Roth's Theorem on Arithmetic Progressions

Chelsea Edmonds, Angeliki Koutsoukou-Argyraki and Lawrence C. Paulson
Computer Laboratory, University of Cambridge CB3 0FD
{cle47,ak2110,lp15}@cam.ac.uk

May 14, 2024

Abstract

We formalise a proof of Roth's Theorem on Arithmetic Progressions, a major result in additive combinatorics on the existence of 3-term arithmetic progressions in subsets of natural numbers. To this end, we follow a proof using graph regularity. We employ our recent formalisation of Szemerédi's Regularity Lemma, a major result in extremal graph theory, which we use here to prove the Triangle Counting Lemma and the Triangle Removal Lemma. Our sources are Yufei Zhao's MIT lecture notes "Graph Theory and Additive Combinatorics"[1] and W.T. Gowers's Cambridge lecture notes "Topics in Combinatorics".[2] We also refer to the University of Georgia notes by Stephanie Bell and Will Grodzicki "Using Szemerédi's Regularity Lemma to Prove Roth's Theorem".[3]

Contents

Acknowledgements

The authors were supported by the ERC Advanced Grant ALEXANDRIA (Project 742178) funded by the European Research Council.

[1] https://yufeizhao.com/gtacbook/ and https://yufeizhao.com/gtac/gtac.pdf
[2] https://www.dpmms.cam.ac.uk/~par31/notes/tic.pdf
[3] http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.432.327

1 Roth's Theorem on Arithmetic Progressions

theory *Roth-Arithmetic-Progressions*
imports *Szemeredi-Regularity.Szemeredi*
Random-Graph-Subgraph-Threshold.Subgraph-Threshold
Ergodic-Theory.Asymptotic-Density
HOL−Library.Ramsey HOL−Library.Nat-Bijection

begin

1.1 Miscellaneous Preliminaries

lemma *sum-prod-le-prod-sum*:
fixes $a :: 'a \Rightarrow 'b$::*linordered-idom*
assumes $\bigwedge i.\ i \in I \Longrightarrow a\ i \geq 0 \wedge b\ i \geq 0$
shows $(\sum i{\in}I.\ \sum j{\in}I.\ a\ i * b\ j) \leq (\sum i{\in}I.\ a\ i) * (\sum i{\in}I.\ b\ i)$
⟨*proof*⟩

lemma *real-mult-gt-cube*: $A \geq (X :: real) \Longrightarrow B \geq X \Longrightarrow C \geq X \Longrightarrow X \geq 0 \Longrightarrow A * B * C \geq X\hat{}3$
⟨*proof*⟩

lemma *triple-sigma-rewrite-card*:
assumes *finite X finite Y finite Z*
shows *card* $\{(x,y,z)\ .\ x \in X \wedge (y,z) \in Y \times Z \wedge P\ x\ y\ z\} = (\sum x{\in}\ X\ .\ card\ \{(y,z) \in Y \times Z.\ P\ x\ y\ z\})$
⟨*proof*⟩

lemma *all-edges-between-mono1*:
$Y \subseteq Z \Longrightarrow$ *all-edges-between Y X G* $\subseteq$ *all-edges-between Z X G*
⟨*proof*⟩

lemma *all-edges-between-mono2*:
$Y \subseteq Z \Longrightarrow$ *all-edges-between X Y G* $\subseteq$ *all-edges-between X Z G*
⟨*proof*⟩

lemma *uwellformed-alt-fst*:
assumes *uwellformed G* $\{x, y\} \in$ *uedges G*
shows $x \in$ *uverts G*
⟨*proof*⟩

lemma *uwellformed-alt-snd*:
assumes *uwellformed G* $\{x, y\} \in$ *uedges G*
shows $y \in$ *uverts G*
⟨*proof*⟩

lemma *all-edges-between-subset-times*: *all-edges-between X Y G* $\subseteq (X \cap \bigcup(uedges\ G)) \times (Y \cap \bigcup(uedges\ G))$
⟨*proof*⟩

lemma *finite-all-edges-between'*:
assumes *finite* (*uverts G*) *uwellformed G*
shows *finite* (*all-edges-between X Y G*)
⟨*proof*⟩

lemma *all-edges-between-E-diff*:
all-edges-between X Y (*V*,*E*−*E'*) = *all-edges-between X Y* (*V*,*E*) − *all-edges-between X Y* (*V*,*E'*)
⟨*proof*⟩

lemma *all-edges-between-E-Un*:
all-edges-between X Y (*V*,*E*∪*E'*) = *all-edges-between X Y* (*V*,*E*) ∪ *all-edges-between X Y* (*V*,*E'*)
⟨*proof*⟩

lemma *all-edges-between-E-UN*:
all-edges-between X Y (*V*, ⋃*i*∈*I*. *E i*) = (⋃*i*∈*I*. *all-edges-between X Y* (*V*,*E i*))
⟨*proof*⟩

lemma *all-edges-preserved*: ⟦*all-edges-between A B G'* = *all-edges-between A B G*; *X* ⊆ *A*; *Y* ⊆ *B*⟧
⟹ *all-edges-between X Y G'* = *all-edges-between X Y G*
⟨*proof*⟩

lemma *subgraph-edge-wf*:
assumes *uwellformed G uverts H* = *uverts G uedges H* ⊆ *uedges G*
shows *uwellformed H*
⟨*proof*⟩

1.2 Preliminaries on Neighbors in Graphs

definition *neighbor-in-graph*:: *uvert* ⇒ *uvert* ⇒ *ugraph* ⇒ *bool*
where *neighbor-in-graph x y G* ≡ (*x* ∈ (*uverts G*) ∧ *y* ∈ (*uverts G*) ∧ {*x*,*y*} ∈ (*uedges G*))

definition *neighbors* :: *uvert* ⇒ *ugraph* ⇒ *uvert set* **where**
neighbors x G ≡ {*y* ∈ *uverts G* . *neighbor-in-graph x y G*}

definition *neighbors-ss*:: *uvert* ⇒ *uvert set* ⇒ *ugraph* ⇒ *uvert set* **where**
neighbors-ss x Y G ≡ {*y* ∈ *Y* . *neighbor-in-graph x y G*}

lemma *all-edges-betw-sigma-neighbor*:
uwellformed G ⟹ *all-edges-between X Y G* = (*SIGMA x*:*X*. *neighbors-ss x Y G*)
⟨*proof*⟩

lemma *card-all-edges-betw-neighbor*:
assumes *finite X finite Y uwellformed G*
shows *card* (*all-edges-between X Y G*) = (∑*x*∈*X*. *card* (*neighbors-ss x Y G*))

⟨*proof*⟩

1.3 Preliminaries on Triangles in Graphs

definition *triangle-in-graph*:: *uvert* ⇒ *uvert* ⇒ *uvert* ⇒ *ugraph* ⇒ *bool*
where *triangle-in-graph x y z G*
≡ ({*x*,*y*} ∈ *uedges G*) ∧ ({*y*,*z*} ∈ *uedges G*) ∧ ({*x*,*z*} ∈ *uedges G*)

definition *triangle-triples*
where *triangle-triples X Y Z G* ≡ {(*x*,*y*,*z*) ∈ *X* × *Y* × *Z*. *triangle-in-graph x y z G*}

lemma *triangle-commu1*:
assumes *triangle-in-graph x y z G*
shows *triangle-in-graph y x z G*
⟨*proof*⟩

lemma *triangle-vertices-distinct1*:
assumes *wf*: *uwellformed G*
assumes *tri*: *triangle-in-graph x y z G*
shows *x* ≠ *y*
⟨*proof*⟩

lemma *triangle-vertices-distinct2*:
assumes *uwellformed G triangle-in-graph x y z G*
shows *y* ≠ *z*
⟨*proof*⟩

lemma *triangle-in-graph-edge-point*:
assumes *uwellformed G*
shows *triangle-in-graph x y z G* ⟷ {*y*, *z*} ∈ *uedges G* ∧ *neighbor-in-graph x y G* ∧ *neighbor-in-graph x z G*
⟨*proof*⟩

definition
unique-triangles G
≡ ∀ *e* ∈ *uedges G*. ∃!*T*. ∃ *x y z*. *T* = {*x*,*y*,*z*} ∧ *triangle-in-graph x y z G* ∧ *e* ⊆ *T*

definition *triangle-free-graph*:: *ugraph* ⇒ *bool*
where *triangle-free-graph G* ≡ ¬(∃ *x y z*. *triangle-in-graph x y z G*)

lemma *triangle-free-graph-empty*: *uedges G* = {} ⟹ *triangle-free-graph G*
⟨*proof*⟩

lemma *edge-vertices-not-equal*:
assumes *uwellformed G* {*x*,*y*} ∈ *uedges G*
shows *x* ≠ *y*
⟨*proof*⟩

lemma *triangle-in-graph-verts*:
assumes *uwellformed G triangle-in-graph x y z G*
shows *x* ∈ *uverts G y* ∈ *uverts G z*∈ *uverts G*
⟨*proof*⟩

definition *triangle-set* :: *ugraph* ⇒ *uvert set set*
where *triangle-set G* ≡ { {*x*,*y*,*z*} | *x y z*. *triangle-in-graph x y z G*}

fun *mk-triangle-set* :: (*uvert* × *uvert* × *uvert*) ⇒ *uvert set*
where *mk-triangle-set* (*x*,*y*,*z*) = {*x*,*y*,*z*}

lemma *finite-triangle-set*:
assumes *fin*: *finite* (*uverts G*) **and** *wf*: *uwellformed G*
shows *finite* (*triangle-set G*)
⟨*proof*⟩

lemma *card-triangle-3*:
assumes *t* ∈ *triangle-set G uwellformed G*
shows *card t* = *3*
⟨*proof*⟩

lemma *triangle-set-power-set-ss*: *uwellformed G* ⟹ *triangle-set G* ⊆ *Pow* (*uverts G*)
⟨*proof*⟩

lemma *triangle-in-graph-ss*:
assumes *uedges Gnew* ⊆ *uedges G*
assumes *triangle-in-graph x y z Gnew*
shows *triangle-in-graph x y z G*
⟨*proof*⟩

lemma *triangle-set-graph-edge-ss*:
assumes *uedges Gnew* ⊆ *uedges G*
assumes *uverts Gnew* = *uverts G*
shows *triangle-set Gnew* ⊆ *triangle-set G*
⟨*proof*⟩

lemma *triangle-set-graph-edge-ss-bound*:
fixes *G* :: *ugraph* **and** *Gnew* :: *ugraph*
assumes *uwellformed G finite* (*uverts G*) *uedges Gnew* ⊆ *uedges G uverts Gnew* = *uverts G*
shows *card* (*triangle-set G*) ≥ *card* (*triangle-set Gnew*)
⟨*proof*⟩

1.4 The Triangle Counting Lemma and the Triangle Removal Lemma

We begin with some more auxiliary material to be used in the main lemmas.

lemma *regular-pair-neighbor-bound*:
fixes ε::*real*
assumes *finG*: *finite* (*uverts G*)
assumes *xss*: $X \subseteq$ *uverts G* **and** *yss*: $Y \subseteq$ *uverts G* **and** *card X* $>$ *0*
and *wf*: *uwellformed G*
and *eg0*: $\varepsilon > 0$ **and** ε−*regular-pair X Y G* **and** *ed*: *edge-density X Y G* $\geq 2*\varepsilon$
defines $X' \equiv \{x \in X.$ *card* (*neighbors-ss x Y G*) $<$ (*edge-density X Y G* $-$ ε) $*$ *card* (*Y*)$\}$
shows *card X'* $< \varepsilon *$ *card X*
(**is** *card* (*?X'*) $< \varepsilon *$ -)
⟨*proof*⟩

lemma *neighbor-set-meets-e-reg-cond*:
fixes ε::*real*
assumes *edge-density X Y G* $\geq 2*\varepsilon$
and *card* (*neighbors-ss x Y G*) $\geq$ (*edge-density X Y G* $-$ ε) $*$ *card Y*
shows *card* (*neighbors-ss x Y G*) $\geq \varepsilon *$ *card* (*Y*)
⟨*proof*⟩

lemma *all-edges-btwn-neighbor-sets-lower-bound*:
fixes ε::*real*
assumes *rp2*: ε−*regular-pair Y Z G*
and *ed1*: *edge-density X Y G* $\geq 2*\varepsilon$ **and** *ed2*: *edge-density X Z G* $\geq 2*\varepsilon$
and *cond1*: *card* (*neighbors-ss x Y G*) $\geq$ (*edge-density X Y G* $-$ ε) $*$ *card Y*
and *cond2*: *card* (*neighbors-ss x Z G*) $\geq$ (*edge-density X Z G* $-$ ε) $*$ *card Z*
shows *card* (*all-edges-between* (*neighbors-ss x Y G*) (*neighbors-ss x Z G*) *G*)
$\geq$ (*edge-density Y Z G* $-$ ε) $*$ *card* (*neighbors-ss x Y G*) $*$ *card* (*neighbors-ss x Z G*)
(**is** *card* (*all-edges-between ?Y' ?Z' G*) $\geq$ (*edge-density Y Z G* $-$ ε) $*$ *card ?Y'* $*$ *card ?Z'*)
⟨*proof*⟩

We are now ready to show the Triangle Counting Lemma:

theorem *triangle-counting-lemma*:
fixes ε::*real*
assumes *xss*: $X \subseteq$ *uverts G* **and** *yss*: $Y \subseteq$ *uverts G* **and** *zss*: $Z \subseteq$ *uverts G* **and** *en0*: $\varepsilon > 0$
and *finG*: *finite* (*uverts G*) **and** *wf*: *uwellformed G*
and *rp1*: ε−*regular-pair X Y G* **and** *rp2*: ε−*regular-pair Y Z G* **and** *rp3*: ε−*regular-pair X Z G*
and *ed1*: *edge-density X Y G* $\geq 2*\varepsilon$ **and** *ed2*: *edge-density X Z G* $\geq 2*\varepsilon$ **and** *ed3*: *edge-density Y Z G* $\geq 2*\varepsilon$
shows *card* (*triangle-triples X Y Z G*)
$\geq$ (*1*$-$*2*$*$\varepsilon$) $*$ (*edge-density X Y G* $-$ ε) $*$ (*edge-density X Z G* $-$ ε) $*$ (*edge-density Y Z G* $-$ ε)$*$

card X ∗ card Y ∗ card Z
⟨*proof*⟩

definition *regular-graph* :: *real* ⇒ *uvert set set* ⇒ *ugraph* ⇒ *bool*
(*-−regular′-graph* [*999*]*1000*)
where *ε−regular-graph P G* ≡ ∀ *R S*. *R*∈*P* ⟶ *S*∈*P* ⟶ *ε−regular-pair R S G*
for *ε*::*real*

A minimum density, but empty edge sets are excluded.

definition *edge-dense* :: *nat set* ⇒ *nat set* ⇒ *ugraph* ⇒ *real* ⇒ *bool*
where *edge-dense X Y G ε* ≡ *all-edges-between X Y G* = {} ∨ *edge-density X Y G* ≥ *ε*

definition *dense-graph* :: *uvert set set* ⇒ *ugraph* ⇒ *real* ⇒ *bool*
where *dense-graph P G ε* ≡ ∀ *R S*. *R*∈*P* ⟶ *S*∈*P* ⟶ *edge-dense R S G ε* **for** *ε*::*real*

definition *decent* :: *nat set* ⇒ *nat set* ⇒ *ugraph* ⇒ *real* ⇒ *bool*
where *decent X Y G η* ≡ *all-edges-between X Y G* = {} ∨ (*card X* ≥ *η* ∧ *card Y* ≥ *η*) **for** *η*::*real*

definition *decent-graph* :: *uvert set set* ⇒ *ugraph* ⇒ *real* ⇒ *bool*
where *decent-graph P G η* ≡ ∀ *R S*. *R*∈*P* ⟶ *S*∈*P* ⟶ *decent R S G η*

The proof of the triangle counting lemma requires ordered triples. For each unordered triple there are six permutations, hence the factor of 1/6 here.

lemma *card-convert-triangle-rep*:
fixes *G* :: *ugraph*
assumes *X* ⊆ *uverts G* **and** *Y* ⊆ *uverts G* **and** *Z* ⊆ *uverts G* **and** *fin*: *finite* (*uverts G*)
and *wf*: *uwellformed G*
shows *card* (*triangle-set G*) ≥ *1/6* ∗ *card* {(*x*,*y*,*z*) ∈ *X* × *Y* × *Z* . (*triangle-in-graph x y z G*)}
(**is** *- ≥ 1/6 ∗ card ?TT*)
⟨*proof*⟩

lemma *card-convert-triangle-rep-bound*:
fixes *G* :: *ugraph* **and** *t* :: *real*
assumes *X* ⊆ *uverts G* **and** *Y* ⊆ *uverts G* **and** *Z* ⊆ *uverts G* **and** *fin*: *finite* (*uverts G*)
and *wf*: *uwellformed G*
assumes *card* {(*x*,*y*,*z*) ∈ *X* × *Y* × *Z* . (*triangle-in-graph x y z G*)} ≥ *t*
shows *card* (*triangle-set G*) ≥ *1/6* ∗*t*
⟨*proof*⟩

lemma *edge-density-eq0*:
assumes *all-edges-between A B G* = {} **and** *X* ⊆ *A Y* ⊆ *B*

shows *edge-density X Y G = 0*
⟨*proof*⟩

The following is the Triangle Removal Lemma.

theorem *triangle-removal-lemma*:
fixes ε :: *real*
assumes *egt*: $\varepsilon > 0$
shows $\exists \delta$::*real* > 0. $\forall G$. *card*(*uverts G*) $> 0 \longrightarrow$ *uwellformed G* $\longrightarrow$
card (*triangle-set G*) $\leq \delta *$ *card*(*uverts G*) ^ 3 $\longrightarrow$
($\exists G'$. *triangle-free-graph* $G' \wedge$ *uverts* G' = *uverts G* $\wedge$ *uedges* $G' \subseteq$ *uedges G* $\wedge$
card (*uedges G* − *uedges* G') $\leq \varepsilon *$ (*card* (*uverts G*))^2)
(**is** $\exists \delta$::*real* > 0. $\forall G$. - $\longrightarrow$ - $\longrightarrow$ - $\longrightarrow$ ($\exists$ *Gnew*. *?Φ G Gnew*))
⟨*proof*⟩

1.5 Roth's Theorem

We will first need the following corollary of the Triangle Removal Lemma.

See https://en.wikipedia.org/wiki/Ruzsa--Szemerédi_problem. Suggested by Yaël Dillies

corollary *Diamond-free*:
fixes ε :: *real*
assumes $0 < \varepsilon$
shows $\exists N{>}0$. $\forall G$. *card*(*uverts G*) $> N \longrightarrow$ *uwellformed G* $\longrightarrow$ *unique-triangles G* $\longrightarrow$
card (*uedges G*) $\leq \varepsilon *$ (*card* (*uverts G*))^2
⟨*proof*⟩

We are now ready to proceed to the proof of Roth's Theorem for Arithmetic Progressions.

definition *progression3* :: *'a*::*comm-monoid-add* ⇒ *'a* ⇒ *'a set*
where *progression3 k d* ≡ {*k*, *k*+*d*, *k*+*d*+*d*}

lemma *p3-int-iff*: *progression3* (*int k*) (*int d*) ⊆ *int* ' *A* ⟷ *progression3 k d* ⊆ *A*
⟨*proof*⟩

We assume that a set of naturals $A \subseteq \{... < N\}$ does not have any arithmetic progression. We will then show that A is of cardinality $o(N)$.

lemma *RothArithmeticProgressions-aux*:
fixes ε::*real*
assumes $\varepsilon > 0$
obtains *M* **where** $\forall N \geq M$. $\forall A \subseteq \{..{<}N\}$. ($\nexists k\ d$. $d{>}0 \wedge$ *progression3 k d* ⊆ *A*) $\longrightarrow$ *card A* $< \varepsilon *$ *real N*
⟨*proof*⟩

We finally present the main statement formulated using the upper asymptotic density condition.

theorem *RothArithmeticProgressions*:
  assumes *upper-asymptotic-density A > 0*
  shows $\exists k\ d.\ d{>}0 \wedge$ *progression3 k d* $\subseteq A$
⟨*proof*⟩

end