Unavoidable patterns in words

Benny Sudakov
ETH, Zurich

joint with D.Conlon and J. Fox

Definition:

The Ramsey number $r_k(n)$ is the minimum N such that every 2-coloring of the k-tuples of an N-element set contains a monochromatic set of order n.

Ramsey numbers

Definition:

The Ramsey number $r_k(n)$ is the minimum N such that every 2-coloring of the k-tuples of an N-element set contains a monochromatic set of order n.

Theorem: (Ramsey 1930)

For all k, n, the Ramsey number $r_k(n)$ is finite.

Ramsey numbers

Definition:

The Ramsey number $r_k(n)$ is the minimum N such that every 2-coloring of the k-tuples of an N-element set contains a monochromatic set of order n.

Theorem: (Ramsey 1930)

For all k, n, the Ramsey number $r_k(n)$ is finite.

Question:

Estimate the growth rate of $r_k(n)$.

Theorem:

$2^{n/2} \leq r_2(n) \leq 2^{2n}$. *(Erdős'47, Erdős–Szekeres'35)*

Bounds on Ramsey numbers

Theorem:

$$2^{n/2} \leq r_2(n) \leq 2^{2n}. \qquad \text{(Erdős'47, Erdős–Szekeres'35)}$$

$$2^{cn^2} \leq r_3(n) \leq 2^{2^{c'n}}. \qquad \text{(Erdős–Rado'52, Erdős–Hajnal'60s)}$$

Theorem:

$2^{n/2} \leq r_2(n) \leq 2^{2n}$. *(Erdős'47, Erdős–Szekeres'35)*

$2^{cn^2} \leq r_3(n) \leq 2^{2^{c'n}}$. *(Erdős–Rado'52, Erdős–Hajnal'60s)*

Remarks:

- There is a similar gap of one exponential between the upper and the lower bound for $r_k(n)$ for $k > 3$. These bounds are towers of exponentials of height k and $k-1$ respectively.

THEOREM:

$$2^{n/2} \leq r_2(n) \leq 2^{2n}. \qquad \text{(Erdős'47, Erdős–Szekeres'35)}$$

$$2^{cn^2} \leq r_3(n) \leq 2^{2^{c'n}}. \qquad \text{(Erdős–Rado'52, Erdős–Hajnal'60s)}$$

Remarks:

- There is a similar gap of one exponential between the upper and the lower bound for $r_k(n)$ for $k > 3$. These bounds are towers of exponentials of height k and $k-1$ respectively.
- Determining the behavior of $r_3(n)$ will close the gap for all k due to stepping-up lemma of Erdős–Hajnal, which constructs lower bound colorings for uniformity $k+1$ from colorings for uniformity k, effectively gaining an extra exponential each time it is applied.

Definition

- Words and patterns are strings of characters over fixed alphabets.

Definition

- Words and patterns are strings of characters over fixed alphabets.
- A subword of a word is a block of consecutive letters.

Words and patterns

Definition

- Words and patterns are strings of characters over fixed alphabets.
- A subword of a word is a block of consecutive letters.
- A word w contains the pattern P if there is a way to substitute a nonempty word for each letter in P so that the resulting word is a subword of w.

Words and patterns

Definition

- Words and patterns are strings of characters over fixed alphabets.
- A subword of a word is a block of consecutive letters.
- A word w contains the pattern P if there is a way to substitute a nonempty word for each letter in P so that the resulting word is a subword of w.

Example:
The word *mathematics* contains the pattern $xyxz$ with $x = mat$, $y = he$ and $z = ics$.

Definition

- Words and patterns are strings of characters over fixed alphabets.
- A subword of a word is a block of consecutive letters.
- A word w contains the pattern P if there is a way to substitute a nonempty word for each letter in P so that the resulting word is a subword of w.

Example:
The word *mathematics* contains the pattern $xyxz$ with $x = mat$, $y = he$ and $z = ics$.

q-UNAVOIDABILITY

DEFINITION:

A pattern P is q-unavoidable if every sufficiently long word over an alphabet of size q contains a copy of P.

DEFINITION:

A pattern P is q-unavoidable if every sufficiently long word over an alphabet of size q contains a copy of P.

Examples:

- THUE 1906: The pattern xx is 2-unavoidable, but 3-avoidable.

DEFINITION:

A pattern P is q-unavoidable if every sufficiently long word over an alphabet of size q contains a copy of P.

Examples:

- THUE 1906: The pattern xx is 2-unavoidable, but 3-avoidable.
- THUE 1912, MORSE 1921: The pattern xxx is 1-unavoidable, but 2-avoidable.

DEFINITION:

A pattern P is q-unavoidable if every sufficiently long word over an alphabet of size q contains a copy of P.

Examples:

- THUE 1906: The pattern xx is 2-unavoidable, but 3-avoidable.
- THUE 1912, MORSE 1921: The pattern xxx is 1-unavoidable, but 2-avoidable.

  Start with a and recursively substitute $a \mapsto ab$ and $b \mapsto ba$.

Definition:

A pattern P is q-unavoidable if every sufficiently long word over an alphabet of size q contains a copy of P.

Examples and applications of pattern avoidance:

- Combinatorics
- Group theory, e.g, Burnside problem, Undecidability
- Symbolic Dynamics
- Number theory

Unavoidability

Definition:

A pattern P is unavoidable if it is q-unavoidable for all $q \geq 1$.

Unavoidability

Definition:

A pattern P is unavoidable if it is q-unavoidable for all $q \geq 1$.

Ramsey question for patterns:

Which patterns are unavoidable?

Unavoidability

Definition:

A pattern P is unavoidable if it is q-unavoidable for all $q \geq 1$.

Ramsey question for patterns:

Which patterns are unavoidable?

Theorem: (Bean–Ehrenfeucht–McNulty 1979, Zimin 1984)

A word is unavoidable if and only if it is contained in a Zimin word, defined recursively by $Z_1 = x_1$ and $Z_n = Z_{n-1} x_n Z_{n-1}$.

Definition:

A pattern P is unavoidable if it is q-unavoidable for all $q \geq 1$.

Ramsey question for patterns:

Which patterns are unavoidable?

Theorem: (Bean–Ehrenfeucht–McNulty 1979, Zimin 1984)

A word is unavoidable if and only if it is contained in a Zimin word, defined recursively by $Z_1 = x_1$ and $Z_n = Z_{n-1} x_n Z_{n-1}$.

Zimin words:

$Z_1 = x$

Unavoidability

Definition:

A pattern P is unavoidable if it is q-unavoidable for all $q \geq 1$.

Ramsey question for patterns:

Which patterns are unavoidable?

Theorem: (Bean–Ehrenfeucht–McNulty 1979, Zimin 1984)

A word is unavoidable if and only if it is contained in a Zimin word, defined recursively by $Z_1 = x_1$ and $Z_n = Z_{n-1}x_nZ_{n-1}$.

Zimin words:

$Z_1 = x$ $\qquad$ $Z_2 = xyx$

DEFINITION:

A pattern P is unavoidable if it is q-unavoidable for all $q \geq 1$.

RAMSEY QUESTION FOR PATTERNS:

Which patterns are unavoidable?

THEOREM: (BEAN–EHRENFEUCHT–MCNULTY 1979, ZIMIN 1984)

A word is unavoidable if and only if it is contained in a Zimin word, defined recursively by $Z_1 = x_1$ and $Z_n = Z_{n-1}x_nZ_{n-1}$.

Zimin words:

$Z_1 = x$ $\qquad Z_2 = xyx$ $\qquad Z_3 = xyxzxyx$

Unavoidability

Definition:

A pattern P is unavoidable if it is q-unavoidable for all $q \geq 1$.

Ramsey question for patterns:

Which patterns are unavoidable?

Theorem: (Bean–Ehrenfeucht–McNulty 1979, Zimin 1984)

A word is unavoidable if and only if it is contained in a Zimin word, defined recursively by $Z_1 = x_1$ and $Z_n = Z_{n-1} x_n Z_{n-1}$.

Zimin words:

$Z_1 = x$ $\quad Z_2 = xyx$ $\quad Z_3 = xyxzxyx$

$Z_4 = xyxzxyxwxyxzxyx$

Definition:

Let $f(n,q)$ be the smallest natural number such that any word of length $f(n,q)$ over an alphabet of size q contains Zimin word Z_n.

Ramsey numbers for patterns

Definition:

Let $f(n,q)$ be the smallest natural number such that any word of length $f(n,q)$ over an alphabet of size q contains Zimin word Z_n.

Problem: Estimate the asymptotics of $f(n,q)$.

Ramsey numbers for patterns

Definition:

Let $f(n,q)$ be the smallest natural number such that any word of length $f(n,q)$ over an alphabet of size q contains Zimin word Z_n.

Problem: Estimate the asymptotics of $f(n,q)$.

Few upper bounds:

- $f(1,q) = 1$, follows easily from $Z_1 = x$.

Definition:

Let $f(n, q)$ be the smallest natural number such that any word of length $f(n, q)$ over an alphabet of size q contains Zimin word Z_n.

Problem: Estimate the asymptotics of $f(n, q)$.

Few upper bounds:

- $f(1, q) = 1$, follows easily from $Z_1 = x$.
- $f(2, q) = 2q + 1$, follows easily from $Z_2 = xyx$.

Ramsey numbers for patterns

Definition:

Let $f(n, q)$ be the smallest natural number such that any word of length $f(n, q)$ over an alphabet of size q contains Zimin word Z_n.

Problem: Estimate the asymptotics of $f(n, q)$.

Few upper bounds:

- $f(1, q) = 1$, follows easily from $Z_1 = x$.
- $f(2, q) = 2q + 1$, follows easily from $Z_2 = xyx$.
- $f(3, q) \leq q^q$ (Rytter–Shur)

General upper bound

Lemma: (Cooper–Rorabaugh)

$f(n+1, q) \leq (f(n,q)+1)(q^{f(n,q)}+1)-1$

Lemma: (Cooper–Rorabaugh)

$f(n+1,q) \leq (f(n,q)+1)(q^{f(n,q)}+1)-1$

Proof: Given some word w, split it into $m = q^{f(n,q)}+1$ words w_i of length $f(n,q)$, which are separated by single letters:

$$w_1 \, x \, w_2 \, y \ldots z \, w_m$$

Then there are two identical words w_i and w_j, each containing the same copy of some Zimin word Z_n. This forms Z_{n+1}. □

General upper bound

Lemma: (Cooper–Rorabaugh)

$f(n+1,q) \leq (f(n,q)+1)(q^{f(n,q)}+1)-1$

Proof: Given some word w, split it into $m = q^{f(n,q)}+1$ words w_i of length $f(n,q)$, which are separated by single letters:

$$w_1\, x\, w_2\, y \ldots z\, w_m$$

Then there are two identical words w_i and w_j, each containing the same copy of some Zimin word Z_n. This forms Z_{n+1}. $\square$

Theorem:

$$f(n,q) \leq q^{\cdot^{\cdot^{\cdot^{q}}}} \Big\}\, n\text{-}1 \text{ times}.$$

General upper bound

Lemma: (Cooper–Rorabaugh)

$f(n+1,q) \leq (f(n,q)+1)(q^{f(n,q)}+1)-1$

Proof: Given some word w, split it into $m = q^{f(n,q)}+1$ words w_i of length $f(n,q)$, which are separated by single letters:

$$w_1\, x\, w_2\, y \ldots z\, w_m$$

Then there are two identical words w_i and w_j, each containing the same copy of some Zimin word Z_n. This forms Z_{n+1}. □

Theorem:

$$f(n,q) \leq q^{\cdot^{\cdot^{\cdot^{q}}}} \Big\}\, n\text{-}1 \text{ times}\,.$$

Proof: Apply lemma recursively, starting with $f(3,q) \leq q^q$. □

Lower bounds?

Lemma: (Cooper–Rorabaugh)

$$f(n, q) \geq q^{2^{n-1}(1+o(1))},$$

where the $o(1)$ term depends on both q and n.

Lower bounds?

Lemma: (Cooper–Rorabaugh)

$$f(n, q) \geq q^{2^{n-1}(1+o(1))},$$

where the $o(1)$ term depends on both q and n.

Remark: Note that for any fixed n this bound is only constant degree polynomial in q versus upper bound which is tower-type!

Lower bounds?

Lemma: (Cooper–Rorabaugh)

$$f(n,q) \geq q^{2^{n-1}(1+o(1))},$$

where the $o(1)$ term depends on both q and n.

Remark: Note that for any fixed n this bound is only constant degree polynomial in q versus upper bound which is tower-type!

Question:

What is the maximum length of a word not containing the n-th Zimin word?

Tight lower bounds

Theorem: (Conlon–Fox–S.)

- For fixed n and large q,

$$f(n,q) \geq q^{\cdot^{\cdot^{\cdot^{q-o(q)}}}} \Big\} n\text{-1 times} .$$

Tight lower bounds

Theorem: (Conlon–Fox–S.)

- For fixed n and large q,

$$f(n,q) \geq q^{\cdot^{\cdot^{\cdot^{q-o(q)}}}} \Big\} n\text{-1 times}.$$

- $$f(n,2) \geq 2^{\cdot^{\cdot^{\cdot^{2}}}} \Big\} n\text{-4 times}.$$

Tight lower bounds

Theorem: (Conlon–Fox–S.)

- For fixed n and large q,

$$f(n,q) \geq q^{\cdot^{\cdot^{\cdot^{q-o(q)}}}} \Big\} n\text{-1 times} .$$

- $$f(n,2) \geq 2^{\cdot^{\cdot^{\cdot^{2}}}} \Big\} n\text{-4 times} .$$

- For $q \geq 5$,

$$0.75 \cdot 2^q \, q! \leq f(3,q) \leq 5 \cdot 2^q \, q! \,.$$

Tight lower bound for large alphabets

Theorem: (Conlon–Fox–S.)

For fixed n and large q,

$$f(n,q) \geq q^{\cdot^{\cdot^{\cdot^{q-o(q)}}}} \Big\} n\text{-1 times}.$$

Tight lower bound for large alphabets

Theorem: (Conlon–Fox–S.)

For fixed n and large q,

$$f(n,q) \geq q^{\cdot^{\cdot^{\cdot^{q-o(q)}}}} \Big\} n\text{-1 times}.$$

Ingredients of the proof:

- Analyze a certain random construction using Lovász Local Lemma to show that there are $q^{q^{q-o(q)}}$ words avoiding Z_3.

Tight lower bound for large alphabets

Theorem: (Conlon–Fox–S.)

For fixed n and large q,

$$f(n,q) \geq \left. q^{\cdot^{\cdot^{\cdot^{q-o(q)}}}} \right\} n\text{-1 times}.$$

Ingredients of the proof:

- Analyze a certain random construction using Lovász Local Lemma to show that there are $q^{q^{q-o(q)}}$ words avoiding Z_3.
- To go from n to $n+1$ develop an iterative step-up construction which gains an extra exponential at every step.

The step-up construction

Lemma:

Let $S(n, q)$ denote the set of all words w over an alphabet of size q which avoid Z_n and have a distinguished letter, say d, such that any subword of w not containing the letter d avoids Z_{n-1}. Then

$$|S(n+1, q+2)| \geq |S(n, q)|!$$

The step-up construction

Lemma:

Let $S(n, q)$ denote the set of all words w over an alphabet of size q which avoid Z_n and have a distinguished letter, say d, such that any subword of w not containing the letter d avoids Z_{n-1}. Then

$$|S(n+1, q+2)| \geq |S(n, q)|!$$

Proof: Let $m = |S(n, q)|$ and let $w_1, w_2, \ldots, w_m$ be one of the $m!$ orderings of the words in $S(n, q)$.

The step-up construction

Lemma:

Let $S(n,q)$ denote the set of all words w over an alphabet of size q which avoid Z_n and have a distinguished letter, say d, such that any subword of w not containing the letter d avoids Z_{n-1}. Then

$$|S(n+1,q+2)| \geq |S(n,q)|!$$

Proof: Let $m = |S(n,q)|$ and let $w_1, w_2, \ldots, w_m$ be one of the $m!$ orderings of the words in $S(n,q)$. If c is the distinguished letter in w_i for each i, let u_i be the word formed from w_i by replacing c with c_1 if i is odd and c_0 if i is even.

The step-up construction

Lemma:

Let $S(n, q)$ denote the set of all words w over an alphabet of size q which avoid Z_n and have a distinguished letter, say d, such that any subword of w not containing the letter d avoids Z_{n-1}. Then

$$|S(n+1, q+2)| \geq |S(n, q)|!$$

Proof: Let $m = |S(n, q)|$ and let $w_1, w_2, \ldots, w_m$ be one of the $m!$ orderings of the words in $S(n, q)$. If c is the distinguished letter in w_i for each i, let u_i be the word formed from w_i by replacing c with c_1 if i is odd and c_0 if i is even. Then consider the word

$$u_1 d u_2 d \ldots d u_m$$

where d is a new letter.

Lemma:

Let $S(n, q)$ denote the set of all words w over an alphabet of size q which avoid Z_n and have a distinguished letter, say d, such that any subword of w not containing the letter d avoids Z_{n-1}. Then

$$|S(n+1, q+2)| \geq |S(n, q)|!$$

Proof: Let $m = |S(n, q)|$ and let $w_1, w_2, \ldots, w_m$ be one of the $m!$ orderings of the words in $S(n, q)$. If c is the distinguished letter in w_i for each i, let u_i be the word formed from w_i by replacing c with c_1 if i is odd and c_0 if i is even. Then consider the word

$$u_1 d u_2 d \ldots d u_m$$

where d is a new letter. This word contains $q + 2$ letters and we claim that it is in $S(n+1, q+2)$.

An open problem

EXAMPLES:

- $P = xxx$ is 2-avoidable but not 1-avoidable.

An open problem

Examples:

- $P = xxx$ is 2-avoidable but not 1-avoidable.
- $P = xx$ is 3-avoidable but not 2-avoidable.

An open problem

Examples:

- $P = xxx$ is 2-avoidable but not 1-avoidable.
- $P = xx$ is 3-avoidable but not 2-avoidable.
- $P = xyayzbzxcyxdxz$ is 4-avoidable but not 3-avoidable. (Baker–McNulty–Taylor 1989)

An open problem

Examples:

- $P = xxx$ is 2-avoidable but not 1-avoidable.
- $P = xx$ is 3-avoidable but not 2-avoidable.
- $P = xyayzbzxcyxdxz$ is 4-avoidable but not 3-avoidable. (Baker–McNulty–Taylor 1989)
- $P = xyaxzbyxcyzdzwxewzw$ is 5-avoidable but not 4-avoidable. (Clark 2004)

An open problem

Examples:

- $P = xxx$ is 2-avoidable but not 1-avoidable.
- $P = xx$ is 3-avoidable but not 2-avoidable.
- $P = xyayzbzxcyxdxz$ is 4-avoidable but not 3-avoidable. (Baker–McNulty–Taylor 1989)
- $P = xyaxzbyxcyzdzwxewzw$ is 5-avoidable but not 4-avoidable. (Clark 2004)

Problem:

Do there exist patterns P which are q-unavoidable but $(q+1)$-avoidable for every integer $q \geq 1$?

An open problem

Examples:

- $P = xxx$ is 2-avoidable but not 1-avoidable.
- $P = xx$ is 3-avoidable but not 2-avoidable.
- $P = xyayzbzxcyxdxz$ is 4-avoidable but not 3-avoidable. (Baker–McNulty–Taylor 1989)
- $P = xyaxzbyxcyzdzwxewzw$ is 5-avoidable but not 4-avoidable. (Clark 2004)

Problem:

Do there exist patterns P which are q-unavoidable but $(q+1)$-avoidable for every integer $q \geq 1$?

Remark: No such words have been found so far for $q \geq 5$.


рахмат
danke
謝謝
ngiyabonga
teşekkür ederim
Баярлалаа
spas
mersi
barka
welalin
tack
спасибо
faafetai lava
vinaka
kia ora
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
спасибі
blagodaram
kiitos
dankie
thank you
gracias
хвала
nanni
nandri
dhanyavad
mauruuru
köszönöm
asante
manana
obrigada
murakoze
tenki
bedankt
bayarlalaa
gràcie
hvala
akun dankon
aciú
enkosi
djiere dieuf
tau
mochchakkeram
дякую
mamnun
dziękuję
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
go raibh maith agat
chokrane
sobodi
děkuji
sagolun
sukriya
kop khun krap
obrigado
mèsi
didi madloba
ありがとう
arigatô
takk
dakujem
trugarez
najis tuke
tanemirt
grazie
terima kasih
rahmet
kam sah hamnida
rahmat
diolch
dhanyavadagalu
shukriya
mercé
мерси
তোমাকে ধন্যবাদ
감사합니다
xièxie
ευχαριστώ
merci
תודה רבה