

AUTOMATA & FORMAL LANGUAGES

MONDAY 9 OCTOBER 2023

Lecture II

TODAY

SUMMER RESEARCH FESTIVAL

2 рш - 6 рш

MR 2, 3, 4, 5, 9,
8, 15

MR 3

5:20 pm

A motivated theorem prover

5:40 pm

5:40 pm
How to avoid
arithmetic
progressions



**UNIVERSITY OF
CAMBRIDGE**

Summer Research Festival 2023
Monday 9 October 2023, Centre for Mathematical Sciences



**UNIVERSITY OF
CAMBRIDGE**

Summer Research Festival 2023
Monday 9 October 2023, Centre for Mathematical Sciences

"Click on the title link to jump to a short summary"

Meeting Room 1	Meeting Room 2	Meeting Room 3
14.00 Learning Algebraic Varieties From Data <i>Terence Tao (Cambridge), John Bruna (Google)</i> 14.15 Determining Dynamic Brain Networks From Data with Changepoints <i>Alfred Aichwald (Cambridge), John Bruna</i> 14.30 Missing Data <i>David Joffe & Daniel Gusakovsky (Cambridge), Michael Barmann</i>	The dense region in isoperimetric diagrams <i>Patrick Del Moral (The University of Edinburgh)</i> Realizability of tropical curves <i>David Hestenes (Oxford), Benjamin Smith</i> Sampling Guarantees for High-dimensional Posterior Measures <i>David Joffe & Daniel Gusakovsky (Cambridge), Michael Barmann</i>	Hyperbolic metrics on algebraic varieties <i>Wing Yee Lee (Cambridge), Henry Wilton</i> The Geometry of Non-Positive 2-Complexes and Moduli of Surfaces <i>Lorenz Disinger (Cambridge), Henry Wilton</i> Strengthening tests in conformally non-positively curved 2-complexes <i>Patrick Del Moral (Cambridge), Henry Wilton</i>

"Click on the title link to jump to a short summary"

Meeting Room 1	Meeting Room 2	Meeting Room 3
14.00 Learning Algebraic Varieties From Data <i>Terence Tao (Cambridge), John Bruna (Google)</i> 14.15 Determining Dynamic Brain Networks From Data with Changepoints <i>Alfred Aichwald (Cambridge), John Bruna</i> 14.30 Missing Data <i>David Joffe & Daniel Gusakovsky (Cambridge), Michael Barmann</i>	The dense region in isoperimetric diagrams <i>Patrick Del Moral (The University of Edinburgh)</i> Realizability of tropical curves <i>David Hestenes (Oxford), Benjamin Smith</i> Sampling Guarantees for High-dimensional Posterior Measures <i>David Joffe & Daniel Gusakovsky (Cambridge), Michael Barmann</i>	Hyperbolic metrics on algebraic varieties <i>Wing Yee Lee (Cambridge), Henry Wilton</i> The Geometry of Non-Positive 2-Complexes and Moduli of Surfaces <i>Lorenz Disinger (Cambridge), Henry Wilton</i> Strengthening tests in conformally non-positively curved 2-complexes <i>Patrick Del Moral (Cambridge), Henry Wilton</i>

"Click on the title link to jump to a short summary"

Meeting Room 1	Meeting Room 2	Meeting Room 3
14.00 Polynomials with many preperiodic points <i>John Bruna (Cambridge), John Bruna</i> 14.15 Constructing algebraic varieties <i>David Hestenes (Oxford), Benjamin Smith</i> 14.30 Decomposing q-adic Representations as q-adic Representations <i>John Bruna (Cambridge), John Bruna</i>	Reconstructing Point Sets from Random Sets of Distances <i>David Hestenes (Oxford), Benjamin Smith</i> The densest subgraph and q-adic <i>David Hestenes (Oxford), Benjamin Smith</i>	Extremal Points of the Primary Polygons <i>David Hestenes (Oxford), Benjamin Smith</i> Extremal Points of the Primary Polygons <i>David Hestenes (Oxford), Benjamin Smith</i>

"Click on the title link to jump to a short summary"

Meeting Room 1	Meeting Room 2	Meeting Room 3
14.00 On the generalized Turán problem for odd cycles <i>Terence Tao (Cambridge), John Bruna</i> 14.15 A motivated theorem prover <i>David Hestenes (Oxford), Benjamin Smith</i> 14.30 How to avoid arithmetic progressions <i>David Hestenes (Oxford), Benjamin Smith</i>	Random point sets from Random Sets of Distances <i>David Hestenes (Oxford), Benjamin Smith</i> Extremal Points of the Primary Polygons <i>David Hestenes (Oxford), Benjamin Smith</i>	Extremal Points of the Primary Polygons <i>David Hestenes (Oxford), Benjamin Smith</i> Extremal Points of the Primary Polygons <i>David Hestenes (Oxford), Benjamin Smith</i>



Meeting Room 3, 5:00pm - 6:00pm

5:00pm

Csongor Beke

On the generalized Turán problem for odd cycles

• Combinatorics
 • Combinatorics

The generalized Turán problem is concerned with estimating, for graphs T and H and a positive integer n , the maximum number of copies of T that can be found in a graph with n vertices. In this graph can have n vertices. I will talk about the new results we found during the project, in the case of T and H being odd cycles, n smaller than T .

5:30pm

Mantas Baksys & Jovan Gerbscheid

A motivated theorem prover

• Combinatorics
 • Combinatorics
 • Automated and interactive theorem proving

Most mathematical theorems are great at connecting the reader from its results are true but do not show how to come up with proofs. The ability to motivate proofs is crucial for both mathematical research and the eventual dream of human-oriented automated theorem proving. During our project, we built an interactive program that enables the user to find motivated proofs.

5:40am

Yael Dillies

How to avoid arithmetic progressions

• Combinatorics
 • Combinatorics
 • CMF / External
 • Dept of Computer Science & Technology

How many integers between 1 and n can we have without creating an arithmetic progression of length three? I will present the complicated answer to this seemingly innocuous question, and how I am convincing a computer that this answer is correct.

Brief reminder of Lecture I:

Pages 9 & 10.

NOTATION & REMINDER

(1) $\mathbb{N} = \{0, 1, 2, \dots\}$

$n = \{0, \dots, n-1\}$

(2) For any set X , we write

X^n for the sequence / tuples / strings of length n of elements of X

Notation

X^n : X -strings of length n

(3) In particular, X^0 consists only of the empty string. We call this ϵ (EPSILON).

(4) Note that if $\alpha \in X^n$, α is a function with $\text{dom}(\alpha) = n$ and $\text{ran}(\alpha) \subseteq X$.

ϵ : empty string

(5) If $\alpha \in X^n$ and $k \leq n$, then $\alpha \upharpoonright k \in X^k$ is the unique initial segment of α of length k .

$|\alpha|$: length of α

(6) $|\alpha| := \text{dom}(\alpha)$ length of α .

X^* : finite X -strings

(7) $X^* := \bigcup_{n \in \mathbb{N}} X^n$ The set of finite X -sequences

$\alpha\beta$: concatenation of α and β .

(8) Concatenation:

$\alpha \in X^n, \beta \in X^m$ we can define $\alpha\beta \in X^{n+m}$ by

$$\alpha\beta(k) := \begin{cases} \alpha(k) & k < n \\ \beta(k-n) & k = n+l \text{ with } l < m \end{cases}$$

(9) Slightly incorrectly, I'll write x for the length one seq. with value x .

This means that αx is just

$$\underbrace{a_0 \dots a_n}_{=\alpha} x$$

NOTATION & REMINDERS (continued)

(10) Define recursively

$$\alpha^0 := \varepsilon$$

$$\alpha^{n+1} := \alpha^n \alpha$$

(11) If $\alpha = x$, then x^n is the length n sequence consisting of x 's.

(12) $X^+ := X^* \setminus \{\varepsilon\}$

set of non-empty finite X -strings

(13) X is **infinite** if there is an inj. $\mathbb{N} \rightarrow X$
 X is **countable** if there is a surj. $\mathbb{N} \rightarrow X$
or $X = \emptyset$

X is **finite** if not infinite

X is **uncountable** if not countable

(14) X, Y ctable $\Rightarrow X \times Y$ is countable

[Remember Cantor zigzag function

$$z: \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$$

$$\text{bij.}$$

Fix u , find i, j

if $\pi_X: \mathbb{N} \rightarrow X$

$\pi_Y: \mathbb{N} \rightarrow Y$

surjections

s.t. $z(i, j) = u$

$$u \mapsto (\pi_X(i), \pi_Y(j)).$$

(15) X is countable
 $\Rightarrow X^*$ is countable

[Since X^{n+1} is in bij. with $X^n \times X$, it follows by ind. from (14) that all sets X^n are countable.

But $X^* = \bigcup_{n \in \mathbb{N}} X^n$, so countable as a countable union of countable sets. [N&S].]

(16) If $X \neq \emptyset$, then X^* is infinite.

[If $x \in X$, then $n \mapsto x^n$ is an inj. from $\mathbb{N}$ to X^* .]

Sequence of length n
with constant value x .

(17) Cantor's Thm.

If X is infinite, then
 $\mathcal{P}(X)$ power set of X
 $\equiv$ set of all subsets of X
 is uncountable.

[Very famous proof:
DIAGONALISATION

Start with any function

$$f: \mathbb{N} \rightarrow \mathcal{P}(X)$$

and define $D \subseteq X$ s.t.

$D \neq \text{ran}(f)$.] *check typed lecture notes for the proof.*

(18) If X is countable, then so is the
 set of finite subsets of X , say $\text{Fin}(X)$.

[Know by (15) that X^* is countable, so
 only need a surj. from X^* onto $\text{Fin}(X)$.

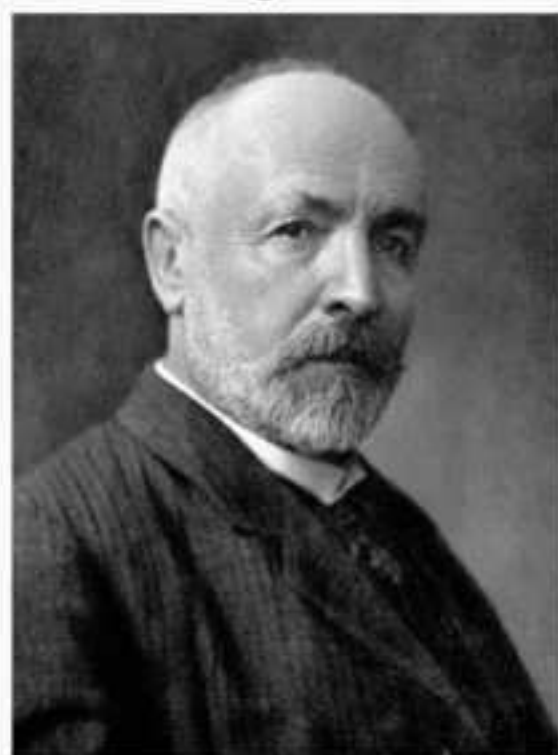
If $\alpha \in X^*$, let $\pi(\alpha) := \text{ran}(\alpha)$.

That is a surjection.]

*Some subtlety: to show "surjection", need for arbitrary $F \in \text{Fin}(X)$
 some α s.t. $\pi(\alpha) = F$. How do you define α ?*

[See typed lecture notes.]

Georg Cantor



Cantor, c. 1910

Born	Georg Ferdinand Ludwig Philipp Cantor
	3 March 1845
	Saint Petersburg , Russian Empire
Died	6 January 1918 (aged 72)
	Halle , Province of Saxony, German Empire
Nationality	German

§ 1.2 Formal Languages & Rewrite Systems

Ω finite set of symbols

We call any $L \subseteq \Omega^*$ a
(formal) Ω -language

Compare to natural languages:

$$\Omega = \{a, b, c, d, \dots, z\}$$

then words become elements of Ω^* and
a dictionary is a subset of Ω^* , thus
a language.

Clearly, that's finite.

[There is little mathematical theory for finite languages.
So?]

FORMAL LANGUAGES & REWRITE SYSTEMS

Chomsky:

A fundamental principle of language is

LINGUISTIC RECURSION
GENERATION

Languages are best thought of as infinitely generated, even though only a finite fragment of these is practically relevant.

Noam Chomsky



Chomsky in 2017

Born	Avram Noam Chomsky December 7, 1928 (age 94) Philadelphia , Pennsylvania, U.S.
Spouses	Carol Schatz (m. 1949; died 2008) Valeria Wasserman (m. 2014)
Children	3, including Aviva
Parent	William Chomsky (father)

E.g., the process of taking subordinate clauses is a *productive* feature of language. In principle, no matter how complex a sentence is, one can increase its complexity by prefixing it with "X observes that". So, we form an infinite sequence of grammatical sentences

B likes *A*.

C believes that *B* likes *A*.

D reports that *C* believes that *B* likes *A*.

E observes that *D* reports that *C* believes that *B* likes *A*.

etc.

Chomsky: The most salient feature of language is its recursive / productive nature which necessarily implies that languages should be conceived of as recursively generated & thus infinite.



GRAMMATICAL

COLOURLESS GREEN IDEAS SLEEP FURIOUSLY

FURIOUSLY SLEEP IDEAS GREEN COLOURLESS

UNGRAMMATICAL

Both sentences are nonsensical and have no meaning, but we can still clearly say that the first one is grammatical and the second one not.

If Ω is our set of symbols, we call elements of

$$\Omega^+ \times \Omega^*$$

REWRITE RULES
PRODUCTION RULES

We write

$$\alpha \longrightarrow \beta$$

for the rule
(α, β).

INFORMAL INTERPRETATION :

"Whenever a string contains α as a substring, we can replace it with β ."

Definition A pair $R = (\Omega, P)$ is called a **REWRITE SYSTEM** if P is a finite set of rewrite rules.

Prop. For any fixed Ω , there are countably many rewrite systems (Ω, P) .

Proof. By (15), Ω^* and Ω^+ are ctbl.
So by (14), $\Omega^+ \times \Omega^*$ is ctbl.

There is 1-1 correspondence between rewrite systems and $\text{Fin}(\Omega^+ \times \Omega^*)$ which is ctbl by (18). q.e.d.

NOTATION

If $R = (\Omega, P)$ is a rewrite system, $\sigma, \tau \in \Omega^*$, then we write

$$\sigma \xrightarrow{R} \tau$$

R rewrites σ in one step as τ

$\iff$

$$\exists \alpha, \beta, \gamma, \delta \in \Omega^* \text{ s.t.}$$

$$\sigma = \alpha\beta\gamma,$$

$$\tau = \alpha\delta\gamma,$$

$$\beta \rightarrow \delta \in P.$$

We let the relation $\xrightarrow{R}$ be the transitive and reflexive closure of $\xrightarrow{R}$.

This means $\sigma \xrightarrow{R} \tau$ iff

①

$$\sigma = \tau$$

②

there is a sequence of strings

$$\sigma_0, \dots, \sigma_n \text{ s.t.}$$

$$\sigma = \sigma_0, \tau = \sigma_n \text{ and } \sigma_k \xrightarrow{R} \sigma_{k+1}.$$

Note that a derivation of length n is a sequence of length $n+1$.

an R -derivation of τ from σ

This is called derivation of length n .

R derives τ from σ .

R rewrites σ as τ

R produces τ from σ

$$D(R, \alpha) :=$$

$$\{\beta; \alpha \xrightarrow{R} \beta\}$$

GRAMMARS

Def. $G = (\Sigma, V, P, S)$ is called a
(formal) grammar over Σ

if

1. $\Sigma \cap V = \emptyset$

2. $\Omega := \Sigma \cup V$ is finite nonempty set of symbols

terminal
symbols
or
letters

nonterminal
symbols
or
variables

Σ is called the alphabet

3. (Ω, P) is a rewrite system

4. $S \in V$ called the start symbol

We call $W := \Sigma^*$ the set of WORDS.

The definitions for rewrite systems still make sense:

$$\alpha \xrightarrow{G} \beta, \mathcal{D}(G, \alpha)$$

$$\alpha \xrightarrow{G} \beta$$

$$L(G) := W \cap \mathcal{D}(G, S)$$

LANGUAGE generated by G .

Example

$$G_0 = (\Sigma, V, P_0, S)$$

$$\Sigma := \{a\}$$

$$V := \{S\}$$

$$P_0 := \{S \rightarrow aaS, S \rightarrow a\}$$

Claim

$$L(G_0) = \{a^{2u+1}; u \in \mathbb{N}\}$$

Easy to see that $\supseteq$;
to see $\subseteq$, observe that every derivable string has odd length.

[Details will be repeated at the beginning of Lecture II.]