

DELETING DIGITS

IOULIA N. BAOULINA, MARTIN KREH, JÖRN STEUDING

1. A PAPER BY SHALLIT ON PRIMES

We consider here the positive integers with respect to their unique decimal expansions, where each $n \in \mathbb{N}$ is given by $n = \sum_{j=0}^k \alpha_j 10^j$ for some non-negative integer k and digit sequence $\alpha_k \alpha_{k-1} \dots \alpha_0$. With slight abuse of notation, we also use n to denote $\alpha_k \alpha_{k-1} \dots \alpha_0$. For such sequences of digits (as well as for the numbers represented by the corresponding expansions) we write $x \triangleleft y$ if x is a subsequence of y , which means that either $x = y$ or x can be obtained from y by deleting some digits of y . For example, $514 \triangleleft 352148$. The main problem is as follows: *given a set $S \subset \mathbb{N}$, find the smallest possible set $M \subset S$ such that for all $s \in S$ there exists $m \in M$ with $m \triangleleft s$* . As a matter of fact, the set

$$M(S) := \{s \in S \mid \{n \in S : n < s, n \triangleleft s\} = \emptyset\}$$

solves this problem; clearly, every element of S contains some element of $M(S)$ as a subword. The set $M(S)$ is said to be the *minimal set* and its elements are also called *minimal*.

Remarkably, $M(S)$ is finite for every possible $S \subset \mathbb{N}$. This follows from a celebrated theorem from the theory of formal languages, the so-called lemma of Higman [1] (see also the Chapter entitled *Subwords* by Jacques Sakarovitch and Imre Simon [2] in Lothaire's encyclopaedia *Combinatorics on Words*).

In general, it seems to be difficult to compute $M(S)$ for a given S , the known proofs of Higman's lemma are ineffective. In the case of the set of prime numbers $\mathbb{P} = \{2, 3, 5, \dots\}$ Jeffrey Shallit [3] succeeded in determining $M(\mathbb{P})$ by elementary means, namely

$$\begin{aligned} M(\mathbb{P}) = \{ & 2, 3, 5, 7, 11, 19, 41, 61, 89, 409, 449, 499, 881, 991, 6469, \\ & 6949, 9001, 9049, 9649, 9949, 60649, 666649, 946669, \\ & 60000049, 66000049, 66600049 \}. \end{aligned}$$

Already by a first glimpse on the elements of $M(\mathbb{P})$ one can guess the kind of reasoning necessary to prove this explicit form for the minimal set of primes. However, in the case of the set $2^{\mathbb{N}} = \{2^n : n \in \mathbb{N}_0\}$ consisting of the powers of 2 Shallit [3] conjectured

$$M(2^{\mathbb{N}}) = \{1, 2, 4, 8, 65536\}.$$

Moreover, he observed that this is true if every number 16^m with $m \geq 4$ has at least one digit from $\{1, 2, 4, 8\}$. This type of problem seems to be difficult to tackle. Of course, we know that the powers of 2 are distributed according to Benford's law (i.e., the sequence $n \log_{10} 2$ is uniformly distributed modulo one), nevertheless, this just indicates - in an appropriate probabilistic framework - that a power of 2 without any digit from $\{1, 2, 4, 8\}$ must be a very rare event.

In this note we present our new results on the minimal sets of a few other arithmetically interesting sets of positive integers (in Section 2 & 3), we furthermore address questions on size and shape of minimal sets in general and indicate the absence of structure behind (Section 4), and finally we conclude with another example, namely the hypothetical minimal set for perfect numbers and analogous question for other bases (Section 5). Our main aim is to make the beautiful theorem of Higman and Shallit's paper more popular in order to attract more research in this direction.

2. SUMS OF SQUARES AND QUADRATIC RESIDUES

In this section we look at three subsets of \mathbb{N} for which the minimal sets can be determined easily. Whereas in the first example the reason for this is that the examined set contains many digits, in the second and third examples we exploit the fact that the elements of the sets are in some sense well-distributed. For the first example, this means that we have almost all digits lying in the set, which means that there are just few numbers that we are left to examine. In the other two examples, the elements of the examined sets lie in certain residue classes for some modulus, which will be helpful. Other explicit constructions of minimal sets of residue classes and some discussions about the structure can be found in [4].

Theorem 1. *Let*

$$\boxed{3} := \{n \in \mathbb{N} \mid \exists x, y, z \in \mathbb{N}_0 : n = x^2 + y^2 + z^2\}.$$

Then

$$M(\boxed{3}) = \{1, 2, 3, 4, 5, 6, 8, 9, 70, 77\}.$$

Proof. It is easy to see that each of the numbers 1, 2, 3, 4, 5, 6, 8, 9, 70, 77 can be written as a sum of three squares, so they belong to $\boxed{3}$. Clearly 1, 2, 3, 4, 5, 6, 8, 9 belong to $M(\boxed{3})$. Since $7 \notin \boxed{3}$, 70 and 77 also belong to $M(\boxed{3})$. If $n \in \boxed{3}$ is different from 1, 2, 3, 4, 5, 6, 8, 9, 70, 77, then either $d \triangleleft n$ for $d \in \{1, 2, 3, 4, 5, 6, 8, 9\}$ and n cannot belong to $M(\boxed{3})$ or n consists only of digits 0 and 7. In the latter case either $70 \triangleleft n$ or $77 \triangleleft n$, and so $n \notin M(\boxed{3})$. \square

Two strings of digits x and y are called *incomparable* if neither $x \triangleleft y$ nor $y \triangleleft x$. The reader might have noticed that a subset of S consisting of pairwise incomparable elements is in general not contained in $M(S)$ (as the example $S = \mathbb{N}$ and the subset of all two digit numbers shows). However, in the previous proof for any element x in the set $\{1, 2, 3, 4, 5, 6, 8, 9, 70, 77\}$ of pairwise incomparable elements, there was no $y \in \boxed{3}, y \neq x$ with $y \triangleleft x$. In this case the

set of pairwise incomparable elements is indeed contained in $M(S)$. We shall use this idea quite often in the sequel.

Theorem 2. *Let*

$$\square \bmod 6 := \{n \in \mathbb{N} \mid \exists x \in \mathbb{N} : x^2 \equiv n \bmod 6\}.$$

Then

$$M(\square \bmod 6) = \{1, 3, 4, 6, 7, 9, 22, 25, 28, 52, 55, 58, 82, 85, 88\}.$$

Proof. It is easy to check that $n \in \square \bmod 6$ if and only if $n \equiv 1, 3, 4, 6 \bmod 6$. So we have indeed $\{1, 3, 4, 6, 7, 9, 22, 25, 28, 52, 55, 58, 82, 85, 88\} \subset \square \bmod 6$. Moreover, these elements are pairwise incomparable. Let $n \in \square \bmod 6$ be arbitrary. If $d \triangleleft n$ with $d \in \{1, 3, 4, 6, 7, 9\}$, we are done. So suppose that n contains none of these digits. Assume first that n contains exactly two digits. Then $22, 25, 28, 52, 55, 58, 82, 85, 88 \in \square \bmod 6$ and $20, 50, 80 \notin \square \bmod 6$. If n has more than two digits, then either two of its digits are nonzero and so there is an $x \in M(\square \bmod 6)$ with $x < n$, $x \triangleleft n$, or n has only one nonzero digit. In the latter case, this digit is congruent to 2 modulo 3, hence $n \equiv 2 \bmod 6$, and therefore $n \notin \square \bmod 6$. \square

Theorem 3. *Let*

$$\square \bmod 7 := \{n \in \mathbb{N} \mid \exists x \in \mathbb{N} : x^2 \equiv n \bmod 7\}.$$

Then

$$M(\square \bmod 7) = \{1, 2, 4, 7, 8, 9, 30, 35, 36, 50, 53, 56, 60, 63, 65, 333, 555, 666\}.$$

Proof. Since $n \in \square \bmod 7$ if and only if $n \equiv 1, 2, 4, 7 \bmod 7$, we have $\{1, 2, 4, 7, 8, 9, 30, 35, 36, 50, 53, 56, 60, 63, 65, 333, 555, 666\} \subset \square \bmod 7$; again these elements are pairwise incomparable. Let $n \in \square \bmod 7$. If $d \triangleleft n$ with $d \in \{1, 2, 4, 7, 8, 9\}$, we are done. So suppose that n contains none of these digits. If n has exactly two digits, then $30, 35, 36, 50, 53, 56, 60, 63, 65 \in \square \bmod 7$ and $33, 55, 66 \notin \square \bmod 7$. If n has at least three digits and at least two of them are distinct we are done. It therefore remains to assume that n is of the form $dd \dots dd$ with $d \in \{3, 5, 6\}$. Then either $333 \triangleleft n$ or $555 \triangleleft n$ or $666 \triangleleft n$. \square

3. VALUES OF ARITHMETICAL FUNCTIONS

In this section, we consider sets of the type $S = \{n \in \mathbb{N} \mid \exists x \in \mathbb{N} : f(x) = n\}$ where $f : \mathbb{N} \rightarrow \mathbb{N}$ is an arithmetical function. Recall that the Euler function $\varphi(n)$ is defined to be the number of positive integers not exceeding n which are relatively prime to n . If $n = \prod_{p|n} p^{k_p}$ is the unique prime factorization of n , then $\varphi(n) = \prod_{p|n} (p-1)p^{k_p-1}$.

Theorem 4. *Let*

$$\varphi(\mathbb{N}) := \{n \in \mathbb{N} \mid \exists x \in \mathbb{N} : \varphi(x) = n\}.$$

Then

$$M(\varphi(\mathbb{N})) = \{1, 2, 4, 6, 8, 30, 70, 500, 900, 990, 5590, 9550, 555555555550\}.$$

Proof. Observe that the numbers above are pairwise incomparable. Further, as

$$\begin{aligned}\varphi(1) &= 1, & \varphi(3) &= 2, & \varphi(5) &= 4, & \varphi(7) &= 6, \\ \varphi(16) &= 8, & \varphi(31) &= 30, & \varphi(71) &= 70, & \varphi(625) &= 500, \\ \varphi(1057) &= 900, & \varphi(991) &= 990, & \varphi(5591) &= 5590, & \varphi(9551) &= 9550,\end{aligned}$$

and

$$\varphi(555555555551) = 555555555550,$$

we see that $\{1, 2, 4, 6, 8, 30, 70, 500, 900, 990, 5590, 9550, 555555555550\} \subset \varphi(\mathbb{N})$. Note also that $3, 5, 7, 9 \notin \varphi(\mathbb{N})$. Now assume that $n \in \varphi(\mathbb{N})$ has at least two digits. If there exists $d \in \{1, 2, 4, 6, 8\}$ such that $d \triangleleft n$ then $n \notin M(\varphi(\mathbb{N}))$. Suppose that n contains only the digits 0, 3, 5, 7, 9. Since $\varphi(m)$ is even for $m > 2$, the last digit is 0. It is easy to see that $50, 90 \notin \varphi(\mathbb{N})$. Now assume that n has at least three digits. If n contains the digits 3 or 7, then $30 \triangleleft n$ or $70 \triangleleft n$, respectively. Consequently, $n \notin M(\varphi(\mathbb{N}))$. Next assume that n contains only the digits 0, 5, 9. It is easily verified that $550, 590, 950 \notin \varphi(\mathbb{N})$. Suppose that n has at least four digits. Then we have the following possibilities:

- n contains at least two zeros. Then either $500 \triangleleft n$ or $900 \triangleleft n$.
- $99 \triangleleft n$. Then $990 \triangleleft n$.
- $559 \triangleleft n$. Then $5590 \triangleleft n$.
- $955 \triangleleft n$. Then $9550 \triangleleft n$.
- $n = 5950$. Then $n \notin \varphi(\mathbb{N})$.
- $n = \underbrace{55 \dots 5}_\ell 0$ with $\ell \geq 3$.

Therefore, it remains to show that if $n = \underbrace{55 \dots 5}_\ell 0$ with $3 \leq \ell \leq 10$, then $n \notin \varphi(\mathbb{N})$. To this end, assume that $\varphi(m) = n$ for some $m \in \mathbb{N}$. Since $4 \nmid n$, we have $4 \nmid m$ and m has exactly one odd prime divisor, that is $m = p^k$ or $m = 2p^k$ with $k \in \mathbb{N}$ and odd $p \in \mathbb{P}$. In both cases we must have $p^{k-1}(p-1) = n$. Using a computer algebra package, we find the unique prime factorization of each of the values of n under consideration, namely,

$$\begin{aligned}5550 &= 2 \cdot 3 \cdot 5^2 \cdot 37, & 55550 &= 2 \cdot 5^2 \cdot 11 \cdot 101, \\ 555550 &= 2 \cdot 5^2 \cdot 41 \cdot 271, & 5555550 &= 2 \cdot 3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 37, \\ 55555550 &= 2 \cdot 5^2 \cdot 239 \cdot 4649, & 555555550 &= 2 \cdot 5^2 \cdot 11 \cdot 73 \cdot 101 \cdot 137, \\ 5555555550 &= 2 \cdot 3^2 \cdot 5^2 \cdot 37 \cdot 333667, & 55555555550 &= 2 \cdot 5^2 \cdot 11 \cdot 41 \cdot 271 \cdot 9091.\end{aligned}$$

We see that if $p^{k-1} \mid n$ with $k \geq 3$, then $k = 3$ and $p = 3$ or 5 , and so $p^{k-1}(p-1) \neq n$. Further, it is easy to check that there is no prime number p with $p(p-1) = n$. Hence $k = 1$ and $n+1$ is a prime. However, all the numbers $\underbrace{55 \dots 5}_\ell 1$, $3 \leq \ell \leq 10$, are composite, and this gives the desired contradiction. \square

It is not difficult to find minimal sets for some sets of shifted values of the Euler function. For example, it can be proved that

$$M(3 + \varphi(\mathbb{N})) = \{4, 5, 7, 9, 11, 13, 21, 23, 31, 33, 61, 63, 81, 83\},$$

where

$$3 + \varphi(\mathbb{N}) := \{n \in \mathbb{N} \mid \exists x \in \mathbb{N} : 3 + \varphi(x) = n\}.$$

We leave this as an exercise for the interested reader.

Related to Euler's function but less well-known is the Dedekind ψ -function, for $n = \prod_{p|n} p^{k_p}$ defined by

$$\psi(n) := \prod_{p|n} (p+1)p^{k_p-1}.$$

As we shall see in the proof of the following theorem, the determination of the corresponding minimal set is possible, although requires a little more effort than in the case of Euler's φ .

Theorem 5. *Let ψ be the Dedekind ψ -function and*

$$\psi(\mathbb{N}) := \{n \in \mathbb{N} \mid \exists x \in \mathbb{N} : \psi(x) = n\}.$$

Then

$$M(\psi(\mathbb{N})) = \{1, 3, 4, 6, 8, 20, 72, 90, 222, 252, 500, 522, 552, 570, 592, 750, 770, \\ 992, 7000, \underbrace{55 \dots 5}_{69} 0\}.$$

Proof. First note, that the elements given above are pairwise incomparable and we have

$$\begin{array}{llll} \psi(1) & = & 1, & \psi(2) & = & 3, & \psi(3) & = & 4, & \psi(5) & = & 6, \\ \psi(7) & = & 8, & \psi(19) & = & 20, & \psi(71) & = & 72, & \psi(89) & = & 90, \\ \psi(146) & = & 222, & \psi(251) & = & 252, & \psi(499) & = & 500, & \psi(521) & = & 522, \\ \psi(411) & = & 552, & \psi(569) & = & 570, & \psi(511) & = & 592, & \psi(625) & = & 750, \\ \psi(769) & = & 770, & \psi(991) & = & 992, & \psi(6631) & = & 7000 \end{array}$$

and

$$\psi(\underbrace{55 \dots 5}_{68} 49) = \underbrace{55 \dots 5}_{69} 0,$$

since $\underbrace{55 \dots 5}_{68} 49$ is a prime, all these numbers belong to $\psi(\mathbb{N})$.

Now let $n \in \psi(\mathbb{N})$ be arbitrary with at least two digits. If $d \triangleleft n$ with $d \in \{1, 3, 4, 6, 8\}$, then $n \notin M(\psi(\mathbb{N}))$. So n does only contain the digits 2, 5, 7, 9, 0. It follows from the product formula $\psi(m) = \prod_{p|m} (p+1)p^{k_p-1}$ defining ψ that $\psi(m)$ is even for $m > 2$. First we note that 22, 50, 52, 70, 92, 292, 502, 550, 700, 922, 952 $\notin \psi(\mathbb{N})$. Now suppose that n has at least four digits. Then one of the following cases holds:

- n ends with a 2 and contains a 7. Then $72 \triangleleft n$.

- n ends with a 2, contains a zero and does not contain a 7. Then $20 \triangleleft n$ or $90 \triangleleft n$ or $502 \triangleleft n$. In the latter case n contains one of the following strings:

5002, 2502, 5202, 5022, 5502, 5052, 9502, 5902, 5092.

Then $20 \triangleleft n$ or $90 \triangleleft n$ or $500 \triangleleft n$ or $522 \triangleleft n$ or $552 \triangleleft n$ or $592 \triangleleft n$.

- n ends with a 2 and contains only the digits 2, 5 and 9. Then $222 \triangleleft n$ or $252 \triangleleft n$ or $522 \triangleleft n$ or $552 \triangleleft n$ or $992 \triangleleft n$.
- n ends with a zero and contains a 2 or a 9. Then $20 \triangleleft n$ or $90 \triangleleft n$.
- n ends with a zero and contains a 7 and a 5. Then $570 \triangleleft n$ or $750 \triangleleft n$.
- n ends with a zero and contains only the digits 7 and 0. Then $770 \triangleleft n$ or $7000 \triangleleft n$.
- n ends with a zero and contains only the digits 5 and 0. Then $500 \triangleleft n$ or $550 \triangleleft n$.

Thus, the only numbers n that need to be examined are of the form $5 \dots 50$. It suffices to show that if $n = \underbrace{55 \dots 5}_\ell 0$ with $3 \leq \ell \leq 68$, then $n \notin \psi(\mathbb{N})$.

We note, that if m has k distinct odd prime factors, then $2^k \mid \psi(m)$. In our case, $n \equiv 2 \pmod{4}$, so if there is $m \in \mathbb{N}$ with $\psi(m) = n$, then $m = p^k$ or $m = 2p^k$ with $p \in \mathbb{P}$ odd. Hence either $p^{k-1}(p+1) = n$ or $3p^{k-1}(p+1) = n$ with $k \in \mathbb{N}$ and odd $p \in \mathbb{P}$. Note that $4 \nmid (p+1)$. A computer search shows that if $p^{k-1} \mid n$ with $k \geq 3$ and $4 \nmid (p+1)$, then $k = 3$ and $p = 5$. This implies that either $n = 150$ or $n = 450$, which is a contradiction. Next, one can easily verify that there is no prime p with $n = p(p+1)$ or $3p(p+1)$, and so $k \neq 2$. Finally, note that neither $n-1$ nor $(n/3)-1$ is a prime, which yields $k \neq 1$. This concludes the proof. \square

4. THE ABSENCE OF STRUCTURE

The examples from the previous sections already indicate that minimal sets behave rather unexpectedly. It is easy to obtain some sets in explicit form, while for other sets of *higher arithmetical complexity* the explicit form of the corresponding minimal set may be only achieved conditionally. Already the size of minimal sets seems to be almost unpredictable.

For $k \in \mathbb{N}$ define $S_k := \mathbb{N} \cap [10^{k-1}, 10^k)$ (the set of positive integers with k digits in base 10). One easily verifies $M(S_k) = S_k$ which shows that the minimal set can be as large as we please even for finite sets. Moreover, S_k is minimal among all sets S with $\sharp M(S) = 9 \cdot 10^{k-1}$. In general it seems to be difficult to prove an upper bound for the number of minimal elements in an arbitrary set.

One might be tempted to try to build up some kind of set theory for minimal sets. The following statement is trivial:

$$M(S \cup T) \subset M(S) \cup M(T).$$

The left-hand side above can be equal to the right-hand side (if, for example, $S = 2 + 10\mathbb{N}_0$ and $T = 3 + 10\mathbb{N}_0$) and smaller (if, for example, $S = \{2\} \cup (\mathbb{P} \cap (1 + 4\mathbb{N}_0))$ and $T = \mathbb{P} \cap (3 + 4\mathbb{N}_0)$). The same inclusion with \cap in place of \cup on the right hand-side is in general false as follows from the example $S = \mathbb{P}$ and $T = 1 + 4\mathbb{N}_0$. Moreover, even $M(S \cap T) \subset M(S) \cup M(T)$ is false as the counterexample $S = 7 + 10\mathbb{N}$ and $T = \mathbb{P}$ shows.

Probably, the main obstacle in order to prove something 'structural' (whatever that means) is that $S \mapsto M(S)$ is not monotone: e.g., $\mathbb{P} \subset \mathbb{N}$ but $\#M(\mathbb{P}) > \#M(\mathbb{N})$. We conclude that minimal sets indeed behave erratically (but would be glad if the reader could disprove our non-mathematical statement).

5. AN ODD END

Choosing another base than 10 would not make things easier in general (see [5] for recent results in this direction). Of course, with the binary expansion in place of the decimal expansion a few minimal sets would look more simple, e.g., $M(\mathbb{P}) = \{10_2, 11_2\}$. However, there are still plenty of sets, interesting from a number-theoretical point of view, where the corresponding minimal set seems to be difficult to describe.

For instance, the set of perfect numbers $\text{Perfect} = \{6, 28, 496, \dots\}$ is conjectured to contain no odd numbers. Recall that a positive integer is called *perfect* if it equals the sum of its proper divisors. As already Euclid knew (in different language), every Mersenne prime $2^p - 1$ (with prime p) yields an even perfect number $2^{p-1}(2^p - 1)$, and Leonhard Euler proved that every even perfect number is of this form (see Dickson [6] for this and further details about and the history of perfect numbers). Édouard Lucas [7] showed that every even perfect number different from 6 and 496 ends with decimal digits 16, 28, 36, 56 or 76. Hence, *if there is no odd perfect number, then the minimal set of perfect numbers is given by $M(\text{Perfect}) = \{6, 28\}$* . If there exists an odd perfect number, although much can be said about its hypothetical multiplicative structure, we cannot exclude the possibility that this odd number is a minimal element. In base 2 we would have $M(\text{Perfect}) = \{110_2\}$ if there is no odd perfect number, and we leave the computation of the minimal set with respect to other bases (under the same assumption) to the interested reader.

ACKNOWLEDGMENT

The authors thank the referee for careful reading of the manuscript and many helpful suggestions.

REFERENCES

- [1] G. HIGMAN, Ordering by divisibility in abstract algebras, *Proc. London Math. Soc.* **2** (1952), pp. 326-336

- [2] J. SAKAROVITCH, I. SIMON, Subwords, in: *Combinatorics on Words*, edited by M. LOTHAIRE, Encyclopedia of Mathematics and its Applications, Vol. 17, Addison-Wesley, 1983.
- [3] J. O. SHALLIT, Minimal Primes, *J. Recreational Math.* **30** (2000), pp. 113-117.
- [4] M. KREH, Minimal Sets, *J. Integer Seq.* **18** (2015), Article 15.5.3.
- [5] C. BRIGHT, R. DEVILLERS and J. SHALLIT, Minimal elements for the prime numbers, *Exp. Math.* **25** (2016) pp. 321-331.
- [6] L.E. DICKSON, *History of Number Theory. Vol. I*, Carnegie Institution of Washington, 1919.
- [7] É. LUCAS, Sur les nombres parfaits, *Mathesis* **10** (1890), pp. 74-76.

Ioulia N. Baoulina, Department of Mathematics, Moscow State Pedagogical University, Krasnoprudnaya str. 14, Moscow 107140, Russia, jbaulina@mail.ru

Martin Kreh, Department for Algebra and Number Theory, University of Hildesheim, Samelsonplatz 1, 31141 Hildesheim, Germany, kreh@imai.uni-hildesheim.de

Jörn Steuding, Department of Mathematics, Würzburg University, Emil-Fischer-Str. 40, 97 074 Würzburg, Germany, steuding@mathematik.uni-wuerzburg.de