State Complexity and Jacobsthal's Function

Jeffrey Shallit
Department of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada
shallit@graceland.uwaterloo.ca
http://www.math.uwaterloo.ca/~shallit

State Complexity

The *state complexity* of a regular language L, $\mathrm{sc}(L)$, is the minimum number of states needed to accept it by a DFA.

The problem:
Given languages L, L' with state complexity n, n' respectively, what are good bounds on the state complexity of $L \cup L'$, LL', L^*, etc.?

For the state complexity of intersection, we have the following bound:

Proposition. *We have*

$$\mathrm{sc}(L \cap L') \leq \mathrm{sc}(L)\mathrm{sc}(L').$$

Proof. Use the usual direct product construction.

State Complexity of Intersection

The upper bound of $\mathrm{sc}(L)\mathrm{sc}(L')$ can be achieved if L, L' are over an alphabet of size at least 2:

Proposition. (S. Yu.) *Define*

$$L := \{x \in \{a,b\}^* \ : \ |x|_a \equiv 0 \pmod{n}\};$$

$$L' := \{y \in \{a,b\}^* \ : \ |y|_b \equiv 0 \pmod{n'}\}.$$

Then

$$\mathrm{sc}(L \cap L') = nn'.$$

But what if L, L' are *unary*, that is, defined over an alphabet of one symbol?

Clearly if $\gcd(n, n') = 1$ then the bound nn' can again be achieved, by taking $L = (a^n)^*$ and $L' = (a^{n'})^*$.

But what if $\gcd(n, n') > 1$?

Note: Nicaud [1999] has recently investigated the *average* state complexity for various operations on unary languages, including intersection.

State Complexity of Intersection for Unary Languages

A connected unary DFA has the property that its transition diagram consists of

- a *tail* of $t \geq 0$ states and
- a *cycle* of $c \geq 1$ states.

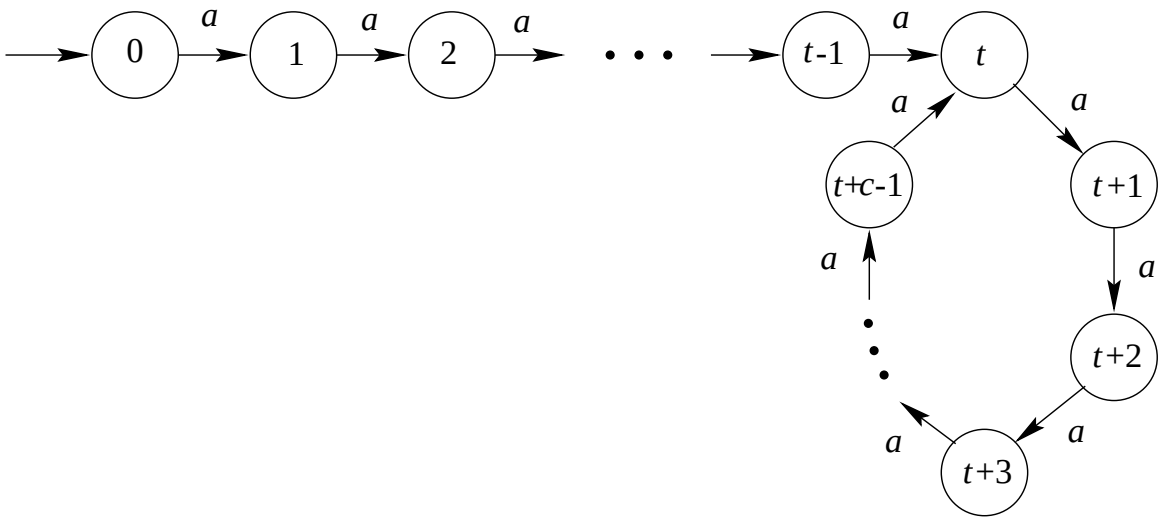

Figure 1: Transition diagram of M (accepting states not shown)

Then it is not difficult to prove the following

Theorem. *Let M, M' be unary DFA's with tails of size t, t' and cycles of size c, c', respectively. If L, L' are the corresponding languages, we have*

$$\mathrm{sc}(L \cap L') \leq \max(t, t') + \mathrm{lcm}(c, c'). \qquad (1)$$

Furthermore, for all $t, t' \geq 0$ and $c, c' \geq 1$ there exist unary languages for which the bound (1) is achieved.

For example, if $t \geq t'$, take

$$\begin{aligned} L &= a^{t+c-1}(a^c)^*; \\ L' &= a^r(a^{c'})^*; \\ r &= t - 1 \bmod c'. \end{aligned}$$

Note: this theorem was obtained independently by PIGHIZZINI [2000].

Two New Number-Theoretic Functions

Thus, to estimate the worst-case behavior for the state complexity of intersection of unary languages with n and n' states, respectively, we must estimate the function

$$F(n,n') = \max_{\substack{1\le c\le n\\ 1\le c'\le n'}} \left(\max(n-c, n'-c') + \mathrm{lcm}(c,c') \right).$$

This in turn suggests studying the somewhat simpler and more natural function

$$G(n,n') = \max_{\substack{1\le c\le n\\ 1\le c'\le n'}} \mathrm{lcm}(c,c').$$

1	2	3	4	5	6	7	8	9	10	11	12	13
2	2	6	6	10	10	14	14	18	18	22	22	26
3	6	6	12	15	15	21	24	24	30	33	33	39
4	6	12	12	20	20	28	28	36	36	44	44	52
5	10	15	20	20	30	35	40	45	45	55	60	65
6	10	15	20	30	30	42	42	45	45	66	66	78
7	14	21	28	35	42	42	56	63	70	77	84	91
8	14	24	28	40	42	56	56	72	72	88	88	104
9	18	24	36	45	45	63	72	72	90	99	99	117
10	18	30	36	45	45	70	72	90	90	110	110	130
11	22	33	44	55	66	77	88	99	110	110	132	143
12	22	33	44	60	66	84	88	99	110	132	132	156
13	26	39	52	65	78	91	104	117	130	143	156	156

Two New Number-Theoretic Functions

- The asymptotic behavior of F and G is still not known precisely
- There is a relation to JACOBSTHAL'S function $g(n)$, which is the *least integer r such that every set of r consecutive integers contains at least one integer relatively prime to n.*
- IWANIEC proved [1978] using the linear sieve that $g(n) = O((\log n)^2)$.
- It then follows that if $n \leq n'$, we have

$$F(n, n') \geq G(n, n') \geq nn' - c_1(\log n)^2 n$$

for some constant c_1.

Visibility of Lattice Points

Our problem is closely related to another one about visibility of lattice points.

Consider the set $\mathcal{D}_n = \{(i,j) \ : \ 0 \leq i,j < n\}$, an $n \times n$ square array of lattice points in the plane.

We say a point $P \in \mathcal{D}_n$ is visible from a point $Q \in \mathcal{D}_n$ if $P = Q$ or if there is no lattice point of $\mathcal{D}_n$ between P and Q on the line joining P and Q.

It is easy to see that $P = (a,b)$ is visible from $Q = (c,d)$ iff $\gcd(a-c, b-d) \leq 1$.

Now let $\mathcal{A}, \mathcal{B}$ be subsets of D_n. We say $\mathcal{A}$ is visible from $\mathcal{B}$ if, for each $P \in \mathcal{A}$ there exists $Q \in \mathcal{B}$ such that P is visible from Q.

Visibility of Lattice Points

ABBOTT [1974] raised the following problem: what is the cardinality $f(n)$ of the smallest subset $\mathcal{S}$ of $\mathcal{D}_n$ such that $\mathcal{D}_n$ is visible from $\mathcal{S}$?

He proved that $f(2) = 2$ (see Figure 2) and

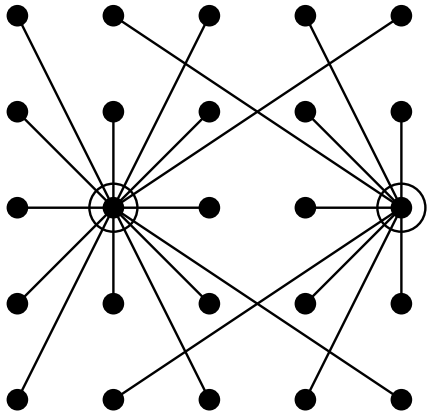

Figure 2: Proving that $f(2) \leq 2$

$$\frac{\log n}{2 \log \log n} < f(n) < 4 \log n$$

for all n sufficiently large.

Visibility of Lattice Points

Adhikari and Balasubramanian [1996] proved that there exist constants c_6 and c_7 such that if we take

$$S = \{(x, y) \ : \quad 1 \leq x \leq c_6 \log\log\log n, \\ 1 \leq y \leq c_7 (\log n)/(\log\log n)\}.$$

then $\mathcal{D}_n$ is visible from S. It follows that

$$f(n) \leq \frac{c_6 c_7 (\log n)(\log\log\log n)}{\log\log n}.$$

Visibility of Lattice Points

We may use the ADHIKARI-BALASUBRAMANIAN results to improve our lower bound in some cases:

Theorem.

(a) If

$$n \leq n' \leq \frac{n \log n}{(\log \log n)(\log \log \log n)},$$

then there exists a constant c_2 such that

$$F(n, n') \geq G(n, n') \geq nn' - c_2 \frac{\log n}{\log \log n} n.$$

(b) If

$$\frac{n \log n}{(\log \log n)(\log \log \log n)} \leq n' \leq \frac{n(\log n)^2}{\log \log \log n}$$

then there exists a constant c_3 such that

$$F(n, n') \geq G(n, n') \geq nn' - c_3(\log \log \log n)n'.$$

An Upper Bound on F

We will now prove an upper bound on F.

Theorem. There exist a constant c_4 and infinitely many distinct pairs n, n' with $n' < n$ such that

$$G(n,n') \leq F(n,n') \leq nn' - c_4\sqrt{\frac{\log n}{\log\log n}}n.$$

First we prove the following lemma.

Lemma. Let n, n' be fixed positive integers. The quantity

$$Q(c,c') := \max(n-c, n'-c') + \mathrm{lcm}(c,c')$$

is maximized ($1 \leq c \leq n$, $1 \leq c' \leq n'$) only if $\gcd(c,c') = 1$.

Proof. Assume not. Then $Q(c,c')$ is maximized for some c, c' with $\gcd(c,c') = g > 1$. Assume without loss of generality that $n \geq n'$. For $n < 11$ the theorem can be verified by a simple computer program. Hence assume $n \geq 11$.

An Upper Bound on F

We have $\max(n-c, n'-c') < n$ and $\operatorname{lcm}(c,c') = \frac{cc'}{g} \leq \frac{n^2}{2}$, so $Q(c,c') < n + \frac{n^2}{2}$.

KANOLD proved that $g(n) < 2^{\omega(n)}$. Hence there exists a k, $1 \leq k \leq 2^{\omega(n)}$ such that $\gcd(n, n-k) = 1$. Since $Q(c,c')$ is a maximum, we have $Q(c,c') \geq n(n-k)+k > n(n-2^{\omega(n)})$. Putting the inequalities for Q together, we get

$$n(n - 2^{\omega(n)}) < n + \frac{n^2}{2},$$

and so $n - 2^{\omega(n)} < 1 + \frac{n}{2}$. Thus $n < 2(2^{\omega(n)} + 1)$.

However, we claim that $n > 2(2^{\omega(n)} + 1)$ for $n \geq 11$. For $11 \leq n \leq 141$ this follows by an explicit calculation. Otherwise $n \geq 142$. We now use a theorem of ROBIN which states $\omega(n) \leq t(n)$ where

$$t(n) := \frac{\log n}{\log\log n} + 1.45743 \frac{\log n}{(\log\log n)^2}.$$

Since $n \geq 142$, we have $\log\log n > 1.6$ and so

$$t(n) < \frac{\log_2 n}{1.6 \log_2 e} + 1.45743 \frac{\log_2 n}{2.56 \log_2 e} < .83 \log_2 n.$$

We thus obtain

$$2(2^{\omega(n)+1}) < 2(2^{t(n)} + 1) < 2(n^{.83} + 1)$$

and it is easily verified that $2(n^{.83}+1) < n$ for $n \geq 70$. This contradiction completes the proof. ∎

Remark. Note that

$$F(n,n') = \max_{\substack{1\leq c\leq n \\ 1\leq c'\leq n'}} (\max(n-c, n'-c') + \mathrm{lcm}(c,c'))$$

does not necessarily achieve its maximum at the same pair (c,c') which maximizes

$$G(n,n') = \max_{\substack{1\leq c\leq n \\ 1\leq c'\leq n'}} \mathrm{lcm}(c,c').$$

For example, $F(148,30) = 4295$, which is uniquely achieved at $(c,c') = (143,30)$, while $G(148,30) = 4292$, which is uniquely achieved at $(c,c') = (148,29)$.

We can now prove our upper bound.

An Upper Bound on F

Theorem. There exist a constant c_4 and infinitely many distinct pairs n, n' with $n' < n$ such that $G(n, n') \leq F(n, n') \leq nn' - c_4\sqrt{\frac{\log n}{\log\log n}}n$.

Proof. Let $d \geq 1$ be a fixed integer. Let $S_d = \{(i, j) : i, j \geq 0 \text{ and } i + j < d\}$. For each pair $(i, j) \in S_d$, choose a distinct prime $q_{i,j}$ from the set $\{p_1, p_2, \ldots, p_v\}$, where p_i denotes the i'th prime and $v = d(d+1)/2$. By the Chinese remainder theorem, we can find n, n' such that $q_{i,j} \mid n - i$ and $q_{i,j} \mid n' - j$ for all pairs $(i, j) \in S_d$. Furthermore, we may choose n and n' such that $K \leq n' < 2K$, $2K \leq n < 3K$, where $K := \prod_{1 \leq i \leq v} p_i$. By the prime number theorem we have $K = e^{(1+o(1))v\log v}$. Hence there exists a constant c_5 such that $d \geq c_5\sqrt{\frac{\log n}{\log\log n}}$.

It follows that $\gcd(n-i, n'-j) > 1$ for all pairs $(i, j) \in S_d$. By a previous Lemma, we know that F cannot achieve its maximum when $(c, c') \in S_d$.

It follows that $F(n, n') \leq \max_{b+c=d}((n-b)(n'-c)+d)$. But

$$\max_{b+c=d}((n-b)(n'-c)+d) \leq nn' - dn' + d^2/4 + d.$$

Hence $F(n, n') \leq n'(n-d) + d^2/4 + d$. Since $n' \geq n/3$, the desired result follows.

Remark. This result suggests defining a function $S(n)$ to be the least positive integer r such that there exists an integer m, $0 \leq m \leq r$, with $\gcd(r-i, m-j) > 1$ for $0 \leq i, j < n$. By an argument similar to that given above, we know that $S(n) < e^{(1+o(1))2n^2 \log n}$. The following table gives the first few values of $S(n)$:

n	$S(n)$	m
1	2	0
2	21	15
3	1310	1276

It is possible to prove through brute force calculation that $450000 < S(4) \leq 172379781$. The upper bound follows from the fact that if $(x, y) = (172379781, 153132345)$, then we have $\gcd(x-i, y-j) > 1$ for $0 \leq i, j < 4$.