Minimal Elements for the Prime Numbers

Curtis Bright\textsuperscript{1}, Jeffrey Shallit\textsuperscript{1}, Raymond Devillers\textsuperscript{2}

\textsuperscript{1}University of Waterloo, \textsuperscript{2}Université libre de Bruxelles

December 7, 2016

Published in \textit{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.

- 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

- 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. $M := \emptyset$
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 $x$
3. return $M$

Caveat: We might not have a nice way of performing operations on $L$. 
Computation of Minimal Sets

Proposed Computation Process

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

1. $M := \emptyset$
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 $x$
3. 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 $\Sigma := \{0, \ldots, 9\}$. 
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.
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.
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,

$$xyz - xz = \sum_{i=1}^{n} (xy_i y_{i+1} \cdots y_n z - xy_{i+1} \cdots y_n z)$$

$$= \sum_{i=1}^{n} 10^{n-i} (xy_i - x)$$

$$= \sum_{i=1}^{n} 10^{n-i-1} (xy_i z - xz)$$

$N$ 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:

\[
\left[\underbrace{4 \cdots 4}_{n} 1\right]_{16} = (8 \cdot 4^n + 7)(8 \cdot 4^n - 7)/15
\]

\[
\left[\underbrace{1 0 \cdots 0}_{n} 1\right]_8 = (2^{n+1} + 1)(4^{n+1} - 2^{n+1} + 1)
\]

Once \(n\) is large enough the right side obviously factors and cannot be prime.
The family 19* in base 17 contains no primes, because

\[
[19\cdots9]_{17}^{2n} = (5 \cdot 17^n + 3)(5 \cdot 17^n - 3)/16
\]

and all \([19\cdots9]_{17}^{2n+1}\) 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 $1\{2,3\}^*1$. 


In the case $|L| = 1$ the family is of the form $xy^*z$, and there is only a single string of each length $\geq |xz|$.

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 $xy^*z$. Of course, this doesn't always happen, but it's at least a reasonable conjecture in the absence of evidence to the contrary.
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.
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.
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

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

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

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

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 $2 \leq b \leq 16$ 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 $2 \leq b \leq 16$ and $b = 18, 20, 22, 23, 24,$ and $30$.
  - The bases $b = 17, 19, 21,$ and $25 \leq b \leq 29$ are solved with the exception of 37 families of the form $xy^*z$. 
### Summary of Results for Bases up to 30

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

*Data based on probable primality tests.*
<table>
<thead>
<tr>
<th>Base</th>
<th>Family</th>
<th>Algebraic form</th>
<th>Base</th>
<th>Family</th>
<th>Algebraic form</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>F19*</td>
<td>$(5 \cdot 821 \cdot 17^n - 3^2)/16$</td>
<td>29</td>
<td>1A*</td>
<td>$(19 \cdot 29^n - 5)/14$</td>
</tr>
<tr>
<td>19</td>
<td>EE16*</td>
<td>$(2^2 \cdot 13 \cdot 307 \cdot 19^n - 1)/3$</td>
<td>68L0*6</td>
<td>AMP*</td>
<td>$(8761 \cdot 29^n - 5^2)/28$</td>
</tr>
<tr>
<td>21</td>
<td>G0*FK</td>
<td>$2^4 \cdot 21^{n+2} + 5 \cdot 67$</td>
<td>F0</td>
<td>F0K</td>
<td>$(3 \cdot 29^n + 2 \cdot 331)/7$</td>
</tr>
<tr>
<td>25</td>
<td>6MF*9</td>
<td>$(1381 \cdot 25^n + 1 - 53)/8$</td>
<td>EE1*</td>
<td>EE1</td>
<td>$(3 \cdot 5 \cdot 29^n + 3 + 139 \cdot 1583)/28$</td>
</tr>
<tr>
<td>CM1*</td>
<td>(59 \cdot 131 \cdot 25^n - 1)/24</td>
<td>F1*E</td>
<td>(337 \cdot 25^{n+1} + 311)/24</td>
<td>F0E*</td>
<td>(3 \cdot 5 \cdot 29^{n+2} + 7573)/28</td>
</tr>
<tr>
<td>F1*F1</td>
<td>$2 \cdot 3 \cdot 61 \cdot 25^n - 1$</td>
<td>F1*F1</td>
<td>$(10^n - 2 \cdot 3 \cdot 877 \cdot 27^n - 2^3)/13$</td>
<td>OOPS*A</td>
<td>$2 \cdot 10453 \cdot 29^n + 1 - 19$</td>
</tr>
<tr>
<td>F0*K0</td>
<td>$3 \cdot 5 \cdot 25^{n+2} + 2^2 \cdot 131$</td>
<td>LOLO*8</td>
<td>$(53 \cdot 83 \cdot 25^{n+1} - 3 \cdot 37)/8$</td>
<td>PC*</td>
<td>$(2 \cdot 89 \cdot 29^n - 3)/7$</td>
</tr>
<tr>
<td>F0K*0</td>
<td>$(5 \cdot 11 \cdot 41 \cdot 25^{n+1} + 19)/6$</td>
<td>LOL*8</td>
<td>$(53 \cdot 83 \cdot 25^{n+1} - 3 \cdot 37)/8$</td>
<td>PPPL*0</td>
<td>$(87103 \cdot 29^n + 3^2)/4$</td>
</tr>
<tr>
<td>26</td>
<td>M1*F1</td>
<td>$(23^2 \cdot 25^{n+2} + 37 \cdot 227)/24$</td>
<td>M10*8</td>
<td>$(19 \cdot 29 \cdot 25^{n+1} + 2^3$</td>
<td>Q*GL</td>
</tr>
<tr>
<td>27</td>
<td>OL*8</td>
<td>$(199 \cdot 25^{n+1} - 3 \cdot 37)/8$</td>
<td>A*6F</td>
<td>$(2 \cdot 26^{n+2} - 7 \cdot 71)/5$</td>
<td>Q*LO</td>
</tr>
<tr>
<td>I*GL</td>
<td>$(2 \cdot 3^2 \cdot 26^{n+2} - 11 \cdot 113)/25$</td>
<td>80*9A</td>
<td>$2^3 \cdot 27^n + 2 + 11 \cdot 23$</td>
<td>RM*G</td>
<td>$(389 \cdot 29^{n+1} - 5 \cdot 19)/14$</td>
</tr>
<tr>
<td>28</td>
<td>999G*</td>
<td>$(101 \cdot 877 \cdot 27^n - 2^3)/13$</td>
<td>CL*E</td>
<td>$(3^2 \cdot 37 \cdot 27^{n+1} - 7 \cdot 29)/26$</td>
<td></td>
</tr>
</tbody>
</table>