*Acta Universitatis Matthiae Belii ser. Mathematics*, 15(2009), 21–35.
*Received: 26 June 2009, Accepted: 24 November 2009.*

# SUPERPRIMES AND GENERALIZED DIRICHLET THEOREM

MIROSLAV HAVIAR AND PETER MALIČKÝ

*Dedicated to the 70th birthday of Alfonz Haviar*

Abstract. A concept of a superprime meaning a prime number whose all digits are prime numbers is introduced and a question whether there is an infinite number of superprimes is raised. A positive answer to this and a few related questions is conjectured and supported by several observations and computations via Mathematica. Among the conjectures is a generalized version of Dirichlet's Theorem on primes which implies certain conjectures presented here as well as the famous conjectures about the infinite number of Mersenne and Fermat primes.

## 1. The main problem

There are several different proofs of the fact that there is an infinite number of primes [1], the best known being likely the one due to Euclid. In this note we introduce a more specific notion of a superprime and ask if there is still an infinite number of superprimes.

**Definition 1.1.** By a *superprime* we mean a prime number whose all digits (in its decimal representation) are prime numbers.

We note that instead of the decimal representation one can consider base $m$ positional notation for $m \geq 4$. (The case $m = 3$ is not interesting as it gives us only one prime digit 2.)

**Example 1.2.** The numbers 2, 3, 5, 7, 23, 37, 53, 73, 223, 227, 233, 257, 277, 337, 353, 373, 523, 557, 577, 727, 733, 757 and 773 are all superprimes among the natural numbers up to one thousand.

**Problem 1.** *Is there an infinite number of superprimes?*

---

2000 *Mathematics Subject Classification.* Primary 06D15; Secondary 08A40.
*Key words and phrases.* Prime number, Dirichlet's Theorem, Prime Number Theorem, superprime, Mersenne prime, Fermat prime.
The first author was supported by VEGA grant 1/0485/09 and by grant APVV-51-009605. The second author was supported by VEGA grants 2/7139/27 and 1/0855/08.

**Example 1.3.** A simple way to generate (and print) all superprimes having at most $r$ digits is to use, within *Mathematica*, the following command:

```
p[0] = 2;p[1] = 3;p[2] = 5;p[3] = 7;
Do[number = 0; Do[m = n; q = 0; Do[z = Mod[m, 4]; m = Floor[m/4];
q = p[z] * 10^i + q, {i, 0, k - 1}]; If[PrimeQ[q], number++; Print[{k, q}]],
{n, 0, 4^k - 1}]; Print[number], {k, 1, r}]
```

Here are the four-digit superprimes obtained:
2237, 2273, 2333, 2357, 2377, 2557, 2753, 2777, 3253, 3257, 3323, 3373, 3527, 3533, 3557, 3727, 3733, 5227, 5233, 5237, 5273, 5323, 5333, 5527, 5557, 5573, 5737, 7237, 7253, 7333, 7523, 7537, 7573, 7577, 7723, 7727, 7753, 7757.

In the table below, $P_k$ is the number of $k$-digit superprimes for $1 \leq k \leq 15$. From this table one can conjecture that $P_k > 3^k$ for $k \geq 10$.

| $k$ | $P_k$ | $\sqrt[k]{P_k}$ | $k$ | $P_k$ | $\sqrt[k]{P_k}$ | $k$ | $P_k$ | $\sqrt[k]{P_k}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 4 | 4.000000000 | 6 | 389 | 2.701831538 | 11 | 214432 | 3.052549327 |
| 2 | 4 | 2.000000000 | 7 | 1325 | 2.792742150 | 12 | 781471 | 3.097961899 |
| 3 | 15 | 2.466212074 | 8 | 4643 | 2.873094002 | 13 | 2884201 | 3.139966685 |
| 4 | 38 | 2.482823796 | 9 | 16623 | 2.944202734 | 14 | 10687480 | 3.177331457 |
| 5 | 128 | 2.639015822 | 10 | 59241 | 3.000974037 | 15 | 39838489 | 3.211344203 |

Based on the computations above we now state the following two conjectures:

**Conjecture 1.** There is an infinite number of superprimes.

**Conjecture 2.** For any integer $k > 0$ there is a $k$-digit superprime.

**Remark 1.** We note that it would be interesting to find the limit $L := \lim_{k\to\infty} \sqrt[k]{P_k}$. It is likely that $L > 3$ and one cannot refute that $L = 4$. For the limit $L$ we have the asymptotic inequality $P_k > (L - \varepsilon)^k$ for every $\varepsilon > 0$.

**Remark 2.** We also note that in the base $m$ positional notation for $m \geq 4$ the situation seems to be analogous: the number $P_k^m$ denoting the number of $k$-digit superprimes in the base $m$ positional notation has been calculated for $4 \leq m \leq 12$ and it turns out that it grows roughly as $a^k$, where $a$ is slightly smaller than the number $\pi(m-1)$. (Here $\pi(x)$ is the prime-counting function, so $\pi(m-1)$ is the number of primes used in the base $m$ positional notation.) Hence Conjecture 1 and Conjecture 2 formulated above with respect to the decimal representation can analogously be formulated in the base $m$ positional notation for arbitrary $m \geq 4$.

## 2. Generalized Dirichlet Theorem

The well-known Dirichlet's Theorem on Primes in Arithmetic Progressions was first published in 1837. In his article [4], P.G.L. Dirichlet stated it as follows: *"each unlimited arithmetic progression, with the first member and the difference being coprime, will contain infinitely many primes."* We present it formally as a theorem below.

**Dirichlet's Theorem.** Assume that $a, b$ are coprime positive numbers. There is an infinite number of primes in the arithmetical sequence

$$a, a+b, a+2b, a+3b, \dots .$$

Our aim here is to conjecture a Generalized Dirichlet Theorem. Let us call by *Dirichlet sequence* the sequence in Dirichlet's Theorem. We shall consider the sequence $(x_n)_{n=0}^{\infty}$ defined by the recursive formula

$$x_{n+1} = a x_n + b \tag{1}$$

where $a$, $b$ and $x_0$ are integers such that $b$ is coprime to $a \cdot x_0$. We shall call it *Generalized Dirichlet sequence.* We note that one obtains Dirichlet sequence from it in the special case $a = 1$.

For our Generalized Dirichlet (GD) sequence we have the explicit formula

$$x_n = \begin{cases} a^n \left( x_0 + \frac{b}{a-1} \right) - \frac{b}{a-1} = a^n x_0 + \frac{b(a^n-1)}{a-1} & \text{for } a \neq 1, \\ x_0 + n\, b & \text{for } a = 1. \end{cases}$$

We note that the cases $a \in \{-1, 0\}$ are trivial and in case $b = -(a-1)x_0$ our sequence is constant. So, we shall assume $a \notin \{-1, 0, 1\}$ and $b \neq -(a-1)x_0$. Throughout this section we shall also need to consider as primes all members $x_n$ for which $|x_n|$ is prime - so even the negative integers. (This is not unusual, we note that also Mathematica treats prime numbers the same way and the commands PrimeQ[-3] or ProvablePrimeQ[-3] give answers 'True'). Such a consideration of prime numbers will help us to simplify the statements in this section and yet it will not negatively influence our main goal here which is to conjecture a Generalized Dirichlet Theorem.

**Remark 3.** If $a \neq -1$ a $b \neq -(a-1)x_0$, then the explicit formula above yields that all members of the GD sequence are different.

Also, the explicit formula above yields immediately the following statement.

**Proposition 2.1.** *Let $(x_n)_{n=0}^{\infty}$ be a GD sequence. If $a \cdot x_0$ and $b$ have a divisor $d > 1$, then all $x_n$ for $n \geq 1$ are divisible by $d$, and consequently, the GD sequence contains at most three primes.*

**Example 2.2.** Let us take in a GD sequence $a = 3$, $b = -12$ and $x_0 = 5$. Then the GD sequence $(5, 3, -3, -21, -75, \dots)$ contains only the primes 5, 3 and $-3$.

**Remark 4.** We note that also a partial converse to Proposition 2.1 is true: if two consecutive members $x_n$ and $x_{n+1}$ of the GD sequence have a common divisor $d > 1$, then $a \cdot x_0$ and $b$ have the divisor $d$, too.

**Example 2.3.** (i) Let us take in the GD sequence $x_0 = 3$, $a = 5$ and $b = 1$. Hence $x_{n+1} = 5x_n + 1$. Then $x_{n+2} = 25x_n + 6$, which means that all members $x_{2k}$ are divisible by 3. Since $x_1 = 16$, all members $x_{2k+1}$ are even and the sequence $(x_n)_{n=0}^{\infty}$ contains only the prime $x_0 = 3$.

(ii) Let us consider $x_0 = 14$, $a = 16$ and $b = 1$. We have $x_{n+1} = 16x_n + 1$. Then $x_{n+3} = 4096x_n + 273$. We note that $x_0 = 14 = 7 \cdot 2$, $x_1 = 225 = 3 \cdot 75$, $x_2 = 3601 = 13 \cdot 277$ and $273 = 3 \cdot 7 \cdot 13$.

It can easily be seen from the formulas above that all members $x_{3k}$ are divisible by 7, all $x_{3k+1}$ are divisible by 3 and all $x_{3k+2}$ are divisible by 13. The GD sequence $(x_n)_{n=0}^{\infty}$ does not contain any prime.

The previous examples show that even if $b$ is coprime to $a \cdot x_0$, the GD sequence $(x_n)_{n=0}^{\infty}$ can be partitioned into $k$ subsequences, of which each has its own nontrivial divisor, and so the sequence $(x_n)_{n=0}^{\infty}$ contains only a finite number of primes. Our aim is to study conditions forcing the GD sequence to contain only a finite number of primes.

Let us put

$$\begin{aligned} A_k &:= 1 + a + \dots a^{k-1} = \frac{a^k - 1}{a - 1} \text{ for } k \geq 0 \\ B_k &:= A_k b \\ y_n^{(k,j)} &:= x_{kn+j} \text{ for } k \geq 2 \text{ and } 0 \leq j \leq k-1 \end{aligned}$$

Obviously, $y_{n+1}^{(k,j)} = A_k y_n^{(k,j)} + B_k$ and $y_0^{(k,j)} = x_j$.

**Proposition 2.4.** *Assume that there exists $k \geq 2$ such that the following conditions hold:*

($a_k$) *for all $j \in \{0, \dots, k-1\}, A_k$ has a common divisor $d_j > 1$ with $x_j$.*

*Equivalently,*

($b_k$)

*for all $j \in \{0, \dots, k-1\}, A_k$ has a common divisor $d_j > 1$ with $x_0 - A_j b$.*

*Then the GD sequence $(x_n)_{n=0}^{\infty}$ contains only a finite number of primes.*

*Proof.* Assume that there exists $k \geq 2$ such that, for all $0 \leq j \leq k-1$, $A_k$ has a common divisor $d_j > 1$ with $x_j = y_0^{(k,j)}$. Then $d_j$ is a common divisor of $B_k = A_k b$ and $x_j$, which means that, for all $0 \leq j \leq k-1$, the sequence $(y_n^{(k,j)})_{n=0}^{\infty}$ contains a finite number of primes. Consequently, the GD sequence $(x_n)_{n=0}^{\infty}$ contains only a finite number of primes.

Now we show that $(a_k)$ is equivalent to $(b_k)$. If $j = 0$, then $x_j = x_0 - A_j b$, thus $(a_k)$ immediately implies $(b_k)$. For $1 \le j \le k-1$ we continue as follows. If, by $(a_k)$, $A_k$ has a common divisor $d_j > 1$ with $x_j$, it also has the common divisor $d_j$ with $x_j - A_k b = a^j x_0 + (A_j - A_k)b = a^j(x_0 - A_{k-j}b)$. Since $A_k$ is coprime to $a$, it has the common divisor $d_j$ with $(x_0 - A_{k-j}b)$. Hence $A_k$ has a common divisor with $x_0 - A_j b$ for all $1 \le j \le k-1$. Consequently, $(b_k)$ holds.

Conversely, to show that $(b_k)$ implies $(a_k)$, let for all $0 \le j \le k-1$, $A_k$ has a common divisor $d_j > 1$ with $x_0 - A_j b$. If $j = 0$, then $A_k$ has the common divisor $d_0$ with $x_0$. If $1 \le j \le k-1$, then $A_k$ has the common divisor $d_j$ with $a^{k-j}(x_0 - A_j b) = a^{k-j}x_0 - A_{k-j}b + A_k b$. This means that $A_k$ has the common divisor $d_j$ with $a^{k-j}x_0 - A_{k-j}b = x_{k-j}$. Hence $A_k$ has a common divisor with $x_j$ for all $j \in \{0, \ldots, k-1\}$. The proof is complete. □

**Remark 5.** We note that in the conditions $(a_k)$ and $(b_k)$ we could write $B_k$ instead of $A_k$. The equality $B_k = A_k b$ means that if $B_k$ has a common divisor $d_j > 1$ with $x_j$, but $A_k$ is coprime to $x_j$, then $b$ has a common divisor with $x_j$. Then $b$ has a common divisor with $ax_{j-1}$. Now if $b$ has a common divisor with $a$, we may apply Proposition 2.1. If $b$ is coprime to $a$, it has a common divisor with $x_{j-1}$ and then, by induction, it has a common divisor with $x_n$ for $n \ge 0$. So, these cases are in fact already covered by Proposition 2.1.

We also note that it follows from above that if we want the GD sequence to contain an infinite number of primes, then we have to guarantee that the conditions $(a_k)$ and $(b_k)$ are not satisfied for all $k \ge 2$, which is not a simple task. The most convenient way to guarantee it seems to be to show that $A_k$ is coprime to $x_0$ or to $x_1$ as we do it later with respect to Mersenne and Fermat primes (see Remark 8 and Remark 9, respectively). An alternative way is to show that $A_k$ is coprime to $x_0 - b = x_0 - A_1 b$, which is applied at the very end of Section 3.

**Remark 6.** (i) For $k = 2$ the condition $(a_k)$ above means that $a+1$ has common divisors with $x_0$ and $x_1$, and the equivalent condition $(b_k)$ means that $a+1$ has common divisors with $x_0$ and $x_0 - b$.

Our following observations concerning the conditions $(a_k)$ and $(b_k)$ for $k > 2$ are based on our computations via Mathematica and C++.

(ii) For $k = 3$ the conditions $(a_k)$ and $(b_k)$ are not satisfied for all $a$, $b$, $x_0$ with the integer values in the interval from $-15$ to 15 provided $a \cdot x_0$ is coprime to $b$.

(iii) For $k = 5$ the conditions $(a_k)$ and $(b_k)$ are not satisfied for all $a$, $b$, $x_0$ with the integer values in the interval from $-361$ to 361 provided $a \cdot x_0$ is coprime to $b$.

If 361 above is replaced with 1000, then there is exactly one example of $a$, $b$, $x_0$ with the given integer values where the conditions $(a_k)$ and $(b_k)$ are satisfied, namely $a = -139$, $b = 67$, $x_0 = 362$.

(iv) Similarly, for $k = 7$ the conditions $(a_k)$ and $(b_k)$ are not satisfied for all $a$, $b$, $x_0$ with the integer values in the interval from $-1500$ to $1500$ provided that $a \cdot x_0$ is coprime to $b$.

(v) An interesting situation occurs in case $k = 12$. The conditions $(a_{12})$ and $(b_{12})$ are satisfied for the values $a = -11$, $b = 7$, $x = -9$ (as well as for $a = 7$, $b = -23$, $x_0 = 25$), but for the same values the conditions $(a_4)$ and $(a_6)$ (as well as $(b_4)$ and $(b_6)$) are not satisfied. Hence the validity of the conditions is not transferred from $k$'s to their divisors.

The following statements are related to properties of (generalized Mersenne) numbers $\frac{a^n-1}{a-1}$ where $a \notin \{-1, 0, 1\}$, which can be primes only when $n$ is a prime. However, as we shall see, they can be primes for only a finite number of values $n$.

**Proposition 2.5.** *Let $c \notin \{-1, 0, 1\}$ and $m > 1$ be integers. Then $u_n = \frac{c^{mn}-1}{c^m-1}$ is not prime for $n > m$.*

*Proof.* The number $c^{mn} - 1$ is divisible by numbers $c^m - 1$ and $c^n - 1$, and so is divisible by their least common multiple which we denote by $M$. We can consider $M > 0$. Obviously, $M \geq |c^n - 1|$.

First we shall show that $|c^n - 1| > |c^m - 1|$ for $n > m$. We have the inequalities $|c^n - 1| \geq |c|^n - 1 \geq |c|^{m+1} - 1 \geq 2|c|^m - 1 = |c|^m + |c|^m - 1 \geq |c|^m + 3 > |c|^m + 1 \geq |c^m - 1|$. Hence $M > |c^m - 1|$. Further, $|c^{mn} - 1| \geq |c^{mn}| - 1 \geq |c^{2n}| - 1 = (|c^n| - 1)(|c^n| + 1) > |c^m - 1||c^n - 1| \geq M$. So we obtain a non-trivial factorization $\frac{c^{mn}-1}{c^m-1} = \frac{c^{mn}-1}{M} \frac{M}{c^m-1}$. □

**Proposition 2.6.** *Let the GD sequence satisfy the condition*

$$(c)\ a = c^{km}, b = \pm\frac{c^{km} - 1}{c^m - 1}, x_0 = \pm\frac{c^{jm} - 1}{c^m - 1}, \text{ with } c \notin \{-1, 0, 1\}, j \geq 0, k \geq 1, m \geq 2$$

*integer numbers and with a choice of the same signs $\pm$.*

*Then the GD sequence contains only a finite number of primes.*

*Proof.* W.l.o.g., let us choose the sign $+$. We have that $x_n$ is equal to

$$c^{kmn}\frac{c^{jm} - 1}{c^m - 1} + \frac{c^{kmn} - 1}{c^{km} - 1}\frac{c^{km} - 1}{c^m - 1} = \frac{c^{m(j+kn)} - c^{kmn}}{c^m - 1} + \frac{c^{kmn} - 1}{c^m - 1} = \frac{c^{m(j+kn)} - 1}{c^m - 1}.$$

If $j + kn > m$, then $x_n$ is not a prime by Proposition 2.5. □

Though $a = -4c^4$ is not a power of an integer, the next proposition shows that it behaves similarly to $a = c^m$, which is likely related to the fact that $-4c^4 = (1 + \mathrm{i})^4 c^4$.

**Proposition 2.7.** *Let $c \geq 1$ and $n \geq 1$. Then the integer $u_n = \frac{(-4c^4)^n-1}{-4c^4-1}$ is prime only for $c = 1$ and $n = 2$.*

*Proof.* We note that $u_0 = 0$, $u_1 = 1$ and $u_2 = 1 - 4c^4 = (1 - 2c^2)(1 + 2c^2)$. If $2 < n = 2k$, then

$$u_n = \frac{(-4c^4)^{2k} - 1}{-4c^4 - 1} = \frac{((-4c^4)^k - 1)((-4c^4)^k + 1)}{-4c^4 - 1} = \frac{(-4c^4)^k - 1}{-4c^4 - 1}((-4c^4)^k + 1)\,.$$

If $1 < n = 2k + 1$ then, with $x = 2^k c^{2k+1}$, we use the following identity due to Sophie Germain:

$$4x^4 + 1 = (2x^2 + 2x + 1)(2x^2 - 2x + 1).$$

We obtain

$$\begin{aligned} u_n &= \frac{\left(4c^4\right)^{2k+1} + 1}{4c^4 + 1} = \frac{4 \cdot (2^k c^{2k+1})^4 + 1}{4c^4 + 1} \\ &= \frac{(2 \cdot (2^k c^{2k+1})^2 + 2 \cdot 2^k c^{2k+1} + 1)(2 \cdot (2^k c^{2k+1})^2 - 2 \cdot 2^k c^{2k+1} + 1)}{4c^4 + 1}\,. \end{aligned}$$

□

**Proposition 2.8.** *Let the GD sequence satisfies the condition*

(*d*) $a = (-4c^4)^k, b = \pm\frac{(-4c^4)^k - 1}{-4c^4 - 1}, x_0 = \pm\frac{(-4c^4)^j - 1}{-4c^4 - 1}$ *with* $c \geq 1, j \geq 0, k \geq 1$ *integer numbers and with a choice of the same signs* $\pm$.

*Then the GD sequence contains at most one prime.*

*Proof.* W.l.o.g., let us choose the sign +. We have that $x_n$ is equal to

$$(-4c^4)^{kn}\frac{(-4c^4)^j - 1}{-4c^4 - 1} + \frac{(-4c^4)^{kn} - 1}{(-4c^4)^k - 1}\frac{(-4c^4)^k - 1}{-4c^4 - 1} = \frac{(-4c^4)^{(j+kn)} - 1}{-4c^4 - 1}.$$

If $j + kn \neq 2$, then $x_n$ is not a prime by Proposition 2.7. □

The next statements are related to the factorizations of $a^n - b^n$ or to the identity of Sophie Germain that we already used in the proof of Proposition 2.7.

**Proposition 2.9.** *Let* $c \notin \{-1, 0, 1\}$, $d \neq 0$, $\alpha \neq 0$ *and* $m > 1$ *be integers. Then the number* $u_n = \alpha^m c^{mn} - d^m$ *is prime for only a finite number of values* $n$.

*Proof.* The number $\alpha c^n - d$ is obviously a divisor of $u_n$. The equalities $\alpha c^n - d = \pm 1$ and $\alpha c^n - d = \pm u_n$ can be satisfied for only a finite number of values $n$. □

**Proposition 2.10.** *Let the GD sequence satisfies the condition*

(*e*) $a = c^m,\ b = \pm d^m(c^m - 1),\ x_0 = \pm(\alpha^m - d^m)$ *with* $d \neq 0, c \notin \{-1, 0, 1\}, \alpha \neq 0$ *integer numbers and a choice of the same signs* $\pm$.

*Then the GD sequence contains only a finite number of primes.*

*Proof.* W.l.o.g., let us choose the sign $+$. We have

$$x_n = c^{mn}(\alpha^m - d^m) + \frac{d^m(c^m-1)(c^{mn}-1)}{c^m-1} = \alpha^m c^{mn} - d^m.$$

□

**Remark 7.** If $c, d$ are not coprime, then $a, b$ are not coprime, too, so in the previous proposition we could add the condition that $c, d$ are not coprime.

**Proposition 2.11.** *Let $c \geq 2$, $d \geq 1$ and $\alpha \geq 1$ be integers. Then the numbers $u_n = \alpha^4 c^{4n} + 4d^4$ and $v_n = 4\alpha^4 c^{4n} + d^4$ are prime only for $d = 1$, $\alpha = 1$ and $n = 0$.*

*Proof.* Again, the identity due to Sophie Germain,

$$x^4 + 4y^4 = (x^2 + 2xy + 2y^2)(x^2 - 2xy + 2y^2) = \left((x+y)^2 + y^2\right)\left((x-y)^2 + y^2\right),$$

leads to the factorizations

$$u_n = \alpha^4 c^{4n} + 4d^4 = \left((\alpha c^n + d)^2 + d^2\right)\left((\alpha c^n - d)^2 + d^2\right)$$
$$v_n = 4\alpha^4 c^{4n} + d^4 = \left((d + \alpha c^n)^2 + c^{2n}\right)\left((d - \alpha c^n)^2 + \alpha^2 c^{2n}\right).$$

All factors are greater than 1 excepting the case $d = 1$, $\alpha = 1$ and $n = 0$. □

**Proposition 2.12.** *Let the GD sequence satisfies one of the the conditions*

*(f)*

$a = c^4,\ b = \pm 4d^4(1 - c^4),\ x_0 = \pm\left(\alpha^4 + 4d^4\right),$

*(g)*

$a = c^4,\ b = \pm d^4(1 - c^4),\ x_0 = \pm\left(4\alpha^4 + d^4\right),$

*with $c \geq 2$, $d \geq 1, \alpha \geq 1$ integer numbers and a choice of the same signs $\pm$.*

*Then the GD sequence contains at most one prime.*

*Proof.* Assuming that the condition (f) (the condition (g)) is satisfied, one obtains $x_n = \pm u_n$ ($x_n = \pm v_n$) from Proposition 2.11. □

Propositions 2.6, 2.8 2.10 and 2.12 give us another necessary conditions for the GD sequence to contain an infinite number of primes, namely that the integers $a$, $b$ and $x_0$ cannot have the values given in the conditions (c), (d), (e), (f) and (g). We are now ready to conjecture Generalized Dirichlet Theorem.

**Conjecture 3.** (**Generalized Dirichlet Theorem**) Let $a$, $b$ and $x_0$ be integers such that $b$ is coprime to $a \cdot x_0$. Consider the GD sequence $(x_n)_{n=0}^{\infty}$ defined by the recursive formula

$$x_{n+1} = ax_n + b$$

in which none of the following conditions is satisfied:

(i) the conditions ($a_k$) (equivalently ($b_k$)) from 2.4, for $k \geq 2$;

(ii) the condition (c) from 2.6;
(iii) the condition(d) from 2.8;
(iv) the condition(e) from 2.10;
(v) the conditions (f) and (g) from 2.12.

Then the GD sequence $(x_n)_{n=0}^{\infty}$ contains an infinite number of primes.

**Remark 8.** In the special case $x_0 = 0, a = 2, b = 1$ one obtains in the GD sequence $x_n = 2^n - 1$ which is a prime only if $n$ is a prime meaning $x_n$ is a *Mersenne prime*. We note that, for all $k \geq 2$, the condition $(a_k)$ is indeed not satisfied as $A_k = 2^k - 1$ and $x_1 = 1$ are coprime. None of the other conditions (c) - (g) is satisfied, too. Thus our Generalized Dirichlet Theorem implies a well-known conjecture saying that there is an infinite number of Mersenne primes.

**Remark 9.** In the special case $x_0 = 2, a = 2, b = -1$ we obtain the GD sequence $x_n = 2^n + 1$ which is a prime only if $n = 2^k$ meaning $x_n$ is a *Fermat prime*. Now the numbers $A_k = 2^k - 1$ and $x_0 = 2$ are coprime. Hence, similarly, as above, our Generalized Dirichlet Theorem implies a famous conjecture saying that there is an infinite number of Fermat primes.

**Remark 10.** In the case $a = 10$, $b = 1$, $x_0 = 0$, we have that $A_k$ is coprime to $x_1$, and we get the GD sequence with $x_n = (10^n - 1)/9 = 1 \cdots 1$, that is, with members $x_n$ consisting only of the digits 1 for $n \geq 1$. Generalized Dirichlet Theorem implies that there is an infinite number of primes whose decimal representation has only digits 1. Here Mathematica found the primes for

$$n = 2, 19, 23, 317 \text{ and } 1031.$$

More primes in this sequence have not been found.

**Example 2.13.** Let $a = 18$, $b = 1$ and $x_0 = 0$. Then $x_n = \frac{18^n - 1}{17}$. Using Mathematica, we have found out that $x_n$ is not prime for $2 < n \leq 25000$. However, we do not see any explanation for this fact.

Therefore also a weaker form of the conjecture seems to be interesting.

**Conjecture 4.** (**Weak Generalized Dirichlet Theorem**) There are integers $a \neq \pm 1$, $x_0$ and $b \neq (1 - a)x_0$ such that the GD sequence $(x_n)_{n=1}^{\infty}$ defined by

$$x_{n+1} = ax_n + b$$

contains an infinite number of primes.

## 3. Variations of the problem

There are several variations of the main problem regarding the infinite number of superprimes. For example, it looks as there is an infinite number of superprimes consisting only of arbitrary two fixed prime digits. Even a variation of Conjecture 2 saying that there is such specific $k$-digit superprime for any $k > 0$ seems to be true. We look more closely to superprimes consisting of the digits 2 and 3.

**Example 3.1.** To generate, via Mathematica, all superprimes having at most $r$ digits from the set $\{2, 3\}$, one can easily modify the command from Example 1.3. Here is the output obtained for $r = 8$: 2, 3, 23, 223, 233, 2333, 3323, 23333, 32233, 32323, 33223, 222323, 232333, 233323, 323233, 323333, 333233, 333323, 2222333, 2223233, 2232323, 2233223, 2332333, 2333323, 3222223, 3223223, 3223333, 3233323, 3233333, 3332233, 3333233, 22222223, 22223323, 22232233, 22232323, 23222233, 23223223, 23223323, 23322223, 32322223, 32323223, 32333333, 33222223, 33323333.

In the table below, $P_k$ denotes the number of $k$-digit primes of the considered type. One can ask what is the limit $L := \lim_{k\to\infty} \sqrt[k]{P_k}$. For $L$ we again have the asymptotic inequality $P_k > (L - \varepsilon)^k$ for every $\varepsilon > 0$.

| $k$ | $P_k$ | $\sqrt[k]{P_k}$ | $k$ | $P_k$ | $\sqrt[k]{P_k}$ | $k$ | $P_k$ | $\sqrt[k]{P_k}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 2.000000000 | 8 | 13 | 1.377980015 | 15 | 1337 | 1.615878716 |
| 2 | 1 | 1.000000000 | 9 | 39 | 1.502397860 | 16 | 1922 | 1.604111626 |
| 3 | 2 | 1.259921050 | 10 | 52 | 1.484568818 | 17 | 4549 | 1.641237856 |
| 4 | 2 | 1.189207115 | 11 | 104 | 1.525340028 | 18 | 7778 | 1.644975106 |
| 5 | 4 | 1.319507911 | 12 | 197 | 1.553121812 | 19 | 15926 | 1.664039040 |
| 6 | 7 | 1.383087554 | 13 | 382 | 1.579866021 | 20 | 25210 | 1.659887454 |
| 7 | 13 | 1.442562919 | 14 | 618 | 1.582545917 | 21 | 57882 | 1.685729112 |

**Example 3.2.** Here is the command and the output in Mathematica for all numbers $0 < n \leq 11000$ such that there is a superprime with the first digit 2 followed by $n$ digits 3:

$$\text{Do}[p = 2 * 10^n + (10^n - 1)/3; \text{If}[\text{PrimeQ}[q], \text{Print}[n]], \{n, 1, 11000\}];$$

$n = 1, 2, 3, 4, 10, 16, 22, 53, 91, 94, 106, 138, 210, 282, 522, 597, 1049, 2227, 6459, 10582.$

That is, the first five superprimes of this specific form are

$$23, 233, 2333, 23333, 23333333333.$$

Based on the above computation we state a stronger version of Problem 1 and the following two conjectures:

**Problem 2.** *Is there an infinite number of superprimes consisting of the digits* 2 *and* 3*?* More generally, *is there, for any pair of distinct prime digits, an infinite number of superprimes with only these two fixed digits?*

**Conjecture 5.** There is an infinite number of superprimes with the first digit 2 which is followed by $n$ digits 3 ($n > 0$). This is also true if the pair of prime digits $(2, 3)$ is replaced with any pair $(p, q)$ of distinct prime digits where $q \notin \{2, 5\}$.

**Conjecture 6.** For any integer $k > 0$ there is a $k$-digit superprime consisting only of the digits 2 and 3. This is also true if the prime digits $2, 3$ are replaced with any two distinct prime digits.

**Remark 11.** The sequence of numbers 2, 23, 233, 2333, ... from Example 3.2 can be obtained from our Generalized Dirichlet Theorem in the special case $a = 10$, $b = 3$ a $x_0$=2. In this case we have that $x_0 = 2$ is coprime to $A_k = 1 \cdots 1$.

Hence Generalized Dirichlet Theorem implies an affirmative answer to the the first part of Problem 2 and, of course, implies the first part of the Conjecture 5, too.

A stronger version of the main problem we consider here asks if there is an infinite number of superprimes with a stronger property that every subchain of the superprime's decimal representation consisting of the two subsequent digits is again a decimal representation of a prime number. For example, 373 is the first such superprime with 3 digits as both 37 and 73 are primes.

The following example indicates that there might be an infinite number of the superprimes having this stronger property. Let us call them *strong superprimes*. (We note that also strong superprimes can be considered in arbitrary base $m$ positional notation for $m \geq 4$.)

**Example 3.3.** To generate, via Mathematica, all strong superprimes having at most $r$ digits, one can again easily modify the command from Example 1.3. Here is the list of the first 7 strong superprimes with at least 3 digits:

$$373, 237373, 537373, 5373737, 53737373, 53737373737, 237373737373.$$

The output indicates that there are three types of the strong superprimes:

(i) Type A: 23 followed by $n$ copies of 73, the first one is 237373 ($n = 2$);
(ii) Type B: 53 followed by $n$ copies of 73, the first one is 537373 ($n = 2$);
(iii) Type C: 5 followed by $n$ copies of 37, the first one is 5373737 ($n = 3$).

We have generated, via Mathematica, the strong superprimes of the given three types with at most 2000 digits by a modification of the command from Example 3.2:

(i) Type A strong superprimes: $n = 2, 5, 20, 441$;
(ii) Type B strong superprimes: $n = 2, 3, 12, 21, 23, 483$;
(iii) Type C strong superprimes: $n = 3, 5, 8, 11, 15, 24, 53, 369, 710$.

**Conjecture 7.** There is an infinite number of strong superprimes of each of the three types A, B, C described above.

We again note that our Generalized Dirichlet Theorem in section 2 implies the Conjecture 7. To see this, let us put in Generalized Dirichlet Theorem in all three cases $a = 100$. For $b = 73$ and $x_0 = 23$ one obtains the type A, for $b = 73$ and $x_0 = 53$ the type B and for $b = 53$ and $x_0 = 5$ type C. In all three cases we have $A_1 = 1$ and $A_k = 10\ldots 101$ for $k \geq 2$. Since $23 - 73 = -50$, $53 - 73 = -20$ a $5 - 37 = -32$, the requirement that $x_0 - A_1 b$ is coprime to $A_k$ is satisfied in all three cases. Here one can see that sometimes checking the condition $(b_k)$ might be more convenient then checking $(a_k)$. In all three cases the conditions (c)-(g) are obviously not satisfied.

We finally consider the question whether there is an infinite number of superprimes with even a stronger property that all subchains of the superprime's decimal representation consisting of the two and three subsequent digits are again decimal representations of prime numbers. Here 373 is the first such superprime as all of 37, 73 and 373 are primes. It is obviously the only such strong superprime among the types A, B, C, because the number 737 is not prime. Hence strengthening further the concept of a strong superprime introduced here does not seem to be fruitful anymore.

**Remark 12.** We note that in the base $m$ positional notation for $m \geq 4$ the situation is quite different than in the above case $m = 10$. We have been searching (using simple modifications of the given commands in Mathematica) for strong superprimes in the base $m$ positional notation for $4 \leq m \leq 16$.

Just to illustrate our findings, we note that for $m = 5$ there are two 2-digit superprimes $13 = 23_5$ and $17 = 32_5$, one 3-digit strong superprime $67 = 232_5$ and one 5-digit strong superprime $2213 = 32323_5$. We have found a 17-digit strong superprime $1540415445963 = 32323232323232323_5$. For the base $m = 6$ there are two 2-digit superprimes $17 = 25_6$ and $23 = 35_6$; from this it can be easily shown that for $k > 2$ there are no $k$-digit strong superprimes in the base 6 positional notation. For $m = 8$ there are eight 2-digit superprimes: $19 = 23_8$, $23 = 27_8$, $29 = 35_8$, $31 = 37_8$, $43 = 53_8$, $47 = 57_8$, $59 = 73_8$ and $61 = 75_8$. Then also the number of $k$-digit strong superprimes is of course greater in the base $m$ positional notation for $m = 8$ than for $m = 10$.

## 4. Observations on sequences related to the problem

Our first observation in this section concerns the increasing sequence $(a_n)_{n=1}^{\infty}$ of all natural numbers (not necessarily superprimes) consisting only of the prime digits 2, 3, 5 and 7. This is the sequence

$$2, 3, 5, 7, 22, \cdots, 77, 222, \cdots, 777, 2222, \cdots, 7777, \cdots .$$

Assume that the $n$-th member $a_n$ consists of $k$ digits. Then

$$\frac{2}{9}\left(10^k - 1\right) \leq a_n \leq \frac{7}{9}\left(10^k - 1\right)$$

and

$$4 + 16 + \cdots + 4^{k-1} < n \leq 4 + 16 + \cdots + 4^k\,,$$

whence

$$\frac{4^k - 4}{3} < n \leq \frac{4^{(k+1)} - 4}{3}\,.$$

Thus

$$\frac{4^k - 1}{3} \leq n < \frac{4^{(k+1)} - 1}{3}$$

which yields

$$4^k \leq 3n + 1 < 4^{k+1}\,.$$

Therefore

$$k \leq \frac{\log(3n+1)}{\log 4} < k + 1$$

whence

$$k = \left[\frac{\log(3n+1)}{\log 4}\right]\,.$$

Consequently, for $k$ we have

$$\frac{\log(3n+1)}{\log 4} - 1 < k \leq \frac{\log(3n+1)}{\log 4}\,.$$

Hence we obtain the inequalities

$$\frac{2}{9}\left(10^{\frac{\log(3n+1)}{\log 4}-1} - 1\right) < a_n \leq \frac{7}{9}\left(10^{\frac{\log(3n+1)}{\log 4}} - 1\right)\,,$$

which can be rewritten as

$$\frac{2}{90}(3n+1)^{\log_4 10} - \frac{2}{9} < a_n \leq \frac{7}{9}(3n+1)^{\log_4 10} - \frac{7}{9}\,.$$

The last inequalities yield

$$\liminf_{n\to\infty} \frac{a_n}{n^{\log_4 10}} \geq \frac{2}{90}\,3^{\log_4 10}$$

and

$$\limsup_{n\to\infty} \frac{a_n}{n^{\log_4 10}} \geq \frac{7}{9}\,3^{\log_4 10}\,.$$

We conclude that the sequence $(a_n)_{n=1}^{\infty}$ behaves as $(n^a)_{n=1}^{\infty}$, where

$$a = \log_4 10 = 1.660964\ldots.$$

Our second observation concerns the sequence $(b_n)_{n=1}^{\infty}$ of all natural numbers, which are not primes. We first note the well-known Prime Number Theorem says that the number of primes among the first $n$ natural numbers is asymptotically $\frac{n}{\log n}$. This yields that the sequence $(b_n)_{n=1}^{\infty}$ is growing 'slowly'.

More precisely, let $\pi(n)$ denote the number of primes less than or equal to a natural number $n$ and let $p_n$ be the $n$-th prime. It is well-known that

$$\lim_{n\to\infty} \frac{\pi(n)\ln n}{n} = 1\,.$$

This implies that

$$\lim_{n\to\infty} \frac{p_n}{n\,\ln n} = 1\,.$$

Moreover, by [5],

$$p_n > n\,\ln n \qquad \text{for all natural numbers } n\,.$$

We shall show that

$$n\left(1+\frac{1}{\ln n+2}\right) < b_n < n\left(1+\frac{1}{\ln n-5}\right) \text{ for } n > \mathrm{e}^6\,,\ \text{i.e. } n \ge 404\,.$$

We note that the right inequality is an improvement of the asymptotic inequality

$$b_n < (1+\varepsilon)n \text{ for every } \varepsilon > 0.$$

On the other hand, the left inequality shows that the right inequality cannot be essentially improved.

For $n \ge 55$ we have (we refer to [5])

$$\frac{n}{\ln n+2} < \pi(n) < \frac{n}{\ln n-4}\,.$$

This is our starting point for the following observation. Let $n > \mathrm{e}^5 > 55$, hence $\ln n > 5$. Let us denote $m := b_n$. Obviously, $m > n$ and

$$n = m-\pi(m) > m\left(1-\frac{1}{\ln m-4}\right) > m\left(1-\frac{1}{\ln n-4}\right) = m\,\frac{\ln n-5}{\ln n-4}\,.$$

Therefore

$$m < n\,\frac{\ln n-4}{\ln n-5} = n\left(1+\frac{1}{\ln n-5}\right)\,.$$

From this it follows $\ln m < \ln n + \ln\left(1+\frac{1}{\ln n-5}\right) < \ln n + \frac{1}{\ln n-5}$ and

$$n = m-\pi(m) < m\left(1-\frac{1}{\ln m+2}\right) < m\left(1-\frac{1}{\ln n+2+\frac{1}{\ln n-5}}\right) = m\,\frac{\ln^2 n-4\ln n-4}{\ln^2 n-3\ln n-9}\,.$$

Hence

$$m > n\,\frac{\ln^2 n-3\ln n-9}{\ln^2 n-4\ln n-4} = n\left(1+\frac{\ln n-5}{\ln^2 n-4\ln n-4}\right)\,.$$

Under the condition that $\ln n \ge 6$ we have

$$\frac{\ln n-5}{\ln^2 n-4\ln n-4} \ge \frac{1}{\ln n+2}\,.$$

Thus
$$m > n\left(1+\frac{1}{\ln n+2}\right).$$
So we conclude that the growth of the sequence $(b_n)_{n=1}^{\infty}$ is comparable with the growth of the sequence of the natural numbers and yet it does not contain any prime. The sequence $(p_n)_{n=1}^{\infty}$ is growing a bit faster than the sequence of the natural numbers and yet it contains all (and only) primes. From this it follows that having a sequence of natural numbers, one cannot conclude anything about as whether it contains primes or not.

**Remark 13.** The inequalities above can even be slightly improved. We note that for $17 \le n < \mathrm{e}^{100}$ as well as for $n > \mathrm{e}^{200}$ we have (we refer again to [5])
$$\frac{n}{\ln n} < \pi(n) < \frac{n}{\ln n - 2}.$$
From this one can analogously as above derive
$$n\left(1+\frac{1}{\ln n}\right) < a_n < n\left(1+\frac{1}{\ln n-3}\right) \text{ for } n > \mathrm{e}^{200}.$$

## References


1. M. Aigner and G.M. Ziegler, *Proofs from THE BOOK*, Berlin-Heidelberg: Springer-Verlag, 1998.
2. F. Borneman, *PRIMES is in P: a breakthrough for "Everyman"*, Notices of AMS **50** (2003), pp. 545–552.
3. J. Derbyshire, *Prime Obsession*, Washington, DC: Joseph Henry Press, 2003.
4. P.G.L. Dirichlet, *There are infinitely many prime numbers in all arithmetic progressions with first term and difference coprime*, Abhandlungen der Königlich Preussischen Akademie der Wissenschaften von 1837, (1837), 45–81. (An English translation from German by R. Stephan available at http://arxiv.org/PS_cache/arxiv/pdf/0808/0808.1408v1.pdf.)
5. J.B. Rosser, *The n-th prime is greater than* $n \log n$, Proc. London Math. Soc. (2) **45** (1939), 21–44.
6. J.B. Rosser, *Explicit bounds for some functions of prime numbers*, Amer. J. Math. **63** (1941), 211–232.
7. J.B. Rosser and L. Schoenfeld, *Approximate Formulas for Some Functions of Prime Numbers*, Illinois Journal Math. **6** (1962), 64–94.
8. M. Sautoy, *The Music of the Primes*, Harper Collins, 2003.



Department of Mathematics, Matej Bel University, Tajovského 40, 974 01 Banská Bystrica, Slovak Republic
*E-mail address*: mhaviar@fpv.umb.sk

Department of Mathematics, Matej Bel University, Tajovského 40, 974 01 Banská Bystrica, Slovak Republic
*E-mail address*: malicky@fpv.umb.sk