The long arm of the law

Kari Eloranta

Institute of Mathematics
Aalto University, Finland

Representing Streams, Dec. 10 - 14, 2012
Lorentz Center, Leiden, the Netherlands

The problem

Definition

Consider the spaces of infinite 1-dimensional sequences of symbols from $S = \{1, 2, 3, \ldots, d\}$ with an **exclusion rule***:*

$$X_{(d,f)} = \left\{ x \in S^{\mathbf{Z}} \mid x_i \neq x_{i+f(n)},\ i \in \mathbf{Z},\ n \in \mathbf{N} \right\} \tag{1}$$

where $f : \mathbf{N} \to \mathbf{N}$ is a strictly increasing function.

One-sided case $X^{+}_{(d,f)}$: **Z** in (1) replaced by **N**.

Basic questions: When is $X_{(d,f)}$ non-empty? Can it be of exponential size? What are generic elements like? If only finite sequences, what are they like?

The model was originally proposed by Mike Keane with $f(n) = n^2$.

The problem

Definition

Consider the spaces of infinite 1-dimensional sequences of symbols from $S = \{1, 2, 3, \dots, d\}$ *with an* **exclusion rule***:*

$$X_{(d,f)} = \left\{ x \in S^{\mathbf{Z}} \,|\, x_i \neq x_{i+f(n)},\ i \in \mathbf{Z},\ n \in \mathbf{N} \right\} \tag{1}$$

where $f : \mathbf{N} \to \mathbf{N}$ *is a strictly increasing function.*

One-sided case $X^+_{(d,f)}$: $\mathbf{Z}$ in (1) replaced by $\mathbf{N}$.

Basic questions: When is $X_{(d,f)}$ non-empty? Can it be of exponential size? What are generic elements like? If only finite sequences, what are they like?

The model was originally proposed by Mike Keane with $f(n) = n^2$.

Examples, linear f

$S = \{1, 2\}$ and $f(n) = 2n$.
$x_0 = 1$ implies $x_{2k} = 2, \forall k \neq 0$. But $x_2 = 2$ implies $x_{2m} = 1, \ \forall m \neq 1$, a contradiction. So $X_{(2,2n)} = \emptyset$.

In fact $X_{(d,kn)} = \emptyset$ for all $d, k \geq 2$. Just exhaust S:

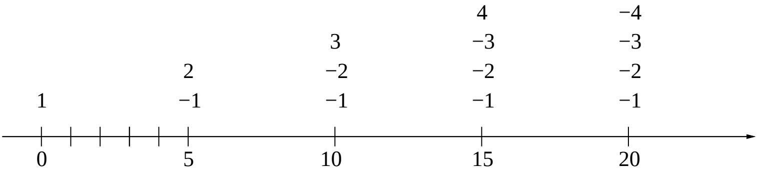


Figure: $X_{(d,5n)} = \emptyset$.

But for $S = \{1, 2\}$ and $f(n) = 2n - 1$ we have periodic points $(12)^*$, hence $X_{(2,2n-1)} \neq \emptyset$. Favourable parity!

Examples, linear f

$S = \{1, 2\}$ and $f(n) = 2n$.
$x_0 = 1$ implies $x_{2k} = 2, \forall k \neq 0$. But $x_2 = 2$ implies $x_{2m} = 1, \ \forall m \neq 1$, a contradiction. So $X_{(2,2n)} = \emptyset$.

In fact $X_{(d,kn)} = \emptyset$ for all $d, k \geq 2$. Just exhaust S:

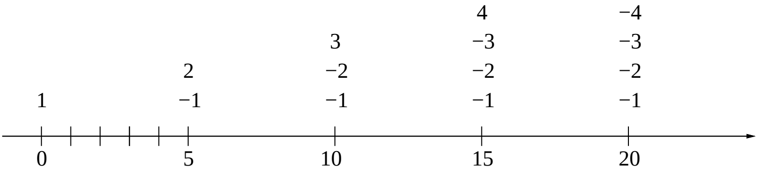


Figure: $X_{(d,5n)} = \emptyset$.

But for $S = \{1, 2\}$ and $f(n) = 2n - 1$ we have periodic points $(12)^*$, hence $X_{(2,2n-1)} \neq \emptyset$. Favourable parity!

Examples, faster growing f

$S = \{1, 2\}$ and $f(n) = n^r$, $r = 2, 3, \ldots$
If $x_0 = 1$ then $x_{2i} = 1$, $\forall i \in \mathbf{Z}$ so in particular $x_{2^r} = 1$, a contradiction.
Therefore $X_{(2,n^r)} = \emptyset$.

Suppose there is $m \in \mathbf{N}$ which does not divide any of the values $f(n)$, $n \in \mathbf{N}$.
Then for $d \geq m$ we can have periodic points.
For example $X_{(3,2^n)}$ and $X^+_{(4,\{primes\})}$ are nonempty.

$X^+_{(d,n!)}$, $d = 2$ or 3 can immediately be seen to be empty. But ...

Examples, faster growing f

$S = \{1, 2\}$ and $f(n) = n^r$, $r = 2, 3, \ldots$
If $x_0 = 1$ then $x_{2i} = 1$, $\forall i \in \mathbf{Z}$ so in particular $x_{2^r} = 1$, a contradiction.
Therefore $X_{(2,n^r)} = \emptyset$.

Suppose there is $m \in \mathbf{N}$ which does not divide any of the values $f(n)$, $n \in \mathbf{N}$.
Then for $d \geq m$ we can have periodic points.
For example $X_{(3,2^n)}$ and $X^+_{(4,\{primes\})}$ are nonempty.

$X^+_{(d,n!)}$, $d = 2$ or 3 can immediately be seen to be empty. But ...

Examples, faster growing f

$S = \{1, 2\}$ and $f(n) = n^r$, $r = 2, 3, \ldots$
If $x_0 = 1$ then $x_{2i} = 1$, $\forall i \in \mathbf{Z}$ so in particular $x_{2^r} = 1$, a contradiction.
Therefore $X_{(2,n^r)} = \emptyset$.

Suppose there is $m \in \mathbf{N}$ which does not divide any of the values $f(n)$, $n \in \mathbf{N}$.
Then for $d \geq m$ we can have periodic points.
For example $X_{(3,2^n)}$ and $X^+_{(4,\{primes\})}$ are nonempty.

$X^+_{(d,n!)}$, $d = 2$ or 3 can immediately be seen to be empty. But ...

Non-trivial example

... $X^{+}_{(4,n!)}$ could be non-trivial. There is a period (of length 25) which repeats almost until the exclusion would violate it for the first time at 5041.

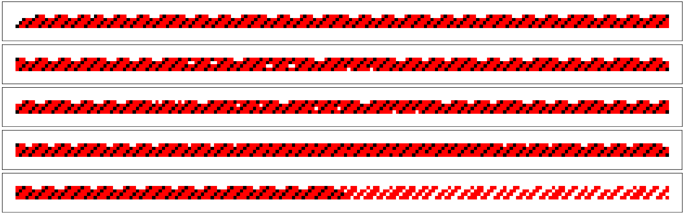


Figure: Lexicographically generated candidate for $X^{+}_{(4,n!)}$ (from $x_1 = 1$). Segments, from top: 1-200, 4950-5150, 10000-10200, 362950-363150, 499900-500100.

Periodicity contradicted at intervals of length $n!, n = 7, 10, 11, 12 \ldots$
but the sequence generation survives them at least half a million steps.

Languages

Proposition

If for any natural m there is a natural n such that we have $m|f(n)$ *then the words satisfying the exclusion do not form a context-free language. Hence the sequences do not form a regular language (sofic shift) either.*

Proof by showing that the validity of the Pumping lemma is dependent on the (non)divisibility property.

Beyond this... need detailed info on f-residues.

Languages

Proposition

If for any natural m there is a natural n such that we have $m|f(n)$ then the words satisfying the exclusion do not form a context-free language. Hence the sequences do not form a regular language (sofic shift) either.

Proof by showing that the validity of the Pumping lemma is dependent on the (non)divisibility property.

Beyond this... need detailed info on f-residues.

Powers

For $d \geq 3$ and $f(n) = n^r$, $r = 2, 3, \ldots$ it is known that:

- None of the corresponding languages are context-free (by the Proposition).
- For $X^+_{(3,n^3)}$ sequences of length at least 300 can be generated.
- $X^+_{(3,n^2)}$ and $X_{(3,n^2)}$ are empty (elementary argument).
- $X^+_{(4,n^2)} = \emptyset$ by a computer assisted proof. Max sequence length is 47.
- For $d = 5$ one can generate sequences of length at least 170.
- Random generation of sequences for $X^+_{(d,n^2)}$, $d = 5, 6, 7, 10, 15$ and 20 suggest strongly that all these spaces are empty.

Powers

For $d \geq 3$ and $f(n) = n^r$, $r = 2, 3, \ldots$ it is known that:

- None of the corresponding languages are context-free (by the Proposition).
- For $X^+_{(3,n^3)}$ sequences of length at least 300 can be generated.
- $X^+_{(3,n^2)}$ and $X_{(3,n^2)}$ are empty (elementary argument).
- $X^+_{(4,n^2)} = \emptyset$ by a computer assisted proof. Max sequence length is 47.
- For $d = 5$ one can generate sequences of length at least 170.
- Random generation of sequences for $X^+_{(d,n^2)}$, $d = 5, 6, 7, 10, 15$ and 20 suggest strongly that all these spaces are empty.

Powers

For $d \geq 3$ and $f(n) = n^r$, $r = 2, 3, \ldots$ it is known that:

- None of the corresponding languages are context-free (by the Proposition).
- For $X^+_{(3,n^3)}$ sequences of length at least 300 can be generated.
- $X^+_{(3,n^2)}$ and $X_{(3,n^2)}$ are empty (elementary argument).
- $X^+_{(4,n^2)} = \emptyset$ by a computer assisted proof. Max sequence length is 47.
- For $d = 5$ one can generate sequences of length at least 170.
- Random generation of sequences for $X^+_{(d,n^2)}$, $d = 5, 6, 7, 10, 15$ and 20 suggest strongly that all these spaces are empty.

Algorithm for one-sided sequences

Algorithm v2.0:

0. `set` $M \geq 1$, `let` $S_j = S$ `at each` $j \in \{1, \ldots, M\}$ `and set` $i = 1$.
1. `if` $S_i = \emptyset$ `then` **halt**,
   `else pick uniformly a random symbol` $s \in S_i$.
2. `update` $S_j \leftarrow S_j \setminus \{s\}$ `for all` $j = i + f(n) \in \{i + 1, \ldots, M\}$, $n \in \mathbf{N}$.
3. `if` $i = M$ **halt** `and call` **full length**,

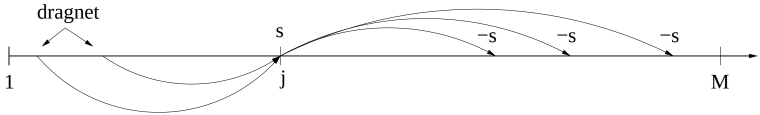
i.e. each coordinate is chosen independently and uniformly but in such a way as to respect the restrictions from all the relevant coordinates in its past.

Probabilistic model for $f(n) = n^2$

Dragnet D_j is the set of coordinates less than j restricting the assignment at j. Its cardinality is a step-function, equal to d from the start of the first **interval** at coordinate $j = d^2 + 1$.

The i^{th} interval is from $(d + i - 1)^2 + 1$ to $(d + i)^2$ (length $l_i = 2(d + i) - 1$).

If the sites on the dragnet D_j support the entire alphabet S then at site j there is a **full block**. First full block is possible at the start of the first interval.



Probabilistic model for n^2

Assume that all the symbols on $\{1, 2, \ldots, j-1\}$ have been laid out independently and uniformly from S. Then

Proposition

Let B_j be the event that one has the first full block at j in the i^{th} interval. Then

(2) $$\mathbf{P}(B_j) = p_i = \frac{1}{d^{d+i-1}} \sum_{\substack{k_r \geq 1,\ r=1,\ldots,d \\ k_1+\cdots+k_d=d+i-1}} \binom{d+i-1}{k_1\ k_2\ \ldots\ k_d}$$

where the sum is d-fold over the given positive integers.

Recall the multinomial: $\binom{a}{b_1\ b_2\ \ldots\ b_d} = \frac{a!}{b_1!b_2!\cdots b_d!}$, $\sum_{i=1}^{d} b_i = a$.
Proof is just combinatorics on the dragnet.

Probabilistic model for n^2

On the interval with dragnet cardinality $d + i - 1$ the sequence extension halts w.p. p_i and its length on the interval $\sim Geom(p_i)$.

Lemma

For an alphabet S of size d one has for all $i \geq 1$

$$1 - p_i < d \left(1 - \frac{1}{d}\right)^{d-1} \left(1 - \frac{1}{d}\right)^{i}.$$

For the proof of the Lemma one has to consider the entries on the $(d + i - 1)^{th}$ level (from the top) of Pascal's d-pyramid. Multinomial Theorem gives the total sum but for $1 - p_i$ we need to bound its boundary sum. Note that $p_i \uparrow 1$ is obvious, but its geometric lower bound requires some work.

Probabilistic model for n^2

On the interval with dragnet cardinality $d + i - 1$ the sequence extension halts w.p. p_i and its length on the interval $\sim Geom(p_i)$.

Lemma

For an alphabet S of size d one has for all $i \geq 1$

$$1 - p_i < d \left(1 - \frac{1}{d}\right)^{d-1} \left(1 - \frac{1}{d}\right)^{i}.$$

For the proof of the Lemma one has to consider the entries on the $(d + i - 1)^{th}$ level (from the top) of Pascal's d-pyramid. Multinomial Theorem gives the total sum but for $1 - p_i$ we need to bound its boundary sum. Note that $p_i \uparrow 1$ is obvious, but its geometric lower bound requires some work.

Probabilistic model for n^2

Theorem

Let the assumptions on the sequence be as above. Then a full block materializes at j i.e. $\mathbf{P}(\textit{sequence generation halts at } j) =$

$$\begin{cases} 0 & 1 \leq j \leq d^2 \\ (1-p_1)^{[j-d^2-1]}\, p_1 & j \text{ in the first interval} \\ \left(\prod_{k=1}^{i-1}(1-p_k)^{l_k}\right)(1-p_i)^{[j-d^2-1-\sum_{k=1}^{i-1} l_k]} p_i & j \text{ in the } i^{th} \text{ interval, } i \geq 2 \end{cases}$$

and the halting time distribution has a geometric tail.
The sequences generated are almost surely of finite length.

The Theorem follows by combining the geometric halting probabilities on the intervals, "uniformizing" them for a tail estimate (l_k are not equal) and finally using Borel-Cantelli.

Probabilistic model for n^2

Theorem

Let the assumptions on the sequence be as above. Then a full block materializes at j i.e. $\mathbf{P}(\textit{sequence generation halts at } j) =$

$$\begin{cases} 0 & 1 \le j \le d^2 \\ (1-p_1)^{[j-d^2-1]}\, p_1 & j \text{ in the first interval} \\ \left(\prod_{k=1}^{i-1}(1-p_k)^{l_k}\right)(1-p_i)^{[j-d^2-1-\sum_{k=1}^{i-1} l_k]} p_i & j \text{ in the } i^{th} \text{ interval, } i \ge 2 \end{cases}$$

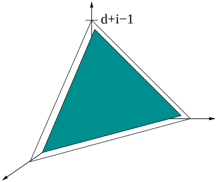
and the halting time distribution has a geometric tail.
The sequences generated are almost surely of finite length.

The Theorem follows by combining the geometric halting probabilities on the intervals, "uniformizing" them for a tail estimate (l_k are not equal) and finally using Borel-Cantelli.

Remarks

- The sum

$$\sum_{\substack{k_r \geq 1,\ r=1,\ldots,d \\ k_1+\cdots+k_d=d+i-1}} \binom{d+i-1}{k_1\ k_2\ \ldots\ k_d}$$



has asymptotically an exponential number of summands both in d and i. To use the Theorem for large d and i one needs to find an efficient way to compute the p_i's.

For small i the sum can be compressed. E.g. for $i = 4$:

$$\binom{d}{1}\binom{d+3}{1\ \ldots\ 1\ 4} + 2\binom{d}{2}\binom{d+3}{1\ \ldots\ 1\ 2\ 3} + \binom{d}{3}\binom{d+3}{1\ \ldots\ 1\ 2\ 2\ 2}$$

but this gets complicated soon... Estimates for the tail if $i \gg 1$.

Remarks

- While $p_i \uparrow 1$ monotonically, the halting distribution is jagged: At the i^{th} jump

$$\frac{\mathbf{P}(halts\ at\ (d+i)^2+1)}{\mathbf{P}(halts\ at\ (d+i)^2)} = \frac{1-p_i}{p_i} p_{i+1} \to 0 \quad as \quad i \to \infty$$

  but far exceeds 1 earlier.
- The independence assumption seems heavy for small alphabet but less so for a large one. But actually...

Reality check for n^2 and $d = 5, 10$ and 15

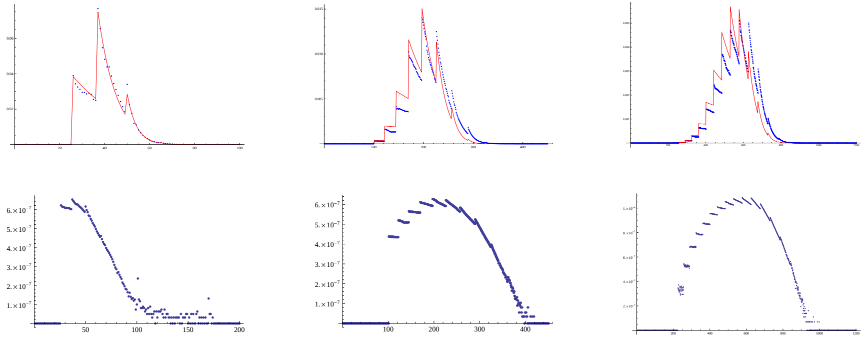


Figure: Top row: empirical (blue/rough) and theoretical (red/smooth) halting probability distributions. Bottom row: log of that above for the data.

Dependencies, squared

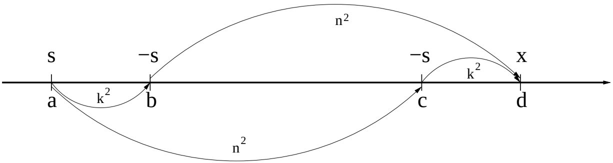


Figure: A dependency mechanism affecting the termination probability. Dragnet at d.

$$\mathbf{P}\left(\textit{full block at } d \;\middle|\; k^2+n^2 \textit{ not square}\right) < \mathbf{P}\left(\textit{full block at } d\right) < \mathbf{P}\left(\textit{full block at } d \;\middle|\; k^2+n^2 \textit{ square}\right)$$

As k and n vary, the non-square case is far more likely to occur than the square case. So termination probabilities of the independent model should major the observed ones.

Dependencies, squared

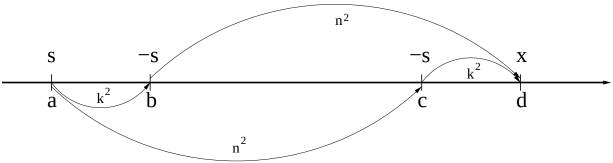


Figure: A dependency mechanism affecting the termination probability. Dragnet at d.

$$\mathbf{P}\left(\textit{full block at } d \;\middle|\; k^2+n^2 \textit{ not square}\right) < \mathbf{P}\left(\textit{full block at } d\right) < \mathbf{P}\left(\textit{full block at } d \;\middle|\; k^2+n^2 \textit{ square}\right)$$

As k and n vary, the non-square case is far more likely to occur than the square case. So termination probabilities of the independent model should major the observed ones.

Statistics of the sequence lengths

Symbols d	Empirical mean	Empirical std. dev.	Model mean	Model std. dev.	Sequences
4	27.2542	5.13374	23.992	5.23924	$50 \cdot 10^6$
5	39.5672	8.28983	39.2172	8.22516	$80 \cdot 10^6$
6	60.8247	13.5813	59.3666	11.9713	$80 \cdot 10^6$
7	89.4687	18.5912	84.982	16.5113	$30 \cdot 10^6$
10	209.315	38.2887	199.562	35.1369	$20 \cdot 10^6$
15	566.87	92.2796	543.291	84.4349	$10 \cdot 10^6$
20	1156.57	170.829	$*$	$*$	$5 \cdot 10^6$

Table: Data from randomly generated one-sided sequences and the probabilistic model. Asterisks are due to missing coefficients (for i large).

Conjecture

Based on the data one might venture to...

Conjecture

(i) All the spaces $X^{+}_{(d,n^2)}$ *and* $X_{(d,n^2)}$, $d \geq 1$ *are empty.*

(ii) Suppose $T^{(d)}$ *is the halting instant of the Algorithm v2.0. For sufficiently rapidly growing* $M(d)$ *there are positive constants a and b such that as* $d \to \infty$

$$\mathbf{P}\left(\frac{T^{(d)} - ad^{5/2}}{bd^{15/7}} \leq x\right) \longrightarrow \Phi(x) \qquad \forall x \in \mathbf{R}$$

where Φ *is the cumulative distribution function of the standard normal* $N(0,1)$.

CLT should hold for the probabilistic model as well (with parameters but not exponents adjusted)

$M(d)$ just needs to outgrow the off-set rate $d^{5/2}$.

Termination details for n^2

One can record when the upcoming termination can be seen for the first time (x-coord.) and how far ahead it will be (y):

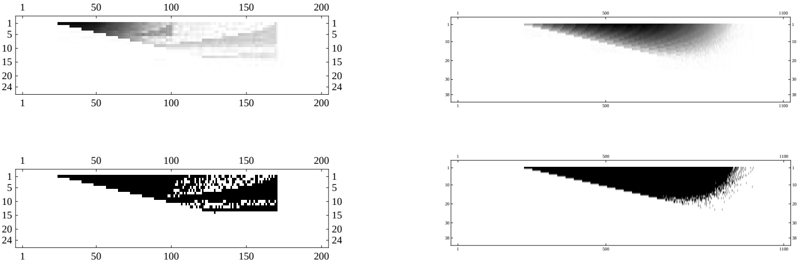


Figure: Terminal jump distribution (log and sign, top and bottom resp.) for $d = 5$ and 15 (20 and 10 million samples).

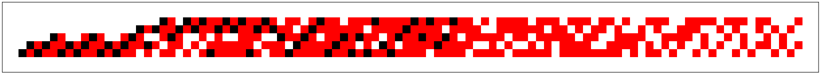
Termination details for n^2

- Left slope is due to the one-sidedness: exactly quantifiable.
- For about $d \leq 7$ some (number theoretic?) constraints rule the interior and the right edge.
- Beyond this range of d the termination seems like a random process.
- Randomness conspires in favor of showing $\emptyset$!

Rest

arXiv:math-ph/1204.3439
or
www.math.hut.fi/∼kve/research.html.en

Thank you!



A **context-free language** is recognized by a non-deterministic pushdown automaton. Such language necessarily satisfies a **Pumping Lemma**:

Lemma

Any sufficiently long string s, say $|s| \geq k$, *can be written as* $s = uvxyz$ *such that*
(i) $|vxy| \leq k$,
(ii) $|vy| \geq 1$,
(iii) $uv^n xy^n z$ *is an allowed string for all natural* n.

If either v or y vanishes but the other is non-trivial (hence (ii) is still valid) this reduces to the Pumping Lemma of **regular languages** (languages recognized by a finite state automaton).

If $p \nmid a$ and the congruence $x^2 \equiv a \pmod{p}$ is soluble then a is called the **quadratic residue modulo** p.

Mod $p = 2$ every integer is a quadratic residue. Let $P = \{1, 2, \dots, p-1\}$. The basic distribution result is

Lemma

Let p be an odd prime. Then exactly half of the integers a on P are quadratic residues modulo p.

Little is known on the distribution of the residues beyond this. 1 is quadratic residue and so is a if a is a square. Maximum number of residues between non-residues is $2\sqrt{p}+1$. If $N(p)$ is the smallest non-residue in P then for large p, $N(p) < p^{1/2+\epsilon}$ (by Burgess, -57). If RH holds then $N(p) = c(\ln p)^2$.

Initializations:

```
(* n2 blocks *)
d = 4; (* number of symbols *)
M = 200; (* length of sequences attempted, must be bigger than imax!  *)
blocks = Table[ If[IntegerQ[Sqrt[i]] == True, 1, 0], i, M]; (* block sites *)


(* ab-array initialization *)
ab = Table[0, d, M];
ab[[1]] = Flatten[Prepend[Drop[-blocks, -1], 1]];
ab[[2]] = Flatten[Prepend[Drop[-blocks, -2], 0, 1]];
col[i] := ab[[1, i]], ab[[2, i]], ab[[3, i]], ab[[4, i]] (* i th column of ab *)


(* for FULL RUN for |S|=4, minimal output!  *)
i = 3; imax = 100; (* max seq.  length constructed *)
base = 0; (* running assumption for the third column *)
maxlength = 0; (* initialization for the maximal length sequence found *)
lowbacktrack = 10; (* highest index from which backtrack is notified *)
```

... and the code...

```
(Label[fwd];
While[ i <= imax,
maxlength = Max[maxlength, i];
locs = Intersection[ Flatten[Position[col[i], 0]],
base + 1, base + 2, base + 3, base + 4]; (* free symbols above base *)
If[Length[locs] == 0, loc = 0, loc = Min[locs] ]; Label[jumpup];
If[loc == 0,
i = i - 1; If[i <= lowbacktrack, Print["cannot assign, will backtrack to:  ", i] ];
Goto[backtr];,
ab[[loc, i ]] = 1; base = loc;(* new symbol assignment *)
blockcols = Flatten[Prepend[Drop[blocks, -i], Table[0, i]]];
ab[[loc]] = ab[[loc]] - blockcols; (* assigning the new blocks *) ];
(* check if full blocks formed *)
Do[ If[ Flatten[ Position[col[Flatten[ Position[blockcols, 1] ][[chkcol]]], 0]] == ,
ab[[loc, i]] = 0;
ab[[loc]] = ab[[loc]] + blockcols; (* taking the new assignment and blocks away *)
If[ loc < Max[locs],
loc = locs[[Flatten[Position[locs, loc]][[1]] + 1]]; Goto[jumpup];,(* try higher symbol *)
i = i - 1; Goto[backtr]; ] ],
chkcol, Length[Flatten[ Position[blockcols, 1]]]  ]; i = i + 1; base = 0; ]; Abort[];
Label[backtr];
If[i == 3 && ab[[3, 3]] == 1,
Print["All done, furthest assignment:  ", maxlength]; Abort[];]
If[i <= lowbacktrack, Print["recalling assignment at:  ", i] ];
base = Flatten[Position[col[i], 1]][[ 1]]; (* position of the assignment to be recalled *)
blockcols = Flatten[Prepend[Drop[blocks, -i], Table[0, i]]];
ab[[base, i]] = 0;
ab[[base]] = ab[[base]] + blockcols; (* taking the assignment and its blocks away *)
If[base == d, i = i - 1; Goto[backtr];, Goto[fwd]; ])
```