Minimal Elements for the Prime Numbers

Curtis Bright[1], Jeffrey Shallit[1], Raymond Devillers[2]

[1]University of Waterloo, [2]Université libre de Bruxelles

December 7, 2016

Published in *Experimental Mathematics* (Vol. 25, Issue 3)

Motivation

Fact

The following 26 numbers are prime:

2, 3, 5, 7, 11, 19, 41, 61, 89, 409, 449, 499, 881, 991, 6469, 6949, 9001, 9049, 9649, 9949, 60649, 666649, 946669, 60000049, 66000049, 66600049

Motivation

Fact

The following 26 numbers are prime:

2, 3, 5, 7, 11, 19, 41, 61, 89, 409, 449, 499, 881, 991, 6469, 6949, 9001, 9049, 9649, 9949, 60649, 666649, 946669, 60000049, 66000049, 66600049

Claim

Give me a prime number and I can remove some of its digits to obtain a prime on this list!

Minimal Primes

- The primes in this list are known as the *minimal primes* because this the smallest list of numbers for which this claim holds.

Minimal Sets

- More generally, any language (set of strings over a finite alphabet) has its own *minimal set* of elements and the minimal primes are the minimal set of the language

$$\{2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots\}.$$

Definitions

- x is a *subword* of y if one can strike out zero or more symbols of y to get x.
- A string of symbols s is *minimal* for a language L if
  1. s is a member of L and
  2. s does not contain another member of L as a subword.
- $M(L)$ denotes the set of minimal elements of L.

Higman–Haines Theorem

- $M(L)$ is finite for every language L.

Computation of Minimal Sets

- Computing $M(L)$ is undecidable in general and can be very difficult to compute even for simple languages.

Computation of Minimal Sets

- Computing $M(L)$ is undecidable in general and can be very difficult to compute even for simple languages.
- Can lead to some strange behaviour...

Computation of Minimal Sets

- Computing $M(L)$ is undecidable in general and can be very difficult to compute even for simple languages.
- Can lead to some strange behaviour...
  - The minimal set for primes of the form $4n+1$ has 146 elements, the largest of which has 79 digits.

Computation of Minimal Sets

- Computing $M(L)$ is undecidable in general and can be very difficult to compute even for simple languages.
- Can lead to some strange behaviour...
  - The minimal set for primes of the form $4n + 1$ has 146 elements, the largest of which has 79 digits.
  - The minimal set for primes of the form $4n + 3$ has 113 elements, the largest of which has 19,153 digits!

Computation of Minimal Sets

Proposed Computation Process

- The following process will determine $M(L)$ *if* it can be implemented:

  1. $M \coloneqq \emptyset$
  2. while $L \neq \emptyset$ do
     3. choose x, a shortest string in L
     4. add x to M
     5. remove from L all words containing the subword x
  6. return M

Computation of Minimal Sets

Proposed Computation Process

- The following process will determine $M(L)$ *if* it can be implemented:

  1. $M \coloneqq \emptyset$
  2. while $L \neq \emptyset$ do
     3. choose x, a shortest string in L
     4. add x to M
     5. remove from L all words containing the subword x
  6. return M

- Caveat: We might not have a nice way of performing operations on L.

Computation of Minimal Sets

Using Overapproximations

- This process also works if L is replaced with an overapproximation L', so long as once no more minimal elements remain to be found we can show that $L' = \emptyset$.

Computation of Minimal Sets

Using Overapproximations

- This process also works if L is replaced with an overapproximation L', so long as once no more minimal elements remain to be found we can show that $L' = \emptyset$.
- In practice, we choose L' to be a regular language, e.g.,

$$\{2, 5\} \cup \Sigma^*\{1, 3, 7, 9\}$$

  is a regular overapproximation to the set of primes over the alphabet $\Sigma \coloneqq \{0, \ldots, 9\}$.

Computation of Minimal Sets

Sample Language

- We will work with overapproximations of the form xL^*z where x and z are strings of digits and L is a set of digits.
- To be able to apply the process previously described, we need to be able to test if xL^*z contains a prime or not.

Computation of Minimal Sets

Sample Language

- We will work with overapproximations of the form xL^*z where x and z are strings of digits and L is a set of digits.
- To be able to apply the process previously described, we need to be able to test if xL^*z contains a prime or not.
- It is unknown if this problem is decidable.

Computation of Minimal Sets

Necessary Operations

- In order to perform the process previously described, we need to perform the following operations on the language xL^*z:
  1. Determine if the language contains a prime.
  2. If so, determine the smallest prime(s) in the language.
  3. If a prime is found, shrink the language under consideration so that it no longer contains that prime.

Computation of Minimal Sets

Necessary Operations

- In order to perform the process previously described, we need to perform the following operations on the language xL^*z:
    1. Determine if the language contains a prime.
    2. If so, determine the smallest prime(s) in the language.
    3. If a prime is found, shrink the language under consideration so that it no longer contains that prime.
        - And any strings which contain that prime as a subword.

Proving that xL^*z contains no primes

Method 1: Find a common divisor

Theorem. If N divides xz and all numbers of the form xLz then N divides all numbers of the form xL^*z.

Proving that xL^*z contains no primes

Method 1: Find a common divisor

Theorem. If N divides xz and all numbers of the form xLz then N divides all numbers of the form xL^*z.

Example. 7 divides 49 and 469 so 7 divides 4669, 46669, and all numbers of the form 46^*9.

Proof

N divides xz and all xLz implies N divides all xL^*z

Say $y \in L^*$ contains the digits $y_1, \ldots, y_n$ and z is a digit. By telescoping,

$$\begin{aligned} xyz - xz &= \sum_{i=1}^{n} \left(xy_i y_{i+1} \cdots y_n z - xy_{i+1} \cdots y_n z \right) \\ &= \sum_{i=1}^{n} 10^{n-i} \left(xy_i - x \right) \\ &= \sum_{i=1}^{n} 10^{n-i-1} \left(xy_i z - xz \right) \end{aligned}$$

N must divide xyz since it divides every other term in this equation.

Proving that xL^*z contains no primes

Method 2: Use an algebraic factorization

Let $[x]_b$ represent the evaluation of the string x in base b; the following are some example algebraic factorizations:

$$[\overbrace{4\cdots4}^{n}1]_{16} = (8\cdot4^n+7)(8\cdot4^n-7)/15$$
$$[1\overbrace{0\cdots0}^{n}1]_{8} = (2^{n+1}+1)(4^{n+1}-2^{n+1}+1)$$

Once n is large enough the right side obviously factors and cannot be prime.

Proving that xL^*z contains no primes

Combination method

The family 19^* in base 17 contains no primes, because

$$\big[1\overbrace{9\cdots9}^{2n}\big]_{17} = (5\cdot17^n+3)(5\cdot17^n-3)/16$$

and all $\big[1\overbrace{9\cdots9}^{2n+1}\big]_{17}$ are even, since $[19]_{17}$ and $[1999]_{17}$ are even.

Proving that xL^*z contains a prime

- In practice, if xL^*z could not be ruled out as only containing composites and $|L| > 1$ then a relatively small prime could always be found in the language.
- Intuitively, this is because there are a large number of small strings in such a language, and at least one is likely to be prime.
  - For example, there are 2^{n-2} strings of length n in the language $1\{2,3\}^*1$.

Searching for primes in xy^*z

- In the case $|L| = 1$ the family is of the form xy^*z, and there is only a single string of each length $\geqslant |xz|$.
- Some families xy^*z could not be ruled out as only containing composites and no primes could be found in the family, even after searching through numbers with over 100,000 digits.

Does xy^*z contain large primes?

- The prime number theorem tells us that the chance that a random n-digit number is prime is approximately $1/n$. If one conjectures the numbers xy^*z behave similarly you would expect $\sum_{n=2}^{\infty} 1/n = \infty$ primes of the form xy^*z.

Does xy^*z contain large primes?

- The prime number theorem tells us that the chance that a random n-digit number is prime is approximately $1/n$. If one conjectures the numbers xy^*z behave similarly you would expect $\sum_{n=2}^{\infty} 1/n = \infty$ primes of the form xy^*z.
- Of course, this doesn't always happen, but it's at least a reasonable conjecture in the absence of evidence to the contrary.

In Practice...

- Many xy^*z families contain no small primes even though they do contain very large primes.
- For example, the smallest prime in the base 23 family $9\mathrm{E}^*$ is $9\mathrm{E}^{800873}$ which when written in decimal contains 1,090,573 digits.

In Practice...

- Many xy^*z families contain no small primes even though they do contain very large primes.
- For example, the smallest prime in the base 23 family $9E^*$ is $9E^{800873}$ which when written in decimal contains 1,090,573 digits.
  - Technically, probable primality tests were used to show this (which have a *very* small chance of making an error) because all known primality tests run far too slowly to run on a number of this size.

Shrinking the Language

- Recall that once a minimal prime has been found we want to shrink the language being searched while still keeping it large enough that it contains all remaining minimal primes.

Shrinking xL^*z

- Say that xyz is discovered to be prime with $y \in L$. Then xL^*z can be replaced with

$$x(L \setminus \{y\})^*z.$$

Shrinking xL^*z

- Say that $xyyz$ is discovered to be prime with $y \in L$. Then xL^*z can be replaced with

$$x(L \setminus \{y\})^*z \quad \cup \quad x(L \setminus \{y\})^*y(L \setminus \{y\})^*z.$$

Shrinking xL^*z

- Say that $xy\hat{y}z$ and $x\hat{y}yz$ are discovered to be prime with y, $\hat{y} \in L$ and $y \neq \hat{y}$. Then xL^*z can be replaced with

$$x(L \setminus \{y\})^* z \quad \cup \quad x(L \setminus \{\hat{y}\})^* z.$$

Shrinking xL^*z

- Say that $xy\hat{y}z$ is discovered to be prime with y, $\hat{y} \in L$ and $y \neq \hat{y}$. Then xL^*z can be replaced with

$$x(L \setminus \{y\})^*(L \setminus \{\hat{y}\})^*z.$$

Exploring xL^*z

- If the methods we've discussed cannot be used to rule out or shrink xL^*z where $L = \{y_1, \ldots, y_n\}$ then we can replace it by

$$xL^*y_1z \quad \cup \quad xL^*y_2z \quad \cup \quad \cdots \quad \cup \quad xL^*y_nz$$

  and re-run the methods on this new language.

Experimental Results

- There is no guarantee that the techniques discussed will ever terminate, but in practice they often do.
- They are able to determine the minimal primes of the form $4n + 1$ and $4n + 3$ and the minimal primes expressed in the bases b for $2 \leqslant b \leqslant 16$ and $b = 18, 20, 22, 23, 24$, and 30.

Experimental Results

- There is no guarantee that the techniques discussed will ever terminate, but in practice they often do.
- They are able to determine the minimal primes of the form $4n+1$ and $4n+3$ and the minimal primes expressed in the bases b for $2 \leqslant b \leqslant 16$ and $b = 18$, 20, 22, 23, 24, and 30.
  - The bases $b = 17$, 19, 21, and $25 \leqslant b \leqslant 29$ are solved with the exception of 37 families of the form xy^*z.

Summary of Results for Bases up to 30

Base	# elements	Max. length	# unsolved families
2	2	2	
3	3	3	
4	3	2	
5	8	5	
6	7	5	
7	9	5	
8	15	9	
9	12	4	
10	26	8	
11	152	45	
12	17	8	
13*	228	32,021	
14	240	86	
15	100	107	
16	483	3545	
17*	$\geqslant$ 1279	$\geqslant$ 111,334	1
18	50	33	
19*	$\geqslant$ 3462	$\geqslant$ 110,986	1
20	651	449	
21*	$\geqslant$ 2600	$\geqslant$ 479,150	1
22	1242	764	
23*	6021	800,874	
24	306	100	
25*	$\geqslant$ 17,597	$\geqslant$ 136,967	12
26	$\geqslant$ 5662	$\geqslant$ 8773	2
27*	$\geqslant$ 17,210	$\geqslant$ 109,006	5
28*	$\geqslant$ 5783	$\geqslant$ 94,538	1
29*	$\geqslant$ 57,283	$\geqslant$ 174,240	14
30	220	1024	

*Data based on probable primality tests.

Unsolved Families

Base	Family	Algebraic form	Base	Family	Algebraic form
17	F19*	$(5 \cdot 821 \cdot 17^n - 3^2)/16$	29	1A*	$(19 \cdot 29^n - 5)/14$
19	EE16*	$(2^2 \cdot 13 \cdot 307 \cdot 19^n - 1)/3$		68L0*6	$7 \cdot 757 \cdot 29^{n+1} + 2 \cdot 3$
21	G0*FK	$2^4 \cdot 21^{n+2} + 5 \cdot 67$		AMP*	$(8761 \cdot 29^n - 5^2)/28$
25	6MF*9	$(1381 \cdot 25^{n+1} - 53)/8$		C*FK	$(3 \cdot 29^{n+2} + 2 \cdot 331)/7$
	CM1*	$(59 \cdot 131 \cdot 25^n - 1)/24$		F*0PF	$(3 \cdot 5 \cdot 29^{n+3} + 139 \cdot 1583)/28$
	EE1*	$(8737 \cdot 25^n - 1)/24$		FKI*	$(6379 \cdot 29^n - 3^2)/14$
	E1*E	$(337 \cdot 25^{n+1} + 311)/24$		F*0P	$(3 \cdot 5 \cdot 29^{n+2} + 7573)/28$
	EF0*	$2 \cdot 3 \cdot 61 \cdot 25^n - 1$		LP09*	$(31 \cdot 16607 \cdot 29^n - 3^2)/28$
	F1*F1	$(19^2 \cdot 25^{n+2} + 37 \cdot 227)/24$		00PS*A	$2 \cdot 10453 \cdot 29^{n+1} - 19$
	F0*K0	$3 \cdot 5 \cdot 25^{n+2} + 2^2 \cdot 131$		PC*	$(2 \cdot 89 \cdot 29^n - 3)/7$
	F0K*0	$(5 \cdot 11 \cdot 41 \cdot 25^{n+1} + 19)/6$		PPPL*0	$(87103 \cdot 29^{n+1} + 3^2)/4$
	L0L*8	$(53 \cdot 83 \cdot 25^{n+1} - 3 \cdot 37)/8$		Q*GL	$(13 \cdot 29^{n+2} - 3 \cdot 1381)/14$
	M1*F1	$(23^2 \cdot 25^{n+2} + 37 \cdot 227)/24$		Q*L0	$(13 \cdot 29^{n+2} - 19 \cdot 109)/14$
	M10*8	$19 \cdot 29 \cdot 25^{n+1} + 2^3$		RM*G	$(389 \cdot 29^{n+1} - 5 \cdot 19)/14$
	0L*8	$(199 \cdot 25^{n+1} - 3 \cdot 37)/8$			
26	A*6F	$(2 \cdot 26^{n+2} - 7 \cdot 71)/5$			
	I*GL	$(2 \cdot 3^2 \cdot 26^{n+2} - 11 \cdot 113)/25$			
27	80*9A	$2^3 \cdot 27^{n+2} + 11 \cdot 23$			
	999G*	$(101 \cdot 877 \cdot 27^n - 2^3)/13$			
	CL*E	$(3^2 \cdot 37 \cdot 27^{n+1} - 7 \cdot 29)/26$			
	EI*F8	$(191 \cdot 27^{n+2} - 2^3 \cdot 149)/13$			
	F*9FM	$(3 \cdot 5 \cdot 27^{n+3} - 113557)/26$			
28	0A*F	$(2 \cdot 7 \cdot 47 \cdot 28^{n+1} + 5^3)/27$			