced Grant ALEXANDRIA (Project GA 742178) funded by the European Research Council and led by Lawrence C. Paulson (University of Cambridge, Department of Computer Science and Technology). Mantas Bakšys received funding for his internship supervised by Koutsoukou-Argyraki by the Cambridge Mathematics Placements (CMP) Programme and by the ALEXANDRIA Project. Chelsea Edmonds is jointly funded by the Cambridge Trust (Cambridge Australia Scholarship) and a Cambridge Department of Computer Science and Technology Premium Research Studentship.

1 Miscellaneous technical lemmas

theory *Miscellaneous-Lemmas*
imports
HOL−Library.Indicator-Function
HOL−Analysis.Convex
begin

lemma *set-pairs-filter-subset*: $A \subseteq B \Longrightarrow \{p\ .\ p \in A \times A \wedge P\ p\} \subseteq \{p.\ p \in B \times B \wedge P\ p\}$
⟨*proof*⟩

lemma *card-set-ss-indicator*:
assumes $A \subseteq B$
assumes *finite* B
shows *card* $A = (\sum\ p \in B.\ indicator\ A\ p)$
⟨*proof*⟩

lemma *card-cartesian-prod-square*: *finite* $X \Longrightarrow card\ (X \times X) = (card\ X)\hat{}2$
⟨*proof*⟩

lemma (**in** *ordered-ab-group-add*) *diff-strict1-mono*:
assumes $a > a'\ b \le b'$
shows $a - b > a' - b'$
⟨*proof*⟩

lemma *card-cartesian-product-6*: $card\ (A \times A \times A \times A \times A \times A) = (card\ A)\ \hat{}\ 6$
⟨*proof*⟩

lemma *card-cartesian-product3*: $card\ (X \times Y \times Z) = card\ X * card\ Y * card\ Z$
⟨*proof*⟩

lemma *card-le-image-div*:
fixes A:: *'a set* **and** B:: *'b set* **and** f:: *'a ⇒ 'b set* **and** r:: *real*
assumes *finite* B **and** *pairwise* $(\lambda\ s\ t.\ disjnt\ (f\ s)\ (f\ t))\ A$ **and** $\forall\ d \in A.\ (card\ (f\ d)) \ge r$
and $\forall\ d \in A.\ f\ d \subseteq B$ **and** $r > 0$
shows $card\ A \le card\ B\ /\ r$

⟨*proof*⟩

lemma *list-middle-eq*:
length xs = length ys ⟹ hd xs = hd ys ⟹ last xs = last ys
⟹ butlast (tl xs) = butlast (tl ys) ⟹ xs = ys
⟨*proof*⟩

lemma *list2-middle-singleton*:

assumes *length xs* = *3*
shows *butlast* (*tl xs*) = [*xs* ! *1*]
⟨*proof*⟩

lemma *le-powr-half-mult*:
fixes *x y z*:: *real*
assumes $x \,\hat{}\, 2 \leq y * z$ **and** $0 \leq y$ **and** $0 \leq z$
shows $x \leq y$ *powr*$(1/2) * z$ *powr* $(1/2)$
⟨*proof*⟩

lemma *Cauchy-Schwarz-ineq-sum2*:
fixes *f g*:: *'a* ⇒ *real* **and** *A*:: *'a set*
shows $(\sum d \in A.\ f\ d * g\ d) \leq$
$(\sum d \in A.\ (f\ d)\hat{}2)$ *powr* $(1/2) * (\sum d \in A.\ (g\ d)\hat{}2)$ *powr* $(1/2)$
⟨*proof*⟩

end

2 Background material for the graph-theoretic aspects of the main proof

This section includes a number of lemmas on project specific definitions for graph theory, building on the general undirected graph theory library [1]

theory *Graph-Theory-Preliminaries*
imports
Miscellaneous-Lemmas
Undirected-Graph-Theory.Bipartite-Graphs
Undirected-Graph-Theory.Connectivity
Random-Graph-Subgraph-Threshold.Ugraph-Misc
begin

2.1 On graphs with loops

context *ulgraph*

begin

definition *degree-normalized*:: *'a* ⇒ *'a set* ⇒ *real* **where**
degree-normalized v S ≡ *card* (*neighbors-ss v S*) / (*card S*)

lemma *degree-normalized-le-1*: *degree-normalized x S* ≤ *1*

⟨*proof*⟩

end

2.2 On bipartite graphs

context *bipartite-graph*
begin

definition *codegree*:: $'a \Rightarrow 'a \Rightarrow nat$ **where**
codegree v u $\equiv$ *card* $\{x \in V$. *vert-adj v x* $\wedge$ *vert-adj u x*$\}$

lemma *codegree-neighbors*: *codegree v u* = *card* (*neighborhood v* $\cap$ *neighborhood u*)
⟨*proof*⟩

lemma *codegree-sym*: *codegree v u* = *codegree u v*
⟨*proof*⟩

definition *codegree-normalized*:: $'a \Rightarrow 'a \Rightarrow 'a$ *set* $\Rightarrow$ *real* **where**
codegree-normalized v u S $\equiv$ *codegree v u* / *card S*

lemma *codegree-normalized-altX*:
assumes $x \in X$ **and** $x' \in X$
shows *codegree-normalized* $x\ x'\ Y$ = *card* (*neighbors-ss* $x\ Y \cap$ *neighbors-ss* $x'\ Y$) / *card* Y

⟨*proof*⟩

lemma *codegree-normalized-altY*:
assumes $y \in Y$ **and** $y' \in Y$
shows *codegree-normalized* $y\ y'\ X$ = *card* (*neighbors-ss* $y\ X \cap$ *neighbors-ss* $y'\ X$) / *card* X

⟨*proof*⟩

lemma *codegree-normalized-sym*: *codegree-normalized u v S* = *codegree-normalized v u S*
⟨*proof*⟩

definition *bad-pair*:: $'a \Rightarrow 'a \Rightarrow 'a$ *set* $\Rightarrow$ *real* $\Rightarrow$ *bool* **where**
bad-pair v u S c $\equiv$ *codegree-normalized v u S* $< c$

lemma *bad-pair-sym*:
assumes *bad-pair v u S c* **shows** *bad-pair u v S c*
⟨*proof*⟩

definition *bad-pair-set*:: $'a$ *set* $\Rightarrow 'a$ *set* $\Rightarrow$ *real* $\Rightarrow$ $('a \times 'a)$ *set* **where**
bad-pair-set S T c $\equiv \{(u, v) \in S \times S$. *bad-pair u v T c*$\}$

lemma *bad-pair-set-ss*: *bad-pair-set S T c* $\subseteq S \times S$
⟨*proof*⟩

lemma *bad-pair-set-filter-alt*:

bad-pair-set S T c = *Set.filter* (λ *p* . *bad-pair* (*fst p*) (*snd p*) *T c*) (*S* × *S*)
⟨*proof*⟩

lemma *bad-pair-set-finite*:
assumes *finite S*
shows *finite* (*bad-pair-set S T c*)
⟨*proof*⟩

lemma *codegree-is-path-length-two*:
codegree x x′ = *card* {*p* . *connecting-path x x′ p* ∧ *walk-length p* = *2*}
⟨*proof*⟩

lemma *codegree-bipartite-eq*:
∀ *x* ∈ *X*. ∀ *x′* ∈ *X*. *codegree x x′* = *card* {*y* ∈ *Y*. *vert-adj x y* ∧ *vert-adj x′ y*}
⟨*proof*⟩

lemma (**in** *fin-bipartite-graph*) *bipartite-deg-square-eq*:
∀ *y* ∈ *Y*. (∑ *x′* ∈ *X*. ∑ *x* ∈ *X*. *indicator* {*z*. *vert-adj x z* ∧ *vert-adj x′ z*} *y*)
= (*degree y*)^*2*
⟨*proof*⟩

lemma (**in** *fin-bipartite-graph*) *codegree-degree*:
(∑ *x′* ∈ *X*. ∑ *x* ∈ *X*. (*codegree x x′*)) = (∑ *y* ∈ *Y*. (*degree y*)^*2*)

⟨*proof*⟩

lemma (**in** *fin-bipartite-graph*) *sum-degree-normalized-X-density*:
(∑ *x* ∈ *X*. *degree-normalized x Y*) / *card X* = *edge-density X Y*
⟨*proof*⟩

lemma (**in** *fin-bipartite-graph*) *sum-degree-normalized-Y-density*:
(∑ *y* ∈ *Y*. *degree-normalized y X*) / *card Y* = *edge-density X Y*
⟨*proof*⟩

end
end

3 Auxiliary probability space results

theory *Prob-Space-Lemmas*
imports
Random-Graph-Subgraph-Threshold.Prob-Lemmas
begin

context *prob-space*

begin

lemma *expectation-uniform-count*:

assumes *M* = *uniform-count-measure X* **and** *finite X*
shows *expectation f* = ($\sum$ *x* ∈ *X*. *f x*) / *card X*

⟨*proof*⟩

A lemma to obtain a value for x where the inequality is satisfied

lemma *expectation-obtains-ge*:
fixes *f* :: *'a* ⇒ *real*
assumes *M* = *uniform-count-measure X* **and** *finite X*
assumes *expectation f* ≥ *c*
obtains *x* **where** *x* ∈ *X* **and** *f x* ≥ *c*

⟨*proof*⟩

The following is the variation on the Cauchy-Schwarz inequality presented in Gowers's notes before Lemma 2.13 [3].

lemma *cauchy-schwarz-ineq-var*:
fixes *X* :: *'a* ⇒ *real*
assumes *integrable M* (λ*x*. (*X x*)^*2*) **and** *X* ∈ *borel-measurable M*
shows *expectation* (λ *x*. (*X x*)^*2*) ≥ (*expectation* (λ *x* . (*X x*)))^*2*

⟨*proof*⟩

lemma *integrable-uniform-count-measure-finite*:
fixes *g* :: *'a* ⇒ *'b*::{*banach*, *second-countable-topology*}
shows *finite A* ⟹ *integrable* (*uniform-count-measure A*) *g*
⟨*proof*⟩

lemma *cauchy-schwarz-ineq-var-uniform*:
fixes *X* :: *'a* ⇒ *real*
assumes *M* = *uniform-count-measure S*
assumes *finite S*
shows *expectation* (λ *x*. (*X x*)^*2*) ≥ (*expectation* (λ *x* . (*X x*)))^*2*

⟨*proof*⟩

An equation for expectation over a discrete random variables distribution:

lemma *expectation-finite-uniform-space*:
assumes *M* = *uniform-count-measure S* **and** *finite S*
fixes *X* :: *'a* ⇒ *real*
shows *expectation X* = ($\sum$ *y* ∈ *X* ‘ *S* . *prob* {*x* ∈ *S* . *X x* = *y*} ∗ *y*)

⟨*proof*⟩

lemma *expectation-finite-uniform-indicator*:
assumes *M* = *uniform-count-measure S* **and** *finite S*
shows *expectation* (λ *x*. *indicator* (*T x*) *y*) = *prob* {*x* ∈ *S* . *indicator* (*T x*) *y* = *1*} (**is** *expectation ?X* = -)

⟨*proof*⟩

end
end

4 A triangle inequality for sumsets

theory *Sumset-Triangle-Inequality*
imports
Pluennecke-Ruzsa-Inequality.Pluennecke-Ruzsa-Inequality
begin

context *additive-abelian-group*

begin

We show a useful triangle inequality for sumsets that does *not* follow from the Ruzsa triangle inequality. The proof follows the exposition in Zhao's book [4].

The following auxiliary lemma corresponds to Lemma 7.3.4 in Zhao's book [4].

lemma *triangle-ineq-sumsets-aux*:
fixes $X\ B\ Y :: 'a\ set$
assumes *hX*: *finite X* **and** *hB*: *finite B* **and** *hXG*: $X \subseteq G$ **and** *hBG*: $B \subseteq G$ **and**
hXne: $X \neq \{\}$ **and** *hYX*: $\bigwedge Y.\ Y \subseteq X \Longrightarrow Y \neq \{\} \Longrightarrow$ *card* (*sumset Y B*) / *card Y* $\geq$
card (*sumset X B*) / *card X* **and** *hC*: *finite C* **and** *hCne*: $C \neq \{\}$ **and** *hCG*: $C \subseteq G$
shows *card* (*sumset X* (*sumset C B*)) / *card* (*sumset X C*) $\leq$ *card* (*sumset X B*) / *card X*
⟨*proof*⟩

The following inequality is the result corresponding to Corollary 7.3.6 in Zhao's book [4].

lemma *triangle-ineq-sumsets*:
assumes *hA*: *finite A* **and** *hB*: *finite B* **and** *hC*: *finite C* **and**
hAG : $A \subseteq G$ **and** *hBG*: $B \subseteq G$ **and** *hCG*: $C \subseteq G$
shows *card A* $*$ *card* (*sumset B C*) $\leq$ *card* (*sumset A B*) $*$ *card* (*sumset A C*)

⟨*proof*⟩

end
end

5 Background material in additive combinatorics

This section outlines some background definitions and basic lemmas in additive combinatorics based on the notes by Gowers [3].

theory *Additive-Combinatorics-Preliminaries*
  imports
    Pluennecke-Ruzsa-Inequality.Pluennecke-Ruzsa-Inequality
begin

5.1 Additive quadruples and additive energy

context *additive-abelian-group*

begin

definition *additive-quadruple*:: $'a \Rightarrow 'a \Rightarrow 'a \Rightarrow 'a \Rightarrow bool$ **where**
  additive-quadruple $a\ b\ c\ d \equiv a \in G \wedge b \in G \wedge c \in G \wedge d \in G \wedge a \oplus b = c \oplus d$

lemma *additive-quadruple-aux*:
  assumes *additive-quadruple* $a\ b\ c\ d$
  shows $d = a \oplus b \ominus c$
  ⟨*proof*⟩

lemma *additive-quadruple-diff*:
  assumes *additive-quadruple* $a\ b\ c\ d$
  shows $a \ominus c = d \ominus b$
  ⟨*proof*⟩

definition *additive-quadruple-set*:: $'a\ set \Rightarrow ('a \times 'a \times 'a \times 'a)\ set$ **where**
  additive-quadruple-set $A \equiv \{(a, b, c, d) \mid a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$
  additive-quadruple $a\ b\ c\ d\}$

lemma *additive-quadruple-set-sub*:
  additive-quadruple-set $A \subseteq \{(a, b, c, d) \mid a\ b\ c\ d.\ d = a \oplus b \ominus c \wedge a \in A \wedge b \in A \wedge$
    $c \in A \wedge d \in A\}$ ⟨*proof*⟩

definition *additive-energy*:: $'a\ set \Rightarrow real$ **where**
  additive-energy $A \equiv$ *card* (*additive-quadruple-set* A) / (*card* A)^*3*

lemma *card-ineq-aux-quadruples*:
  assumes *finite* A
  shows *card* (*additive-quadruple-set* A) $\leq$ (*card* A)^*3*

⟨*proof*⟩

lemma *additive-energy-upper-bound*: *additive-energy* $A \leq 1$

⟨*proof*⟩

5.2 On sums

definition *f-sum*:: $'a \Rightarrow 'a\ set \Rightarrow nat$ **where**
$f\text{-}sum\ d\ A \equiv card\ \{(a, b) \mid a\ b.\ a \in A \wedge b \in A \wedge a \oplus b = d\}$

lemma *pairwise-disjnt-sum-1*:
$pairwise\ (\lambda s\ t.\ disjnt\ ((\lambda\ d\ .\{(a, b) \mid a\ b.\ a \in A \wedge b \in A \wedge (a \oplus b = d)\})\ s)$
$((\lambda\ d\ .\{(a, b) \mid a\ b.\ a \in A \wedge b \in A \wedge (a \oplus b = d)\})\ t))\ (sumset\ A\ A)$
⟨*proof*⟩

lemma *pairwise-disjnt-sum-2*:
$pairwise\ disjnt\ ((\lambda\ d.\ \{(a, b) \mid a\ b.\ a \in A \wedge b \in A \wedge a \oplus b = d\})\ `\ (sumset\ A\ A))$
⟨*proof*⟩

lemma *sum-Union-span*:
assumes $A \subseteq G$
shows $\bigcup\ ((\lambda\ d\ .\{(a, b) \mid a\ b.\ a \in A \wedge b \in A \wedge (a \oplus b = d)\})\ `\ (sumset\ A\ A)) = A \times A$

⟨*proof*⟩

lemma *f-sum-le-card*:
assumes *finite* A **and** $A \subseteq G$
shows $f\text{-}sum\ d\ A \leq card\ A$

⟨*proof*⟩

lemma *f-sum-card*:
assumes $A \subseteq G$ **and** *hA*: *finite* A
shows $(\sum\ d \in (sumset\ A\ A).\ (f\text{-}sum\ d\ A)) = (card\ A)\hat{}2$

⟨*proof*⟩

lemma *f-sum-card-eq*:
assumes $A \subseteq G$
shows $\forall\ x \in sumset\ A\ A.\ \ (f\text{-}sum\ x\ A)\hat{}2 =$
$card\ \{(a, b, c, d) \mid a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$
$additive\text{-}quadruple\ a\ b\ c\ d \wedge a \oplus b = x \wedge c \oplus d = x\}$

⟨*proof*⟩

lemma *pairwise-disjoint-sum*:
$pairwise\ (\lambda s\ t.\ disjnt\ ((\lambda\ x.\ \{(a, b, c, d) \mid a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$
$additive\text{-}quadruple\ a\ b\ c\ d \wedge a \oplus b = x \wedge c \oplus d = x\})\ s)$

$((\lambda\ x.\ \{(a,\ b,\ c,\ d)\ |\ a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$
additive-quadruple $a\ b\ c\ d \wedge a \oplus b = x \wedge c \oplus d = x\})\ t))$ (*sumset* $A\ A$)
⟨*proof*⟩

lemma *pairwise-disjnt-quadruple-sum*:
pairwise disjnt $((\lambda\ x.\ \{(a,\ b,\ c,\ d)\ |\ a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$
additive-quadruple $a\ b\ c\ d \wedge a \oplus b = x \wedge c \oplus d = x\})$ ‘ (*sumset* $A\ A$))
⟨*proof*⟩

lemma *quadruple-sum-Union-eq*:
$\bigcup\ ((\lambda\ x.\ \{(a,\ b,\ c,\ d)\ |\ a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$
additive-quadruple $a\ b\ c\ d \wedge a \oplus b = x \wedge c \oplus d = x\})$ ‘ (*sumset* $A\ A$)) =
additive-quadruple-set A

⟨*proof*⟩

lemma *f-sum-card-quadruple-set*:
assumes *hAG*: $A \subseteq G$ **and** *hA*: *finite* A
shows ($\sum\ d \in$ (*sumset* $A\ A$). (*f-sum* $d\ A$)^*2*) = *card* (*additive-quadruple-set* A)

⟨*proof*⟩

lemma *f-sum-card-quadruple-set-additive-energy*: **assumes** $A \subseteq G$ **and** *finite* A
shows ($\sum\ d \in$ *sumset* $A\ A$. (*f-sum* $d\ A$)^*2*) = *additive-energy* A ∗ (*card* A)^*3*
⟨*proof*⟩

definition *popular-sum*:: *'a* ⇒ *real* ⇒ *'a set* ⇒ *bool* **where**
popular-sum $d\ \vartheta\ A\ \equiv$ *f-sum* $d\ A \geq \vartheta$ ∗ *of-real* (*card* A)

definition *popular-sum-set*:: *real* ⇒ *'a set* ⇒ *'a set* **where**
popular-sum-set $\vartheta\ A \equiv \{d \in$ *sumset* $A\ A$. *popular-sum* $d\ \vartheta\ A\}$

5.3 On differences

The following material is directly analogous to the material given previously on sums. All definitions and lemmas are the corresponding ones for differences. E.g. *f-diff* corresponds to *f-sum*.

definition *f-diff*:: *'a* ⇒*'a set* ⇒ *nat* **where**
f-diff $d\ A \equiv$ *card* $\{(a,\ b)\ |\ a\ b.\ a \in A \wedge b \in A \wedge a \ominus b = d\}$

lemma *pairwise-disjnt-diff-1*:
pairwise ($\lambda s\ t.$ *disjnt* $((\lambda\ d\ .\{(a,\ b)\ |\ a\ b.\ a \in A \wedge b \in A \wedge (a \ominus b = d)\})\ s)$
$((\lambda\ d.\ \{(a,\ b)\ |\ a\ b.\ a \in A \wedge b \in A \wedge (a \ominus b = d)\})\ t))$ (*differenceset* $A\ A$)
⟨*proof*⟩

lemma *pairwise-disjnt-diff-2*:

pairwise disjnt $((\lambda\ d.\ \{(a,\ b) \mid a\ b.\ a \in A \wedge b \in A \wedge a \ominus b = d\})$ ‘ (*differenceset* $A\ A))$
⟨*proof*⟩

lemma *diff-Union-span*:
assumes $A \subseteq G$
shows $\bigcup\ ((\lambda\ d\ .\{(a,\ b) \mid a\ b.\ a \in A \wedge b \in A \wedge (a \ominus b = d)\})$ ‘ (*differenceset* $A\ A)) = A \times A$

⟨*proof*⟩

lemma *f-diff-le-card*:
assumes *finite* A **and** $A \subseteq G$
shows *f-diff* $d\ A \leq$ *card* A

⟨*proof*⟩

lemma *f-diff-card*:
assumes $A \subseteq G$ **and** *hA*: *finite* A
shows $(\sum\ d \in$ (*differenceset* $A\ A$). *f-diff* $d\ A) = (card\ A)\hat{}2$

⟨*proof*⟩

lemma *f-diff-card-eq*:
assumes $A \subseteq G$
shows $\forall\ x \in$ *differenceset* $A\ A$. (*f-diff* $x\ A)\hat{}2 =$
card $\{(a,\ b,\ c,\ d) \mid a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$
additive-quadruple $a\ b\ c\ d \wedge a \ominus c = x \wedge d \ominus b = x\}$

⟨*proof*⟩

lemma *pairwise-disjoint-diff*:
pairwise $(\lambda s\ t.$ *disjnt* $((\lambda\ x.\ \{(a,\ b,\ c,\ d) \mid a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$ *additive-quadruple* $a\ b\ c\ d \wedge a \ominus c = x \wedge d \ominus b = x\})\ s)$
$((\lambda\ x.\ \{(a,\ b,\ c,\ d) \mid a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$ *additive-quadruple* $a\ b\ c\ d \wedge a \ominus c = x \wedge d \ominus b = x\})\ t))$ (*differenceset* $A\ A$)
⟨*proof*⟩

lemma *pairwise-disjnt-quadruple-diff*:
pairwise disjnt $((\lambda\ x.\ \{(a,\ b,\ c,\ d) \mid a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$ *additive-quadruple* $a\ b\ c\ d \wedge a \ominus c = x \wedge d \ominus b = x\})$ ‘ (*differenceset* $A\ A$))
⟨*proof*⟩

lemma *quadruple-diff-Union-eq*:
$\bigcup\ ((\lambda\ x.\ \{(a,\ b,\ c,\ d) \mid a\ b\ c\ d.\ a \in A \wedge b \in A \wedge c \in A \wedge d \in A \wedge$ *additive-quadruple* $a\ b\ c\ d \wedge a \ominus c = x \wedge d \ominus b = x\})$ ‘ (*differenceset* $A\ A)) =$

additive-quadruple-set A

⟨*proof*⟩

lemma *f-diff-card-quadruple-set*:
assumes *hAG*: $A \subseteq G$ **and** *hA*: *finite A*
shows $(\sum\ d \in (\textit{differenceset}\ A\ A).\ (\textit{f-diff}\ d\ A)\hat{\ }2) = \textit{card}\ (\textit{additive-quadruple-set}\ A)$

⟨*proof*⟩

lemma *f-diff-card-quadruple-set-additive-energy*: **assumes** $A \subseteq G$ **and** *finite A*
shows $(\sum\ d \in \textit{differenceset}\ A\ A.\ (\textit{f-diff}\ d\ A)\hat{\ }2) = \textit{additive-energy}\ A * (\textit{card}\ A)\hat{\ }3$
⟨*proof*⟩

definition *popular-diff*:: $'a \Rightarrow \textit{real} \Rightarrow {'a}\ \textit{set} \Rightarrow \textit{bool}$ **where**
$\textit{popular-diff}\ d\ \vartheta\ A\ \equiv \textit{f-diff}\ d\ A \geq \vartheta * \textit{of-real}\ (\textit{card}\ A)$

definition *popular-diff-set*:: $\textit{real} \Rightarrow {'a}\ \textit{set} \Rightarrow {'a}\ \textit{set}$ **where**
$\textit{popular-diff-set}\ \vartheta\ A \equiv \{d \in \textit{differenceset}\ A\ A.\ \textit{popular-diff}\ d\ \vartheta\ A\}$

end
end

6 Results on lower bounds on additive energy

theory *Additive-Energy-Lower-Bounds*
imports
Additive-Combinatorics-Preliminaries
Miscellaneous-Lemmas
begin

context *additive-abelian-group*

begin

The following corresponds to Proposition 2.11 in Gowers's notes [3].

proposition *additive-energy-lower-bound-sumset*: **fixes** *C*::*real*
assumes *finite A* **and** $A \subseteq G$ **and** $(\textit{card}\ (\textit{sumset}\ A\ A)) \leq C * \textit{card}\ A$ **and** $\textit{card}\ A \neq 0$
shows $\textit{additive-energy}\ A \geq 1/C$

⟨*proof*⟩

An analogous version of Proposition 2.11 where the assumption is on a difference set is given below. The proof is identical to the proof of *additive-energy-lower-bound-sumset* above (with the obvious modifications).

proposition *additive-energy-lower-bound-differenceset*: **fixes** *C*::*real*

assumes *finite A* **and** $A \subseteq G$ **and** $(card\ (differenceset\ A\ A)) \leq C * card\ A$ **and** $card\ A \neq 0$
shows *additive-energy* $A \geq 1/C$

⟨*proof*⟩

end
end

7 Towards the proof of the Balog–Szemerédi–Gowers Theorem

theory *Balog-Szemeredi-Gowers-Main-Proof*
imports
Prob-Space-Lemmas
Graph-Theory-Preliminaries
Sumset-Triangle-Inequality
Additive-Combinatorics-Preliminaries
begin

context *additive-abelian-group*

begin

After having introduced all the necessary preliminaries in the imported files, we are now ready to follow the chain of the arguments for the main proof as in Gowers's notes [3].

The following lemma corresponds to Lemma 2.13 in Gowers's notes [3].

lemma (**in** *fin-bipartite-graph*) *proportion-bad-pairs-subset-bipartite*:
fixes c::*real*
assumes $c > 0$
obtains X' **where** $X' \subseteq X$ **and** $card\ X' \geq density * card\ X\ /\ sqrt\ 2$ **and** $card\ (bad\text{-}pair\text{-}set\ X'\ Y\ c)\ /\ (card\ X')\hat{}2 \leq 2 * c\ /\ density\hat{}2$
⟨*proof*⟩

The following technical probability lemma corresponds to Lemma 2.14 in Gowers's notes [3].

lemma (**in** *prob-space*) *expectation-condition-card-1*:
fixes X::*'a set* **and** f::*'a* ⇒ *real* **and** δ::*real*
assumes *finite X* **and** $\forall\ x \in X.\ f\ x \leq 1$ **and** $M = uniform\text{-}count\text{-}measure\ X$ **and** *expectation* $f \geq \delta$
shows $card\ \{x \in X.\ (f\ x \geq \delta\ /\ 2)\} \geq \delta * card\ X\ /\ 2$
⟨*proof*⟩

The following technical probability lemma corresponds to Lemma 2.15 in Gowers's notes.

lemma (**in** *prob-space*) *expectation-condition-card-2*:

fixes X::$'a$ *set* **and** β::*real* **and** α::*real* **and** f:: $'a \Rightarrow$ *real*
assumes *finite* X **and** $\bigwedge x.\ x \in X \Longrightarrow f\, x \leq 1$ **and** $\beta > 0$ **and** $\alpha > 0$
and *expectation* $f \geq 1 - \alpha$ **and** $M =$ *uniform-count-measure* X
shows *card* $\{x \in X.\ f\, x \geq 1 - \beta\} \geq (1 - \alpha\ /\ \beta) *$ *card* X

⟨*proof*⟩

The following lemma corresponds to Lemma 2.16 in Gowers's notes [3]. For the proof, we will apply Lemma 2.13 (*proportion-bad-pairs-subset-bipartite*, the technical probability Lemmas 2.14 (*expectation-condition-card-1*) and 2.15 (*expectation-condition-card-2*) as well as background material on graphs with loops and bipartite graphs that was previously presented.

lemma (**in** *fin-bipartite-graph*) *walks-of-length-3-subsets-bipartite*:
obtains X' **and** Y' **where** $X' \subseteq X$ **and** $Y' \subseteq Y$ **and**
card $X' \geq$ (*edge-density* X Y)$\hat{}2 *$ *card* X / 16 **and**
card $Y' \geq$ *edge-density* X Y $*$ *card* Y / 4 **and**
$\forall\ x \in X'.\ \forall\ y \in Y'.$ *card* $\{p.$ *connecting-walk* $x\ y\ p \wedge$ *walk-length* $p = 3\} \geq$
(*edge-density* X Y)$\hat{}6 *$ *card* $X *$ *card* Y / $2\hat{}13$

⟨*proof*⟩

The following lemma corresponds to Lemma 2.17 in Gowers's notes [3].

Note that here we have set(*additive-energy* $A = 2 * c$ (instead of (*additive-energy* $A = c$ as in the notes) and we are accordingly considering c-popular differences (instead of c/2-popular differences as in the notes) so that we will still have ($\vartheta =$ *additive-energy* A / 2.

lemma *popular-differences-card*: **fixes** A::$'a$ *set* **and** c::*real*
assumes *finite* A **and** $A \subseteq G$ **and** *additive-energy* $A = 2 * c$
shows *card* (*popular-diff-set* c A) $\geq c *$ *card* A

⟨*proof*⟩

The following lemma corresponds to Lemma 2.18 in Gowers's notes [3]. It includes the key argument of the main proof and its proof applies Lemmas 2.16 (*walks-of-length-3-subsets-bipartite*) and 2.17 (*popular-differences-card*). In the proof we will use an appropriately defined bipartite graph as an intermediate/auxiliary construct so as to apply lemma *walks-of-length-3-subsets-bipartite*. As each vertex set of the bipartite graph is constructed to be a copy of a finite subset of an Abelian group, we need flexibility regarding types, which is what prompted the introduction and use of the new graph theory library [1] (that does not impose any type restrictions e.g. by representing vertices as natural numbers).

lemma *obtains-subsets-differenceset-card-bound*:
fixes A::$'a$ *set* **and** c::*real*
assumes *finite* A **and** $c>0$ **and** $A \neq \{\}$ **and** $A \subseteq G$ **and** *additive-energy* $A =$
$2 * c$

obtains B **and** A' **where** $B \subseteq A$ **and** $B \neq \{\}$ **and** *card* $B \geq c\hat{}4 * card\ A\ /\ 16$
and $A' \subseteq A$ **and** $A' \neq \{\}$ **and** *card* $A' \geq c\hat{}2 * card\ A\ /\ 4$
and *card* (*differenceset* A' B) $\leq 2\hat{}13 * card\ A\ /\ c\hat{}15$

⟨*proof*⟩

We now show the main theorem, which is a direct application of lemma *obtains-subsets-differenceset-card-bound* and the Ruzsa triangle inequality. (The main theorem corresponds to Corollary 2.19 in Gowers's notes [3].)

theorem *Balog-Szemeredi-Gowers*: **fixes** A::$'a$ *set* **and** c::*real*
assumes *afin*: *finite* A **and** $A \neq \{\}$ **and** $c>0$ **and** *additive-energy* $A = 2 * c$
and *ass*: $A \subseteq G$
obtains A' **where** $A' \subseteq A$ **and** *card* $A' \geq c\hat{}2 * card\ A\ /\ 4$ **and**
card (*differenceset* A' A') $\leq 2\hat{}30 * card\ A\ /\ c\hat{}34$
⟨*proof*⟩

The following is an analogous version of the Balog–Szemerédi–Gowers Theorem for a sumset instead of a difference set. The proof is similar to that of the original version, again using *obtains-subsets-differenceset-card-bound*, however, instead of the Ruzsa triangle inequality we will use the alternative triangle inequality for sumsets *triangle-ineq-sumsets*.

theorem *Balog-Szemeredi-Gowers-sumset*: **fixes** A::$'a$ *set* **and** c::*real*
assumes *afin*: *finite* A **and** $A \neq \{\}$ **and** $c>0$ **and** *additive-energy* $A = 2 * c$
and *ass*: $A \subseteq G$
obtains A' **where** $A' \subseteq A$ **and** *card* $A' \geq c\hat{}2 * card\ A\ /\ 4$ **and**
card (*sumset* A' A') $\leq 2\hat{}30 * card\ A\ /\ c\hat{}34$

⟨*proof*⟩

end
end

8 Supplementary results related to intermediate lemmas used in the proof of the Balog–Szemerédi–Gowers Theorem

theory *Balog-Szemeredi-Gowers-Supplementary*
imports
Balog-Szemeredi-Gowers-Main-Proof
begin

context *additive-abelian-group*

begin

Even though it is not applied anywhere in this development, for the sake of completeness we give the following analogous version of Lemma 2.17 (*pop-*

ular-differences-card) but for popular sums instead of popular differences. The proof is identical to that of Lemma 2.17, with the obvious modifications.

lemma *popular-sums-card*:
fixes A::$'a$ *set* **and** c::*real*
assumes *finite* A **and** *additive-energy* $A = 2 * c$ **and** $A \subseteq G$
shows *card* (*popular-sum-set* c A) $\geq c *$ *card* A

⟨*proof*⟩

The following is an analogous version of lemma *obtains-subsets-differenceset-card-bound* (2.18 in Gowers's notes [3]) but for a sumset instead of a difference set. It is not used anywhere in this development but we provide it for the sake of completeness. The proof is identical to that of lemma *obtains-subsets-differenceset-card-bound* with *f-diff* changed to *f-sum*, *popular-diff* changed to *popular-sum*, ⊕interchanged with ⊖, and instead of lemma *popular-differences-card* we apply its analogous version for popular sums, that is lemma *popular-sums-card*.

lemma *obtains-subsets-sumset-card-bound*: **fixes** A::$'a$ *set* **and** c::*real*
assumes *finite* A **and** $c>0$ **and** $A \neq \{\}$ **and** $A \subseteq G$ **and** *additive-energy* $A = 2 * c$
obtains B **and** A' **where** $B \subseteq A$ **and** $B \neq \{\}$ **and** *card* $B \geq c\hat{}4 *$ *card* A / 16
and $A' \subseteq A$ **and** $A' \neq \{\}$ **and** *card* $A' \geq c\hat{}2 *$ *card* A / 4
and *card* (*sumset* A' B) $\leq 2\hat{}13 *$ *card* A / $c\hat{}15$

⟨*proof*⟩

end
end

References

[1] C. Edmonds. Undirected graph theory. *Archive of Formal Proofs*, September 2022. https://isa-afp.org/entries/Undirected_Graph_Theory.html, Formal proof development.

[2] W. T. Gowers. A new proof of Szemerédi's theorem. *Geometric & Functional Analysis GAFA*, 11(3):465–588, 2001.

[3] W. T. Gowers. Introduction to additive combinatorics, 2022. Lecture notes for Part III of the Mathematical Tripos taught at the University of Cambridge, available at https://drive.google.com/file/d/1ut0mUqSyPMweoxoDTfhXverEONyFgcuO/view.

[4] Y. Zhao. Graph theory and additive combinatorics. Online at https://yufeizhao.com/gtacbook/, 2022. book draft.