Density Theorems for words

Konstantinos Tyros

University of Toronto
Department of Mathematics

2014

Van der Waerden's Theorem 1927

Theorem

For every finite coloring of the natural numbers, one of the colors contains arbitrarily long arithmetic progressions.

Theorem

Let k, r be positive integers. Then there exists some positive integer n_0 such that for every $n \geqslant n_0$ and every r-coloring of the set $\{0, ..., n-1\}$, there exists a monochromatic arithmetic progression of length k.

Van der Waerden's Theorem 1927

Theorem

For every finite coloring of the natural numbers, one of the colors contains arbitrarily long arithmetic progressions.

Theorem

Let k, r be positive integers. Then there exists some positive integer n_0 such that for every $n \geqslant n_0$ and every r-coloring of the set $\{0, \ldots, n-1\}$, there exists a monochromatic arithmetic progression of length k.

Szeméredi's Theorems, 1975

Theorem

Let A be a subset of ω such that

$$\limsup_{n\to\infty} \frac{|A \cap \{0, ..., n-1\}|}{n} > 0.$$

Then A contains arbitrarily long arithmetic progressions.

Theorem

Let k be a positive integer and δ be a positive real. Then there exists some positive integer n_0 such that for every $n \geqslant n_0$ we have that every subset A of the set $\{0, ..., n-1\}$ with $\frac{|A|}{n} \geqslant \delta$ contains an arithmetic progression of length k.

Szeméredi's Theorems, 1975

Theorem

Let A be a subset of ω such that

$$\limsup_{n\to\infty} \frac{|A \cap \{0, ..., n-1\}|}{n} > 0.$$

Then A contains arbitrarily long arithmetic progressions.

Theorem

Let k be a positive integer and δ be a positive real. Then there exists some positive integer n_0 such that for every $n \geqslant n_0$ we have that every subset A of the set $\{0, ..., n-1\}$ with $\frac{|A|}{n} \geqslant \delta$ contains an arithmetic progression of length k.

Szeméredi's Theorems, 1975

- The above result was first conjectured by Erdős and Turán in 1936.
- In 1953 Roth proved it for $k = 3$.
- In 1969 Szemerédi proved it for $k = 4$.
- An proof using Ergodic Theory is given by Furstenberg (1977).
- The best known upper bound for the least such n_0 is due to Gowers (2001).
- Gowers and independently Nagel, Rödl, Schacht and Skokan proved the hypergraph removal lemma, providing a new proof for Szemerédi's Theorem as well as its multidimensional version.

Szeméredi's Theorems, 1975

- The above result was first conjectured by Erdős and Turán in 1936.
- In 1953 Roth proved it for $k = 3$.
- In 1969 Szemerédi proved it for $k = 4$.
- An proof using Ergodic Theory is given by Furstenberg (1977).
- The best known upper bound for the least such n_0 is due to Gowers (2001).
- Gowers and independently Nagel, Rödl, Schacht and Skokan proved the hypergraph removal lemma, providing a new proof for Szemerédi's Theorem as well as its multidimensional version.

Szeméredi's Theorems, 1975

- The above result was first conjectured by Erdős and Turán in 1936.
- In 1953 Roth proved it for $k = 3$.
- In 1969 Szemerédi proved it for $k = 4$.
- An proof using Ergodic Theory is given by Furstenberg (1977).
- The best known upper bound for the least such n_0 is due to Gowers (2001).
- Gowers and independently Nagel, Rödl, Schacht and Skokan proved the hypergraph removal lemma, providing a new proof for Szemerédi's Theorem as well as its multidimensional version.

Szeméredi's Theorems, 1975

- The above result was first conjectured by Erdős and Turán in 1936.
- In 1953 Roth proved it for $k = 3$.
- In 1969 Szemerédi proved it for $k = 4$.
- An proof using Ergodic Theory is given by Furstenberg (1977).
- The best known upper bound for the least such n_0 is due to Gowers (2001).
- Gowers and independently Nagel, Rödl, Schacht and Skokan proved the hypergraph removal lemma, providing a new proof for Szemerédi's Theorem as well as its multidimensional version.

Szeméredi's Theorems, 1975

- The above result was first conjectured by Erdős and Turán in 1936.
- In 1953 Roth proved it for $k = 3$.
- In 1969 Szemerédi proved it for $k = 4$.
- An proof using Ergodic Theory is given by Furstenberg (1977).
- The best known upper bound for the least such n_0 is due to Gowers (2001).
- Gowers and independently Nagel, Rödl, Schacht and Skokan proved the hypergraph removal lemma, providing a new proof for Szemerédi's Theorem as well as its multidimensional version.

Szeméredi's Theorems, 1975

- The above result was first conjectured by Erdős and Turán in 1936.
- In 1953 Roth proved it for $k = 3$.
- In 1969 Szemerédi proved it for $k = 4$.
- An proof using Ergodic Theory is given by Furstenberg (1977).
- The best known upper bound for the least such n_0 is due to Gowers (2001).
- Gowers and independently Nagel, Rödl, Schacht and Skokan proved the hypergraph removal lemma, providing a new proof for Szemerédi's Theorem as well as its multidimensional version.

To state the Hales–Jewett Theorem we need to introduce some notation. Let k be a positive integer and n be a non-negative integer. Moreover, let v be a symbol not belonging to $[k] = \{1, ..., k\}$.

- By $[k]^n$ we denote the set of all sequences $(a_0, ..., a_{n-1})$ of length n with elements from $[k]$.
- We will refer to the elements of $[k]^n$ as constant words of length n over k.

The Hales–Jewett Theorem

To state the Hales–Jewett Theorem we need to introduce some notation. Let k be a positive integer and n be a non-negative integer. Moreover, let v be a symbol not belonging to $[k] = \{1, ..., k\}$.

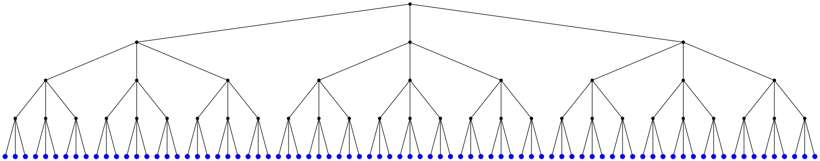
- By $[k]^n$ we denote the set of all sequences $(a_0, ..., a_{n-1})$ of length n with elements from $[k]$.
- We will refer to the elements of $[k]^n$ as constant words of length n over k.

The Hales–Jewett Theorem

To state the Hales–Jewett Theorem we need to introduce some notation. Let k be a positive integer and n be a non-negative integer. Moreover, let v be a symbol not belonging to $[k] = \{1, ..., k\}$.

- By $[k]^n$ we denote the set of all sequences $(a_0, ..., a_{n-1})$ of length n with elements from $[k]$.
- We will refer to the elements of $[k]^n$ as constant words of length n over k.

For example assume that $k = 3$ and $n = 4$.



Moreover, let v be a symbol not belonging to $[k] = \{1, ..., k\}$.

- A variable word over k is a finite sequence $w(v)$ in $[k] \cup \{v\}$ such that v occurs at least once.
- For a variable word $w(v)$ and $a \in [k]$ by $w(a)$ we denote the constant word over k resulting by substituting every occurrence of v by a.
- A combinatorial line is a set of the form $\{w(a) : a \in [k]\}$, where $w(v)$ is a variable word over k.
- Given a variable word $w(v) = (\alpha_0, ..., \alpha_{n-1})$, the set $\{i \in \{0, ..., n-1\} : w_i = v\}$ is called the wildcard set of $w(v)$.

Moreover, let v be a symbol not belonging to $[k] = \{1, ..., k\}$.

- A variable word over k is a finite sequence $w(v)$ in $[k] \cup \{v\}$ such that v occurs at least once.
- For a variable word $w(v)$ and $a \in [k]$ by $w(a)$ we denote the constant word over k resulting by substituting every occurrence of v by a.
- A combinatorial line is a set of the form $\{w(a) : a \in [k]\}$, where $w(v)$ is a variable word over k.
- Given a variable word $w(v) = (\alpha_0, ..., \alpha_{n-1})$, the set $\{i \in \{0, ..., n-1\} : w_i = v\}$ is called the wildcard set of $w(v)$.

Moreover, let v be a symbol not belonging to $[k] = \{1, ..., k\}$.

- A variable word over k is a finite sequence $w(v)$ in $[k] \cup \{v\}$ such that v occurs at least once.
- For a variable word $w(v)$ and $a \in [k]$ by $w(a)$ we denote the constant word over k resulting by substituting every occurrence of v by a.
- A combinatorial line is a set of the form $\{w(a) : a \in [k]\}$, where $w(v)$ is a variable word over k.
- Given a variable word $w(v) = (\alpha_0, ..., \alpha_{n-1})$, the set $\{i \in \{0, ..., n-1\} : w_i = v\}$ is called the wildcard set of $w(v)$.

Moreover, let v be a symbol not belonging to $[k] = \{1, ..., k\}$.

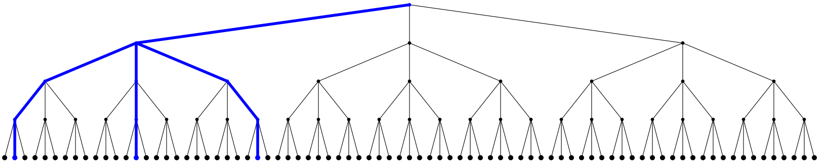
- A variable word over k is a finite sequence $w(v)$ in $[k] \cup \{v\}$ such that v occurs at least once.
- For a variable word $w(v)$ and $a \in [k]$ by $w(a)$ we denote the constant word over k resulting by substituting every occurrence of v by a.
- A combinatorial line is a set of the form $\{w(a) : a \in [k]\}$, where $w(v)$ is a variable word over k.
- Given a variable word $w(v) = (\alpha_0, ..., \alpha_{n-1})$, the set $\{i \in \{0, ..., n-1\} : w_i = v\}$ is called the wildcard set of $w(v)$.

Moreover, let v be a symbol not belonging to $[k] = \{1, ..., k\}$.

- A variable word over k is a finite sequence $w(v)$ in $[k] \cup \{v\}$ such that v occurs at least once.
- For a variable word $w(v)$ and $a \in [k]$ by $w(a)$ we denote the constant word over k resulting by substituting every occurrence of v by a.
- A combinatorial line is a set of the form $\{w(a) : a \in [k]\}$, where $w(v)$ is a variable word over k.
- Given a variable word $w(v) = (\alpha_0, ..., \alpha_{n-1})$, the set $\{i \in \{0, ..., n-1\} : w_i = v\}$ is called the wildcard set of $w(v)$.

For example assume that $k = 3$ and $n = 4$. Also let $w(v) = (1, v, v, 2)$. Then the corresponding combinatorial line is the set

$$\{(1, \mathbf{1}, \mathbf{1}, 2), (1, \mathbf{2}, \mathbf{2}, 2), (1, \mathbf{3}, \mathbf{3}, 2)\}$$



The wildcard set is $\{1, 2\}$.

The Hales–Jewett Theorem

Theorem (Hales and Jewett, 1963)

Let k, r be positive integers. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every r-coloring of $[k]^n$ there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is monochromatic.

- The least such n_0 is denoted by $HJ(k, r)$.
- The best known upper bounds for the numbers $HJ(k, r)$ are primitive recursive and are due to Shelah (1988).

The Hales–Jewett Theorem

Theorem (Hales and Jewett, 1963)

Let k, r be positive integers. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every r-coloring of $[k]^n$ there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is monochromatic.

- The least such n_0 is denoted by $HJ(k, r)$.
- The best known upper bounds for the numbers $HJ(k, r)$ are primitive recursive and are due to Shelah (1988).

The Hales–Jewett Theorem

Theorem (Hales and Jewett, 1963)

Let k, r be positive integers. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every r-coloring of $[k]^n$ there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is monochromatic.

- The least such n_0 is denoted by $HJ(k, r)$.
- The best known upper bounds for the numbers $HJ(k, r)$ are primitive recursive and are due to Shelah (1988).

Taking Van der Waerden's Theorem as a consequence

Let $Q : [k]^n \to \{0, ..., k^n - 1\}$ defined by

$$Q\big((a_0, ..., a_{n-1})\big) = \sum_{i=0}^{n-1} (a_i - 1)k^i$$

Notice that Q is 1-1 and onto.
Moreover, notice that the image through Q of every combinatorial line is an arithmetic progression of length k. In particular, if $w(v) = (\alpha_0, ..., \alpha_{n-1})$ is a variable word over k and X is the wildcard set of $w(v)$ then the first term of the resulting arithmetic progression is $a = \sum_{i \notin X} (w_i - 1)k^i$ and the step is $d = \sum_{i \in X} k^i$.

Let $Q : [k]^n \to \{0, ..., k^n - 1\}$ defined by

$$Q\big((a_0, ..., a_{n-1})\big) = \sum_{i=0}^{n-1} (a_i - 1)k^i$$

Notice that Q is 1-1 and onto.

Moreover, notice that the image through Q of every combinatorial line is an arithmetic progression of length k. In particular, if $w(v) = (\alpha_0, ..., \alpha_{n-1})$ is a variable word over k and X is the wildcard set of $w(v)$ then the first term of the resulting arithmetic progression is $a = \sum_{i \notin X} (w_i - 1)k^i$ and the step is $d = \sum_{i \in X} k^i$.

Taking Van der Waerden's Theorem as a consequence

Let $Q : [k]^n \to \{0, ..., k^n - 1\}$ defined by

$$Q\big((a_0, ..., a_{n-1})\big) = \sum_{i=0}^{n-1} (a_i - 1)k^i$$

Notice that Q is 1-1 and onto.
Moreover, notice that the image through Q of every combinatorial line is an arithmetic progression of length k. In particular, if $w(v) = (\alpha_0, ..., \alpha_{n-1})$ is a variable word over k and X is the wildcard set of $w(v)$ then the first term of the resulting arithmetic progression is $a = \sum_{i \notin X} (w_i - 1)k^i$ and the step is $d = \sum_{i \in X} k^i$.

The Density Hales–Jewett Theorem

Theorem (Furstenberg and Katznelson, 1991)

Let k be positive integer and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every subset A of $[k]^n$ of uniform density at least δ, that is $\frac{|A|}{k^n} \geqslant \delta$, there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is subset of A.

- The least such n_0 is denoted by $DHJ(k, \delta)$.
- Furstenberg and Katznelson proved it using Ergodic Theory and they provide no information on $DHJ(k, \delta)$.
- In 2011, Austin gave another proof using Ergodic theoretic techniques.
- A combinatorial proof is provided by the Polymath paper (2012), giving upper bounds for the numbers $DHJ(k, \delta)$ which have an Ackermann type dependence on k.
- Tao presents another proof on his blog.
- In a joint work with Dodos and Kanellopoulos a simpler combinatorial proof is given.

The Density Hales–Jewett Theorem

Theorem (Furstenberg and Katznelson, 1991)

Let k be positive integer and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every subset A of $[k]^n$ of uniform density at least δ, that is $\frac{|A|}{k^n} \geqslant \delta$, there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is subset of A.

- The least such n_0 is denoted by $DHJ(k, \delta)$.
- Furstenberg and Katznelson proved it using Ergodic Theory and they provide no information on $DHJ(k, \delta)$.
- In 2011, Austin gave another proof using Ergodic theoretic techniques.
- A combinatorial proof is provided by the Polymath paper (2012), giving upper bounds for the numbers $DHJ(k, \delta)$ which have an Ackermann type dependence on k.
- Tao presents another proof on his blog.
- In a joint work with Dodos and Kanellopoulos a simpler combinatorial proof is given.

The Density Hales–Jewett Theorem

Theorem (Furstenberg and Katznelson, 1991)

Let k be positive integer and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every subset A of $[k]^n$ of uniform density at least δ, that is $\frac{|A|}{k^n} \geqslant \delta$, there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is subset of A.

- The least such n_0 is denoted by $DHJ(k, \delta)$.
- Furstenberg and Katznelson proved it using Ergodic Theory and they provide no information on $DHJ(k, \delta)$.
- In 2011, Austin gave another proof using Ergodic theoretic techniques.
- A combinatorial proof is provided by the Polymath paper (2012), giving upper bounds for the numbers $DHJ(k, \delta)$ which have an Ackermann type dependence on k.
- Tao presents another proof on his blog.
- In a joint work with Dodos and Kanellopoulos a simpler combinatorial proof is given.

The Density Hales–Jewett Theorem

Theorem (Furstenberg and Katznelson, 1991)

Let k be positive integer and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every subset A of $[k]^n$ of uniform density at least δ, that is $\frac{|A|}{k^n} \geqslant \delta$, there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is subset of A.

- The least such n_0 is denoted by $DHJ(k, \delta)$.
- Furstenberg and Katznelson proved it using Ergodic Theory and they provide no information on $DHJ(k, \delta)$.
- In 2011, Austin gave another proof using Ergodic theoretic techniques.
- A combinatorial proof is provided by the Polymath paper (2012), giving upper bounds for the numbers $DHJ(k, \delta)$ which have an Ackermann type dependence on k.
- Tao presents another proof on his blog.
- In a joint work with Dodos and Kanellopoulos a simpler combinatorial proof is given.

The Density Hales–Jewett Theorem

Theorem (Furstenberg and Katznelson, 1991)

Let k be positive integer and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every subset A of $[k]^n$ of uniform density at least δ, that is $\frac{|A|}{k^n} \geqslant \delta$, there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is subset of A.

- The least such n_0 is denoted by $DHJ(k, \delta)$.
- Furstenberg and Katznelson proved it using Ergodic Theory and they provide no information on $DHJ(k, \delta)$.
- In 2011, Austin gave another proof using Ergodic theoretic techniques.
- A combinatorial proof is provided by the Polymath paper (2012), giving upper bounds for the numbers $DHJ(k, \delta)$ which have an Ackermann type dependence on k.
- Tao presents another proof on his blog.
- In a joint work with Dodos and Kanellopoulos a simpler combinatorial proof is given.

The Density Hales–Jewett Theorem

Theorem (Furstenberg and Katznelson, 1991)

Let k be positive integer and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every subset A of $[k]^n$ of uniform density at least δ, that is $\frac{|A|}{k^n} \geqslant \delta$, there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is subset of A.

- The least such n_0 is denoted by $DHJ(k, \delta)$.
- Furstenberg and Katznelson proved it using Ergodic Theory and they provide no information on $DHJ(k, \delta)$.
- In 2011, Austin gave another proof using Ergodic theoretic techniques.
- A combinatorial proof is provided by the Polymath paper (2012), giving upper bounds for the numbers $DHJ(k, \delta)$ which have an Ackermann type dependence on k.
- Tao presents another proof on his blog.
- In a joint work with Dodos and Kanellopoulos a simpler combinatorial proof is given.

The Density Hales–Jewett Theorem

Theorem (Furstenberg and Katznelson, 1991)

Let k be positive integer and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 such that for every $n \geqslant n_0$ and every subset A of $[k]^n$ of uniform density at least δ, that is $\frac{|A|}{k^n} \geqslant \delta$, there exists a variable word $w(v)$ of length n such that the set $\{w(a) : a \in [k]\}$ is subset of A.

- The least such n_0 is denoted by $DHJ(k, \delta)$.
- Furstenberg and Katznelson proved it using Ergodic Theory and they provide no information on $DHJ(k, \delta)$.
- In 2011, Austin gave another proof using Ergodic theoretic techniques.
- A combinatorial proof is provided by the Polymath paper (2012), giving upper bounds for the numbers $DHJ(k, \delta)$ which have an Ackermann type dependence on k.
- Tao presents another proof on his blog.
- In a joint work with Dodos and Kanellopoulos a simpler combinatorial proof is given.

A regularity Lemma

Let k, n be positive integers, A a subset of $[k]^n$ and I a subset of $n = \{0, ..., n-1\}$.

- $[k]^n \equiv [k]^I \times [k]^{n \setminus I}$.
- We set $A_x = \{y \in [k]^{n \setminus I} : x \cup y \in A\}$ for all $x \in [k]^I$.

Lemma

Let k, m be positive integers and $0 < \varepsilon \leqslant 1$. Then there exists a positive integer n_0 with the following property. For every $n \geqslant n_0$ and subset A of $[k]^n$ there exists a subinterval I of $\{0, ..., n-1\}$ of length m such that

$$\left| \frac{|A_x|}{|[k]^{n \setminus I}|} - \frac{|A|}{|[k]^n|} \right| < \varepsilon$$

for all $x \in [k]^I$.

A regularity Lemma

Let k, n be positive integers, A a subset of $[k]^n$ and I a subset of $n = \{0, ..., n-1\}$.

- $[k]^n \equiv [k]^I \times [k]^{n \setminus I}$.
- We set $A_x = \{y \in [k]^{n \setminus I} : x \cup y \in A\}$ for all $x \in [k]^I$.

Lemma

Let k, m be positive integers and $0 < \varepsilon \leqslant 1$. Then there exists a positive integer n_0 with the following property. For every $n \geqslant n_0$ and subset A of $[k]^n$ there exists a subinterval I of $\{0, ..., n-1\}$ of length m such that

$$\left| \frac{|A_x|}{|[k]^{n \setminus I}|} - \frac{|A|}{|[k]^n|} \right| < \varepsilon$$

for all $x \in [k]^I$.

A regularity Lemma

Let k, n be positive integers, A a subset of $[k]^n$ and I a subset of $n = \{0, ..., n-1\}$.

- $[k]^n \equiv [k]^I \times [k]^{n \setminus I}$.
- We set $A_x = \{y \in [k]^{n \setminus I} : x \cup y \in A\}$ for all $x \in [k]^I$.

Lemma

Let k, m be positive integers and $0 < \varepsilon \leqslant 1$. Then there exists a positive integer n_0 with the following property. For every $n \geqslant n_0$ and subset A of $[k]^n$ there exists a subinterval I of $\{0, ..., n-1\}$ of length m such that

$$\left| \frac{|A_x|}{|[k]^{n \setminus I}|} - \frac{|A|}{|[k]^n|} \right| < \varepsilon$$

for all $x \in [k]^I$.

A regularity Lemma

Let k, n be positive integers, A a subset of $[k]^n$ and I a subset of $n = \{0, ..., n-1\}$.

- $[k]^n \equiv [k]^I \times [k]^{n \setminus I}$.
- We set $A_x = \{y \in [k]^{n \setminus I} : x \cup y \in A\}$ for all $x \in [k]^I$.

Lemma

Let k, m be positive integers and $0 < \varepsilon \leqslant 1$. Then there exists a positive integer n_0 with the following property. For every $n \geqslant n_0$ and subset A of $[k]^n$ there exists a subinterval I of $\{0, ..., n-1\}$ of length m such that

$$\left| \frac{|A_x|}{|[k]^{n \setminus I}|} - \frac{|A|}{|[k]^n|} \right| < \varepsilon$$

for all $x \in [k]^I$.

Carlson–Simpson Trees

Let k be a positive integer.
By $[k]^{<\omega}$ we denote the set of all words over k.
If c is a constant word over k and $\big(w_q(v)\big)_q$ a sequence of left variable words over k, then the set

$$\{c\} \cup \{c^\frown w_0(a_0)^\frown \ldots ^\frown w_n(a_n) : n \in \omega \text{ and } a_0, \ldots, a_n \in [k]\}$$

is called an infinite Carlson–Simpson tree.
Moreover, if c is a constant word over k and $w_0(v), \ldots, w_{m-1}(v)$ left variable words over k then the set

$$\{c\} \cup \{c^\frown w_0(a_0)^\frown \ldots ^\frown w_n(a_n) : n \in \{0, \ldots, m-1\} \text{ and } a_0, \ldots, a_n \in [k]\}$$

is called a Carlson–Simpson tree of dimension m.

Let k be a positive integer.
By $[k]^{<\omega}$ we denote the set of all words over k.
If c is a constant word over k and $\left(w_q(v)\right)_q$ a sequence of left variable words over k, then the set

$$\{c\} \cup \{c^\frown w_0(a_0)^\frown ...^\frown w_n(a_n) : n \in \omega \text{ and } a_0, ..., a_n \in [k]\}$$

is called an infinite Carlson–Simpson tree.
Moreover, if c is a constant word over k and $w_0(v), ..., w_{m-1}(v)$ left variable words over k then the set

$$\{c\} \cup \{c^\frown w_0(a_0)^\frown ...^\frown w_n(a_n) : n \in \{0, ..., m-1\} \text{ and } a_0, ..., a_n \in [k]\}$$

is called a Carlson–Simpson tree of dimension m.

Let k be a positive integer.
By $[k]^{<\omega}$ we denote the set of all words over k.
If c is a constant word over k and $\big(w_q(v)\big)_q$ a sequence of left variable words over k, then the set

$$\{c\} \cup \{c^\frown w_0(a_0)^\frown \ldots ^\frown w_n(a_n) : n \in \omega \text{ and } a_0, \ldots, a_n \in [k]\}$$

is called an infinite Carlson–Simpson tree.
Moreover, if c is a constant word over k and $w_0(v), \ldots, w_{m-1}(v)$ left variable words over k then the set

$$\{c\} \cup \{c^\frown w_0(a_0)^\frown \ldots ^\frown w_n(a_n) : n \in \{0, \ldots, m-1\} \text{ and } a_0, \ldots, a_n \in [k]\}$$

is called a Carlson–Simpson tree of dimension m.

Carlson–Simpson Trees

Let k be a positive integer.
By $[k]^{<\omega}$ we denote the set of all words over k.
If c is a constant word over k and $\left(w_q(v)\right)_q$ a sequence of left variable words over k, then the set

$$\{c\} \cup \{c^\frown w_0(a_0)^\frown \ldots {}^\frown w_n(a_n) : n \in \omega \text{ and } a_0, \ldots, a_n \in [k]\}$$

is called an infinite Carlson–Simpson tree.
Moreover, if c is a constant word over k and $w_0(v), \ldots, w_{m-1}(v)$ left variable words over k then the set

$$\{c\} \cup \{c^\frown w_0(a_0)^\frown \ldots {}^\frown w_n(a_n) : n \in \{0, \ldots, m-1\} \text{ and } a_0, \ldots, a_n \in [k]\}$$

is called a Carlson–Simpson tree of dimension m.

Theorem (Carlson and Simpson, 1984)

Let k be a positive integer. Then for every finite coloring of $[k]^{<\omega}$ *there exists an monochromatic infinite Carlson–Simpson tree.*

- The Carlson–Simpson Theorem belongs to the circle of results that provide information on the structure of the wildcard set of the variable word obtained by the Hales–Jewett Theorem.

Theorem (Carlson and Simpson, 1984)

Let k be a positive integer. Then for every finite coloring of $[k]^{<\omega}$ *there exists an monochromatic infinite Carlson–Simpson tree.*

- The Carlson–Simpson Theorem belongs to the circle of results that provide information on the structure of the wildcard set of the variable word obtained by the Hales–Jewett Theorem.

The Density Carlson–Simpson Theorem

Theorem (Dodos, Kanellopoulos and K. T., 2013)

Let k be a positive integer. Then every subset A of $[k]^{<\omega}$*, satisfying*

$$\limsup_{n\to\infty} \frac{|A \cap [k]^n|}{|[k]^n|} > 0,$$

(equivalently, there is an infinite subset L of ω *and a real* $0 < \delta \leqslant 1$ *such that*

$$\frac{|A \cap [k]^n|}{|[k]^n|} > \delta$$

for all $n \in L$*)*

contains an infinite Carlson–Simpson tree.

The Density Carlson–Simpson Theorem

Theorem (Dodos, Kanellopoulos and K. T., 2013)

Let k be a positive integer. Then every subset A of $[k]^{<\omega}$, *satisfying*

$$\limsup_{n\to\infty} \frac{|A \cap [k]^n|}{|[k]^n|} > 0,$$

(equivalently, there is an infinite subset L of ω *and a real* $0 < \delta \leqslant 1$ *such that*

$$\frac{|A \cap [k]^n|}{|[k]^n|} > \delta$$

for all $n \in L$*)*
contains an infinite Carlson–Simpson tree.

- The above is consequence of the following.

The Density Carlson–Simpson Theorem

Theorem (Dodos, Kanellopoulos and K. T., 2013)

Let k be a positive integer. Then every subset A of $[k]^{<\omega}$, satisfying

$$\limsup_{n\to\infty} \frac{|A \cap [k]^n|}{|[k]^n|} > 0,$$

(equivalently, there is an infinite subset L of ω and a real $0 < \delta \leqslant 1$ such that

$$\frac{|A \cap [k]^n|}{|[k]^n|} > \delta$$

for all $n \in L$)
contains an infinite Carlson–Simpson tree.

- The above is consequence of the following.

The Density Carlson–Simpson Theorem

Theorem (Dodos, Kanellopoulos and K. T., 2013)

Let k, m be positive integers and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 with the following property. Every subset A of $[k]^{<\omega}$ satisfying

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for at least n_0 many n's, (equivalently there exists a subset L of ω and a real $0 < \delta \leqslant 1$ such that $|L| \geqslant n_0$ and

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for all $n \in L$)

contains a Carlson–Simpson tree of dimension m.

The least such n_0 is denoted by $DCS(k, m, \delta)$.

The Density Carlson–Simpson Theorem

Theorem (Dodos, Kanellopoulos and K. T., 2013)

Let k, m be positive integers and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 with the following property. Every subset A of $[k]^{<\omega}$ satisfying

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for at least n_0 many n's, (equivalently there exists a subset L of ω and a real $0 < \delta \leqslant 1$ such that $|L| \geqslant n_0$ and

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for all $n \in L$)

contains a Carlson–Simpson tree of dimension m.

The least such n_0 is denoted by $DCS(k, m, \delta)$.

Theorem (Dodos, Kanellopoulos and K. T., 2013)

Let k, m be positive integers and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 with the following property. Every subset A of $[k]^{<\omega}$ satisfying

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for at least n_0 many n's, (equivalently there exists a subset L of ω and a real $0 < \delta \leqslant 1$ such that $|L| \geqslant n_0$ and

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for all $n \in L$)
contains a Carlson–Simpson tree of dimension m.

The least such n_0 is denoted by $DCS(k, m, \delta)$.

Theorem (Dodos, Kanellopoulos and K. T., 2013)

Let k, m be positive integers and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 with the following property. Every subset A of $[k]^{<\omega}$ satisfying

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for at least n_0 many n's, (equivalently there exists a subset L of ω and a real $0 < \delta \leqslant 1$ such that $|L| \geqslant n_0$ and

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for all $n \in L$)

contains a Carlson–Simpson tree of dimension m.

Theorem (Dodos, Kanellopoulos and K. T., 2013)

Let k, m be positive integers and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 with the following property. Every subset A of $[k]^{<\omega}$ satisfying

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for at least n_0 many n's, (equivalently there exists a subset L of ω and a real $0 < \delta \leqslant 1$ such that $|L| \geqslant n_0$ and

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for all $n \in L$)
contains a Carlson–Simpson tree of dimension m.

- The least such n_0 is denoted by $DCS(k, m, \delta)$.

The Density Carlson–Simpson Theorem

Theorem (Dodos, Kanellopoulos and K. T., 2013)

Let k, m be positive integers and δ be a real with $0 < \delta \leqslant 1$. Then there exists an integer n_0 with the following property. Every subset A of $[k]^{<\omega}$ satisfying

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for at least n_0 many n's, (equivalently there exists a subset L of ω and a real $0 < \delta \leqslant 1$ such that $|L| \geqslant n_0$ and

$$\frac{|A \cap [k]^n|}{|[k]^n|} \geqslant \delta$$

for all $n \in L$)
contains a Carlson–Simpson tree of dimension m.

- The least such n_0 is denoted by $DCS(k, m, \delta)$.

Relation between the statements

The Density Hales–Jewett Theorem is a consequence of the Density Carlson–Simpson Theorem. In particular, we have that

$$DHJ(k,\delta) \leqslant DCS(k,1,\delta).$$

Probabilistic versions

Let δ be a real such that $0 < \delta \leqslant 1$. Assume that we are given a family $\{A_i : i \in \mathcal{S}\}$ of measurable events in a probability space (Ω, Σ, μ) indexed by a Ramsey space such that $\mu(A_i) \geqslant \delta$ for all $i \in \mathcal{S}$.
Can we find a substructure $\mathcal{S}'$ of $\mathcal{S}$ such that the events $\{A_i : i \in \mathcal{S}'\}$ are highly correlated?

Let δ be a real such that $0 < \delta \leqslant 1$. Assume that we are given a family $\{A_i : i \in \mathcal{S}\}$ of measurable events in a probability space (Ω, Σ, μ) indexed by a Ramsey space such that $\mu(A_i) \geqslant \delta$ for all $i \in \mathcal{S}$.
Can we find a substructure $\mathcal{S}'$ of $\mathcal{S}$ such that the events $\{A_i : i \in \mathcal{S}'\}$ are highly correlated?

Probabilistic version

The simplest example towards this direction is the following.
Let $0 < \delta \leqslant 1$ and $\{A_i : i \in \omega\}$ be a sequence of measurable events in a probability space (Ω, Σ, μ) such that $\mu(A_i) \geqslant \delta$ for all $i \in \omega$. Then for every $0 < \theta < \delta$ there exists a infinite L subset of ω such that for every finite F subset of L we have that

$$\mu\Big(\bigcap_{i \in F} A_i\Big) \geqslant \theta^{|F|}.$$

Probabilistic version

Theorem (Dodos, Kanellopoulos and K.T., 2013)

Let k be a positive integer and δ a real with $0 < \delta \leqslant 1$. Then for every $n \geqslant 1$ there exist a positive real $\theta(k, \delta, n)$ and a positive integer $Cor(k, \delta, n)$ having the following property. For every $m \geqslant 1$ and every family $\{A_t : t \in [k]^{<Cor(k,\delta,m)}\}$ of measurable events such that $\mu(A_t) \geqslant \delta$ for all $t \in [k]^{<Cor(k,\delta,m)}$, there exists an m-dimensional Carlson–Simpson subtree S of $[k]^{<Cor(k,\delta,m)}$ such that for every finite subset F of S we have that

$$\mu\Big(\bigcap_{t \in F} A_t\Big) \geqslant \theta(k, \delta, |F|).$$

Probabilistic version

Theorem (Dodos, Kanellopoulos and K.T., 2013)

Let k be a positive integer and δ a real with $0 < \delta \leqslant 1$. Then for every $n \geqslant 1$ there exist a positive real $\theta(k,\delta,n)$ and a positive integer $Cor(k,\delta,n)$ having the following property. For every $m \geqslant 1$ and every family $\{A_t : t \in [k]^{<Cor(k,\delta,m)}\}$ of measurable events such that $\mu(A_t) \geqslant \delta$ for all $t \in [k]^{<Cor(k,\delta,m)}$,

there exists an m-dimensional Carlson–Simpson subtree S of $[k]^{<Cor(k,\delta,m)}$ such that for every finite subset F of S we have that

$$\mu\Big(\bigcap_{t\in F} A_t\Big) \geqslant \theta(k,\delta,|F|).$$

Probabilistic version

Theorem (Dodos, Kanellopoulos and K.T., 2013)

Let k be a positive integer and δ a real with $0 < \delta \leqslant 1$. Then for every $n \geqslant 1$ there exist a positive real $\theta(k,\delta,n)$ and a positive integer $Cor(k,\delta,n)$ having the following property. For every $m \geqslant 1$ and every family $\{A_t : t \in [k]^{<Cor(k,\delta,m)}\}$ of measurable events such that $\mu(A_t) \geqslant \delta$ for all $t \in [k]^{<Cor(k,\delta,m)}$, there exists an m-dimensional Carlson–Simpson subtree S of $[k]^{<Cor(k,\delta,m)}$ such that for every finite subset F of S we have that

$$\mu\Big(\bigcap_{t\in F} A_t\Big) \geqslant \theta(k,\delta,|F|).$$

Conjecture

For n positive integer, by $[2]^{n\times n}$ we denote the set of all maps with domain $n \times n$ and codomain $[2]$. A variable word is a map with domain $n \times n$ and codomain $[2] \cup \{v\}$ such that the support of v is of the form $X \times X$ for some non-empty subset X of n.

Conjecture (Bergelson)

For every $\delta > 0$ there exists some n_0 with the following property. For every $n \geqslant n_0$ and every subset A of $[2]^{[n]\times[n]}$ there exists a variable word $w(v)$ such that $\{w(a) : a \in [2]\}$ is subset of A.