Lecture 7

Last lecture:

- We sketched a proof of difficult direction of Theorem 1 (Lecture 4).
  For a detailed proof, consult Libkin's book.

In this lecture:

- We discuss the density function of regular languages with examples.

- We formulate and proved Schützenberger's Theorem, stating that regular languages have density functions which satisfy a linear recurrence relation over $\mathbb{Z}$.

- We start our discussion of linear recurrence relations for certain density function.

Linear recurrence relations for density functions of regular languages: Schützenberger's Theorem

Recurrence formulas for density functions.
A case study: Regular languages.

Let L be a language, i.e. a set of words over an alphabet Σ. We will use $\Sigma = \{a, b\}$. L^n denotes the set of words of length exactly n which are in L.

We want to count the number of words $a_L(n) =| L^n |$ in L^n.

Examples:

- $a^* b^*$
- $aab^* a$
- $a^n b^n$
- The set of palindroms

Use auxiliary density functions.

Assume we have density functions

$$\bar{a}(n) = (a_1(n), \ldots, a_m(n))^{tr}$$

and an $(m \times m)$-matrix M such that

$$\bar{a}(n+1) = M\bar{a}(n)$$

We would like to transform this into a recurrence relation for one of the density functions.

Caley's Theorem (for matrices)

The **characteristic polynomial** of an $(m \times m)$-matrix M is

$$\chi_M(\lambda) = \sum_i^m c_i(M)x^i = \sum_i^m c_i x^i = \mathrm{det}(\lambda \cdot \mathbf{1} - M)$$

Caley's Theorem states that, that in the ring of matrices

$$\chi_M(M) = \sum_i^m c_i M^i = 0$$

Hence we have, using that $M^i \bar{a}(n) = \bar{a}(n+i)$,

$$c_m(M)M^m\bar{a}(n) = c_m\bar{a}(n+m) = -\sum_i^{m-1} c_i M^i \bar{a}(n) = -\sum_i^{m-1} c_i \bar{a}(n+i)$$

which is the required recurrence relation for each density function a_k.

How to find the auxiliary density functions?

We can use

- **Pumping Lemma** for regular or context-free languages.
- **Ehrenfeucht-Fraïssé Theorem** for Monadic Second Order Logic
- **Myhill-Nerode Theorem** on congruences closed under concatenation.
- **Büchi's Theorem** identifying regular languages with Mondaic Second Order definable languages.

In the case of words, all these theorems are **inherently related**.

In the case of arbitrary structures the underlying property is the **Feferman-Vaught Theorem for Monadic Second Order Logic**.

Schützenberger's Theorem

Theorem 1 (Schützenberger, 1961)
For every regular language L, the density function $a_L(n)$ satisfies a linear recurrence relation.

Note: There are non-regular languages which also satisfy a linear recurrence relation.

Sketch of proof of Theorem 1.

- Given L, find a non-deterministic automaton A_L with states S, with exactly one accepting state s_L, and with transition table δ_A, which accepts L.
- For each $s \in S$ define A_s to be the automaton with the same states and transition functions as A_L, but with sole accepting state s.
- Use as auxiliary languages the languages $L_s = L(A_s)$ with density functions $a_s(n)$.
- For the transition table δ_A, let $m_{s,t}$ denote the number of transitions from state s to state t.
- Define the $S \times S$-matrix M over non-negative integers with entries $m_{s,t}$.
- Now compute the characteristic polynomial of M to get the linear recurrence relations between the density functions of all the L_s.

Sketch of proof of Theorem 1, continued.

Claim: With $m = |S|$ and $S = \{s_1, \ldots, s_m\}$, and

$$\bar{a}(n) = (a_{s_1}(n), \ldots, a_{s_m}(n))$$

we have

$$\bar{a}(n+1) = M \cdot \bar{a}(n)^{tr} = M \cdot (a_{s_1}(n), \ldots, a_{s_m}(n))^{tr}.$$

Complete the proof !

Properties of density functions.

Given a density function $d(n)$ of

- a regular language,
- a context-free language,
- a herediatry graph property of labeled graphs,
- a monotone graph property of labeled graphs,
- a FOL-definable graph property of labeled graphs,
- a $MSOL$-definable (SOL-definable) graph property of labeled graphs.

What can we say about $d(n)$?

Conversely, given a function $d(n)$, under what conditions is $d(n)$ a density function of any of the above?

From linear recurrences to regular languages

If $d_L(n)$ is a density function of some regular language L over an alphabet Σ, then

- All the values of $d(n)_L$ are non-negative.
- $d(n)_L$ is bounded by $\mid \Sigma \mid^n$.
- $d(n)_L$ satisfies a linear recurrence relation.
- The generating function $f_L(x) = \sum_n d(n)x^n$ is a rational function.

Is every function satisfying the above the density function of some regular language?

Project: The answer is rather complex.

E. Barcucci, A. Del Lungo, A. Frosini and S. Rinaldi, From rational functions to regular languages, in *Formal Power Series and Algebraic Combinatorics*, D. Krob, A.A. Mikhalev and A.V. Mikhalev eds., Springer, 2000, pp. 633-644.

Density functions of graph classes

Let P be a graph property, i.e. a class of graphs closed under isomorphisms, and let $d_P(n)$ be its density function for labeled graphs.

- If $P = Graphs$ consists of all simple graphs,

$$d_{Graphs}(n) = 2^{\binom{n}{2}}$$

  In the unlabeled case the function is rather complicated.

- If $P = LinOrd$ consists of all linear orders.

$$d_{LinOrd}(n) = n!$$

  In the unlabeled case we have the constant function with value 1.

- If $P = SqGrids$ consists of all square grids,

$$d_{SqGrids}(n) = \begin{cases} \frac{n!}{4} & \text{if } n = m^2 \\ 0 & \text{else} \end{cases}$$

  In the unlabeled case we have 1 instead of $n!$.

Growth arguments

Lemma 2
Let $f : \mathbb{Z} \to \mathbb{Z}$ *a function which satisfies a linear recurrence relation* $f(n+1) = \sum_{i=0}^{k} a_i f(n-i)$ *over* $\mathbb{Z}$ *with* $a_{max} = \max_i a_i$.
Then there is a constant $c \in \mathbb{Z}$ *such that* $f(n) \leq 2^{cn}$.

Sketch of Proof:
One can prove this directly by induction with $c = log_2(k \cdot a_{max})$.
Q.E.D.

Corollary 3
For $\mathcal{C} \in \{Graphs, LinOrd, SqGrids\}$,
$f_{\mathcal{C}}(n)$ *does not satisfy a linear recurrence over* $\mathbb{Z}$.

A better estimate of the growth in Lemma 2.

To get a better estimate for c, one uses the spectral radius of the $(k \times k)$-matrix A associated to the recurrence by

$$\begin{pmatrix} f(n+1) \\ \vdots \\ f(n+k)) \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 & \dots & 0 & 0 & a_1 \\ 1 & 0 & \dots & 0 & 0 & a_2 \\ 0 & 1 & \dots & 0 & 0 & a_3 \\ & & \vdots & & & \vdots \\ 0 & 0 & \dots & 0 & 1 & a_k \end{pmatrix}}_{A} \cdot \begin{pmatrix} f(n) \\ \vdots \\ f(n+k-1) \end{pmatrix}$$

By a classical theorem, the sequence of matrices A^n converges iff $\rho(A) < 0$, where $\rho(A)$ denotes the the spectral radius of A, which is the maximum of all the absolute values the eigenvalues of A. Hence, the eigenvalues of A determine the growth rate of the sequence, and from the largest absolute value one can estimate c in the lemma. Q.E.D.

Modular linear recurrences, I

However we note:

- For every $m \in \mathbb{N}$ and for large enough n we have $n! = 0 \pmod{m}$
  Hence, for $n \geq N(m)$ we have
  $$d_{LinOrd}(n+1) = d_{LinOrd}(n) \pmod{m}$$
  and
  $$d_{SqGrid}(n+1) = d_{SqGrid}(n) \pmod{m}$$

We say that a function $f(n)$ satisfies a **trivial modular recurrence**
if for every m there exists N_m such that
if $n > N_m$ then $g(n) \equiv 0 \pmod{m}$.
This is true in particular, and even equivalent to,
if there exist functions $g(n), h(n)$ with $g(n)$ tending to infinity
such that $f(n) = g(n)! \cdot h(n)$.

Clearly, the two examples above are trivial modular recurrences.

Modular linear recurrences, II

Now we look at

$$d_{Graphs}(n+1) = 2^{\binom{n+1}{2}} = 2^{\binom{n}{2}} \cdot 2^n$$

Hence

$$d_{Graphs}(n+m+1) = d_{Graphs}(n) \cdot \prod_{i=0}^{m} 2^{n+i} = d_{Graphs}(n) \cdot 2^{nm} \cdot \prod_{i=0}^{m} 2^i$$

As $nm = 0 \pmod{m}$ we get

$$d_{Graphs}(n+m+1) = d_{Graphs}(n) \cdot \prod_{i=0}^{m} 2^i \pmod{m}$$

This is a non-trivial recurrence.

It is also different for distinct m and m', in other words, non-uniform in m.

Two equal-sized cliques, I

Let $EQ_2CLIQUE$ the class of graphs which consists of two disjoint unions of equal-sized cliques.

We want to study its density function $d_{EQ_2CLIQUE}(n)$. We have

$$d_{EQ_2CLIQUE}(n) = b_2(n) = \begin{cases} \frac{1}{2}\binom{2m}{m} & \text{for } n = 2m \\ 0 & \text{else} \end{cases}$$

The factor $\frac{1}{2}$ is there because we cannot distinguish the choice of the first clique from the choice of its complement.

Proposition 4 (Lucas, 1878)
For every n which is not a power of 2, we have $b_2(n) \equiv 0 \pmod 2$, and for every n which is a power of 2 we have $b_2(n) \equiv 1 \pmod 2$.
In particular, $b_2(n)$ is not ultimately periodic modulo 2.

A proof may be found as Exercise 5.61 in:
R. Graham, D. Knuth and O. Patashnik,
Concrete Mathematics, 2nd ed., Addison-Wesley 1994

There is a generalization of this for p-many equal-sized cliques.

Two equal-sized cliques, II

- We can prove (using the various version of the pebble games) that $EQ_2CLIQUE$ is not definable in FOL, or $\mathcal{L}^{\omega}_{\infty,\omega}$.
- One can also prove (using the $MSOL$-version of the pebble games) that $EQ_2CLIQUE$ is not definable in Monadic Second Order Logic $MSOL$.
- However, $EQ_2CLIQUE$ is definable in Second Order Logic SOL.
- With **two** binary E, M relations we can express in FOL that E is the edge relation of a graph in $EQ_2CLIQUE$, and that M is a matching (bijection) between the two cliques.
  Let us call the class so defined $M_2CLIQUES$.

  But for the density function of $M_2CLIQUES$ we have

$$d_{M_2CLIQUES}(2n) = n! \cdot d_{EQ_2CLIQUES}(2n)$$

  which satisfies the trivial modular recurrence relations.