Experimental methods applied to the computation of integer sequences

Eric Rowland
erowland@math.rutgers.edu

Department of Mathematics
Rutgers University, New Brunswick / Piscataway

April 2, 2009

Overview

Two topics:

- a prime-generating recurrence
- enumeration of binary trees avoiding a given pattern

Theme is speeding up computation of terms in an integer sequence.

Benefits:

- faster algorithm for computing terms in practice
- new information about the structure of the system

In each problem, computing a sequence quickly is related to understanding the structure of some object.

Prime-generating functions

Several functions are known to generate primes.

- Gandhi's formula for the *n*th prime:
  $p_n = \left\lfloor 1 - \log_2 \left(-\frac{1}{2} + \sum_{d \mid \prod_{k=1}^{n-1} p_k} \frac{\mu(d)}{2^d - 1} \right) \right\rfloor$
- Mills' formula: $\left\lfloor \theta^{3^n} \right\rfloor$, where $\theta = 1.3064\ldots$
- multivariate polynomials of Matijasevič and Jones et al., for which the set of positive values assumed by the polynomial is equal to the set of primes

But all known examples are *engineered*.

Naturally occurring functions

Are there “naturally occurring” functions that generate primes?

- Euler’s polynomial $n^2 + n + 41$ is prime for $0 \leq n \leq 39$.

The recurrence

Are there naturally occurring functions that reliably generate primes?

In 2003 Matthew Frank discovered the recurrence

$$a(n) = a(n-1) + \gcd(n, a(n-1)).$$

Consider the initial condition $a(1) = 7$.

First few terms

n	$\gcd(n, a(n-1))$	$a(n)$
1		7
2	1	8
3	1	9
4	1	10
5	5	15
6	3	18
7	1	19
8	1	20
9	1	21
10	1	22
11	11	33
12	3	36
13	1	37
14	1	38
15	1	39
16	1	40
17	1	41
18	1	42
19	1	43
20	1	44

n	$\gcd(n, a(n-1))$	$a(n)$
21	1	45
22	1	46
23	23	69
24	3	72
25	1	73
26	1	74
27	1	75
28	1	76
29	1	77
30	1	78
31	1	79
32	1	80
33	1	81
34	1	82
35	1	83
36	1	84
37	1	85
38	1	86
39	1	87
40	1	88

n	$\gcd(n, a(n-1))$	$a(n)$
41	1	89
42	1	90
43	1	91
44	1	92
45	1	93
46	1	94
47	47	141
48	3	144
49	1	145
50	5	150
51	3	153
52	1	154
53	1	155
54	1	156
55	1	157
56	1	158
57	1	159
58	1	160
59	1	161
60	1	162

$\gcd(n, a(n-1))$ appears to always be 1 or prime.

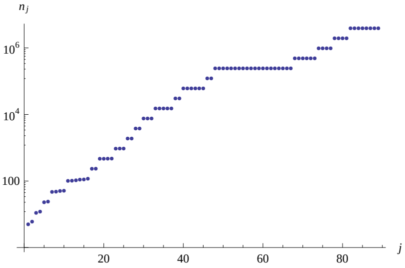
The sequence $\gcd(n, a(n-1))$

1, 1, 1, 5, 3, 1, 1, 1, 1, 11, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 23, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 47, 3, 1, 5, 3,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 101,
3, 1, 1, 7, 1, 1, 1, 1, 11, 3, 1, 1, 1, 1, 1, 13, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 233, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 467, 3, 1, 5, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, . . .

Nontrivial values of gcd$(n, a(n-1))$

5, 3, 11, 3, 23, 3, 47, 3, 5, 3, 101, 3, 7, 11, 3, 13, 233, 3, 467, 3, 5, 3, 941, 3, 7, 1889, 3, 3779, 3, 7559, 3, 13, 15131, 3, 53, 3, 7, 30323, 3, 60647, 3, 5, 3, 101, 3, 121403, 3, 242807, 3, 5, 3, 19, 7, 5, 3, 47, 3, 37, 5, 3, 17, 3, 199, 53, 3, 29, 3, 486041, 3, 7, 421, 23, 3, 972533, 3, 577, 7, 1945649, 3, 163, 7, 3891467, 3, 5, 3, 127, 443, 3, 31, 7783541, 3, 7, 15567089, 3, 19, 29, 3, 5323, 7, 5, 3, 31139561, 3, 41, 3, 5, 3, 62279171, 3, 7, 83, 3, 19, 29, 3, 1103, 3, 5, 3, 13, 7, 124559609, 3, 107, 3, 911, 3, 249120239, 3, 11, 3, 7, 61, 37, 179, 3, 31, 19051, 7, 3793, 23, 3, 5, 3, 6257, 3, 43, 11, 3, 13, 5, 3, 739, 37, 5, 3, 498270791, 3, 19, 11, 3, 41, 3, 5, 3, 996541661, 3, 7, 37, 5, 3, 67, 1993083437, 3, 5, 3, 83, 3, 5, 3, 73, 157, 7, 5, 3, 13, 3986167223, 3, 7, 73, 5, 3, 7, 37, 7, 11, 3, 13, 17, 3, 19, 29, 3, 13, 23, 3, 5, 3, 11, 3, 7972334723, 3, 7, 463, 5, 3, 31, 7, 3797, 3, 5, 3, 15944673761, 3, 11, 3, 5, 3, 17, 3, 53, 3, 139, 607, 17, 3, 5, 3, 11, 3, 7, 113, 3, 11, 3, 5, 3, 293, 3, 5, 3, 53, 3, 5, 3, 151, 11, 3, 31889349053, 3, 63778698107, 3, 5, 3, 491, 3, 1063, 5, 3, 11, 3, 7, 13, 29, 3, 6899, 3, 13, 127557404753, 3, 41, 3, 373, 19, 11, 3, 43, 17, 3, 320839, 255115130849, 3, 510230261699, 3, 72047, 3, 53, 3, 17, 3, 67, 5, 3, 79, 157, 5, 3, 110069, 3, 7, 1020460705907, 3, 5, 3, 43, 179, 3, 557, 3, 167, . . .

Key observations



logarithmic plot of n_j, the jth value of n for which $\gcd(n, a(n-1)) \neq 1$

Ratio between clusters is very nearly 2.

Each cluster is initiated by a large prime p.

Another key observation

n	$\gcd(n, a(n-1))$	$a(n)$	n	$\gcd(n, a(n-1))$	$a(n)$	n	$\gcd(n, a(n-1))$	$a(n)$
1		7	21	1	45	41	1	89
2	1	8	22	1	46	42	1	90
3	1	9	23	23	69	43	1	91
4	1	10	24	3	72	44	1	92
5	5	15	25	1	73	45	1	93
6	3	18	26	1	74	46	1	94
7	1	19	27	1	75	47	47	141
8	1	20	28	1	76	48	3	144
9	1	21	29	1	77	49	1	145
10	1	22	30	1	78	50	5	150
11	11	33	31	1	79	51	3	153
12	3	36	32	1	80	52	1	154
13	1	37	33	1	81	53	1	155
14	1	38	34	1	82	54	1	156
15	1	39	35	1	83	55	1	157
16	1	40	36	1	84	56	1	158
17	1	41	37	1	85	57	1	159
18	1	42	38	1	86	58	1	160
19	1	43	39	1	87	59	1	161
20	1	44	40	1	88	60	1	162

$a(n) = 3n$ whenever $\gcd(n, a(n-1)) \neq 1$.

Local structure

Lemma

Let $n_1 \geq 2$. *Let* $a(n_1) = 3n_1$, *and for* $n > n_1$ *let*

$$a(n) = a(n-1) + \gcd(n, a(n-1)).$$

Let n_2 *be the smallest integer greater than* n_1 *such that* $\gcd(n_2, a(n_2 - 1)) \neq 1$. *Then*

- $\gcd(n_2, a(n_2 - 1)) = p$ *is prime,*
- p *is the smallest prime divisor of* $2n_1 - 1$,
- $n_2 = n_1 + \frac{p-1}{2}$, *and*
- $a(n_2) = 3n_2$.

This lemma provides the inductive step.

Main result

Theorem

Let $a(1) = 7$*, and for* $n > 1$ *let*

$$a(n) = a(n-1) + \gcd(n, a(n-1)).$$

For each $n \geq 2$*,* $\gcd(n, a(n-1))$ *is either* 1 *or prime.*

Is the recurrence a “magical” producer of primes?

No.
Without the shortcut, $\frac{p-3}{2}$ consecutive 1s precede p.
With the shortcut, each step requires finding the smallest prime divisor of $2n-1$.

Other initial conditions

Do all initial conditions produce only 1s and primes? No.
$a(1) = 532$ produces $\gcd(18, a(17)) = \gcd(18, 567) = 9$.
$a(1) = 801$ produces $\gcd(21, a(20)) = \gcd(21, 840) = 21$.

Conjecture

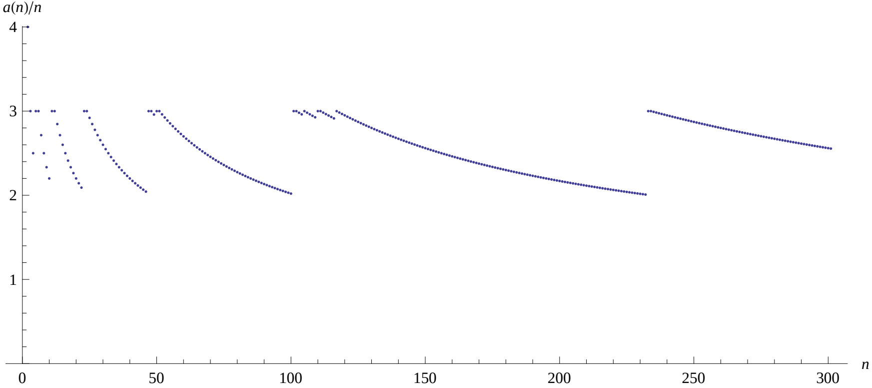
Let $n_1 \geq 1$ and $a(n_1) \geq 1$. For $n > n_1$ let

$$a(n) = a(n-1) + \gcd(n, a(n-1)).$$

Then there exists an N such that for each $n > N$ $\gcd(n, a(n-1))$ is either 1 or prime.

It would suffice to show that $a(n)/n$ always reaches 1, 2, or 3.

Bounding $a(n)/n$ from below



Proposition

If $n_1 \geq 1$ and $a(n_1) > 2n_1 + 1$, then $a(n)/n > 2$ for all $n \geq n_1$.

Bounding $a(n)/n$ from above

Empirically, when $a(n)/n$ is large, it tends to decrease.

Proposition

If $n_1 \geq 1$ and $a(n_1) \geq 1$, then $a(n)/n \leq \lceil a(n_1)/n_1 \rceil$ for all $n \geq n_1$.

Proof.

Let $r = \lceil a(n_1)/n_1 \rceil$. Inductively, assume $a(n-1)/(n-1) \leq r$. Then

$$1 \leq r \leq rn - a(n-1).$$

Since $\gcd(n, a(n-1))$ divides $r \cdot n - a(n-1)$, we have

$$\gcd(n, a(n-1)) \leq rn - a(n-1);$$

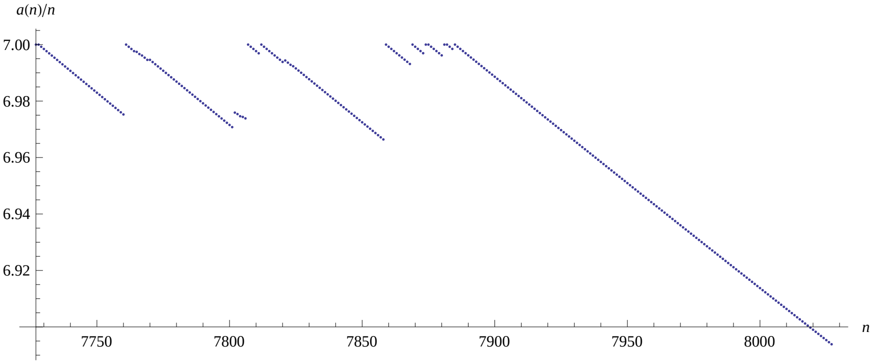
therefore

$$a(n) = a(n-1) + \gcd(n, a(n-1)) \leq rn. \qquad \square$$

Obstruction to the conjecture

It remains to show that $a(n)/n$ cannot remain above 3 indefinitely.

But no global structure is known that might ensure this;
for $a(7727) = 7 \cdot 7727$, $a(n)/n = 7$ reoccurs eleven times.



Transient region

Next best thing: Speed up computation of the transient region.

Let $a(n-1) = n + \Delta$. The recurrence can be interpreted as repeatedly computing the minimal $k \geq 1$ such that $\gcd(n+k, n+\Delta+k) \neq 1$.

Proposition

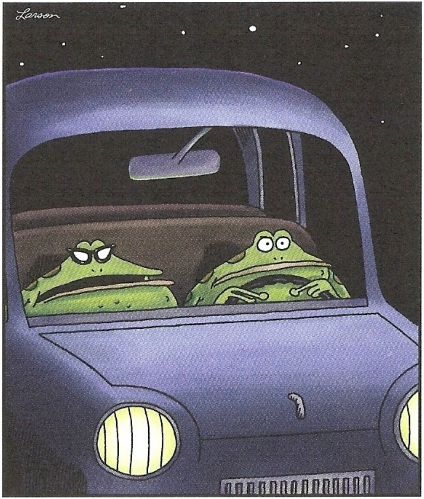
Let $n \geq 0$, $\Delta \geq 2$, and j be integers. Let $k \geq j$ be minimal such that $\gcd(n+k, n+\Delta+k) \neq 1$. Then

$$k = \min \{ \operatorname{mod}_j(-n, p) : p \text{ is a prime dividing } \Delta \},$$

where $\operatorname{mod}_j(x, p) \equiv x \mod p$ *such that* $j \leq \operatorname{mod}_j(x, p) < j + p$.

(The lemma covers the case $\Delta = 2n - 1$ for $j = 1$.)

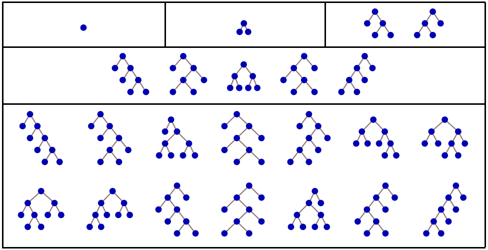
Avoiding trees...



"Watch out for that tree, you idiot! … And *now* you're on the wrong side of the road. Crimony! You're driving like you've been pithed or something."

Binary trees

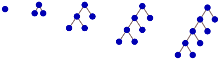
Binary trees with ≤ 5 leaves:

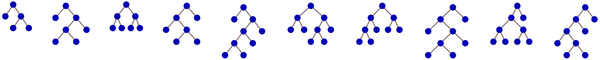


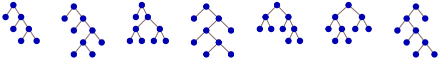
Pattern containment

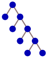
Patterns are contiguous. For example, let $t =$ .

Small binary trees containing...

0 copies of t: 

1 copy of t: 

2 copies of t: 

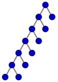
3 copies of t: 

Small patterns

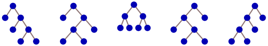
What is the number $a(n)$ of n-leaf binary trees avoiding t?

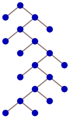
1-leaf tree patterns: $t = \bullet$.
$a(n) = 0$.

2-leaf tree patterns: $t =$ .
$a(1) = 1$; $a(n) = 0$ for $n \geq 2$.

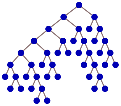
3-leaf tree patterns:  and . "Typical" tree avoiding : .
$a(n) = 1$.

4-leaf tree patterns

4-leaf tree patterns: .

- $t =$ . A "typical" tree avoiding t looks like .

  $a(1) = 1$; $a(n) = 2^{n-2}$ for $n \geq 2$.

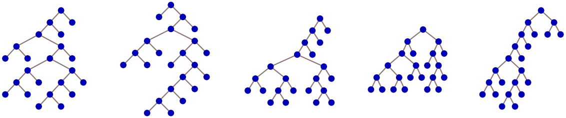
- $t =$ . A "typical" tree avoiding t looks like .

  $a(1) = 1$; $a(n) = 2^{n-2}$ for $n \geq 2$.

These two patterns are equivalent.

The remaining 4-leaf pattern

- $t =$ . Some trees avoiding t:



$a(n) = M_{n-1}$ (a Motzkin number[1]).

---

[1] Robert Donaghey and Louis Shapiro, Motzkin numbers, *Journal of Combinatorial Theory, Series A* **23** (1977) 291–301.

Systematic enumeration

Is there a systematic way to compute $a(n)$ for an arbitrary pattern t?

Let $$\mathrm{Av}_t(x) = \sum_{T \text{ avoids } t} x^{\text{number of vertices in } T} = \sum_{n=0}^{\infty} a(n)x^n.$$

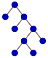
Theorem

$\mathrm{Av}_t(x)$ *is algebraic.*

The proof is constructive.

Example

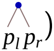
Consider the tree pattern

$t =$  $=$ , where $t_l = \bullet$ and $t_r =$ .

For a given tree pattern p, let

$$\text{weight}(p) := \sum_{T \text{ matches } p \text{ and avoids } t} x^{\text{number of vertices in } T}.$$

Begin with weight($\bullet$) $= x +$ weight(); rewrite weight() by

weight() $=$

$$x \cdot \big(\text{weight}(p_l) \cdot \text{weight}(p_r) - \text{weight}(p_l \cap t_l) \cdot \text{weight}(p_r \cap t_r)\big).$$

System of polynomial equations

$$
\begin{aligned}
\mathrm{weight}(\ ) &= x \cdot \big(\mathrm{weight}(\cdot) \cdot \mathrm{weight}(\cdot) - \mathrm{weight}(\cdot \cap \cdot) \cdot \mathrm{weight}(\cdot \cap \ )\big) \\
&= x \cdot \big(\mathrm{weight}(\cdot)^2 - \mathrm{weight}(\cdot) \cdot \mathrm{weight}(\ )\big)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{weight}(\ ) &= x \cdot \big(\mathrm{weight}(\cdot) \cdot \mathrm{weight}(\ ) - \mathrm{weight}(\cdot \cap \cdot) \cdot \mathrm{weight}(\ \cap \ )\big) \\
&= x \cdot \big(\mathrm{weight}(\cdot) \cdot \mathrm{weight}(\ ) - \mathrm{weight}(\cdot) \cdot \mathrm{weight}(\ )\big)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{weight}(\ ) &= x \cdot \big(\mathrm{weight}(\ ) \cdot \mathrm{weight}(\cdot) - \mathrm{weight}(\ \cap \cdot) \cdot \mathrm{weight}(\cdot \cap \ )\big) \\
&= x \cdot \big(\mathrm{weight}(\ ) \cdot \mathrm{weight}(\cdot) - \mathrm{weight}(\ ) \cdot \mathrm{weight}(\ )\big)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{weight}(\ ) &= x \cdot \big(\mathrm{weight}(\ ) \cdot \mathrm{weight}(\ ) - \mathrm{weight}(\ \cap \cdot) \cdot \mathrm{weight}(\ \cap \ )\big) \\
&= x \cdot \big(\mathrm{weight}(\ ) \cdot \mathrm{weight}(\ ) - \mathrm{weight}(\ ) \cdot \mathrm{weight}(\ )\big)
\end{aligned}
$$

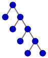
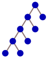
No new variables. Eliminate the four auxiliary variables to obtain

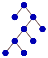
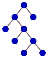
$$x^3 \,\mathrm{weight}(\cdot)^2 - (x^2 - 1)^2 \,\mathrm{weight}(\cdot) - x\,(x^2 - 1) = 0.$$

5-leaf equivalence classes

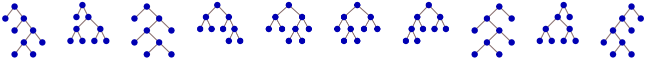
A computer implementation establishes all equivalence classes for binary trees up to 8 leaves.

For 5-leaf tree patterns...

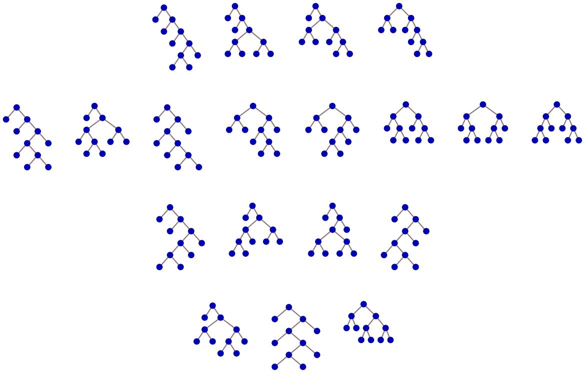
 and  form an equivalence class.

 and  form an equivalence class.

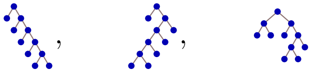
The other 10 tree patterns are equivalent:



6-leaf equivalence classes



Isolated trees:

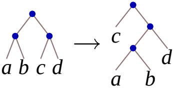


Bijective proofs

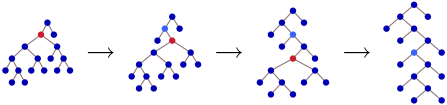
Given two equivalent tree patterns s and t, can we find a bijective proof of the equivalence?

For example,  and  are equivalent.

Let T avoid . Idea: Replace all instances of  with .
How?

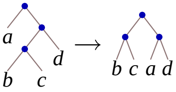


What order? top-down.
For example:

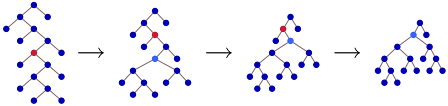


Inverse

The inverse map is a *bottom-up replacement* with the inverse replacement rule:



For example:



Searching for bijections

Generally, we may individually try all $m!$ permutations of leaves.

Permutations whose replacements prove equivalence for pairs of 5-leaf trees:

	t_2	t_3	t_4	t_6	t_7	t_8	t_9	t_{11}	t_{12}	t_{13}
t_2	—	14235		43125						
t_3		—	12534	31245					51234	
t_4		12453	—			41235				
t_6				—	12534	45123				
t_7				12453	—		45123			
t_8				34512		—	31245			
t_9					34512	23145	—			
t_{11}					13452			—	31245	
t_{12}		23451					12453	23145	—	
t_{13}							14532		13425	—

Conjecture

Not every pair of equivalent trees can be shown equivalent by such a bijection. However, the following appears to hold.

Conjecture

Two binary tree patterns s and t are equivalent if and only if there is a sequence of

- *top-down replacements,*
- *bottom-up replacements, and*
- *left–right reflections*

that produces a bijection from binary trees avoiding s to binary trees avoiding t.

The conjecture is true for tree patterns of ≤ 7 leaves.