THE KERNEL OF THE PRIMES FOR NUMBER BASES TWO THROUGH TEN

by

R. Michael Perry
University of Colorado
Boulder, Colorado

CU-CS-138-78 May, 1978

Abstract. The kernels of the languages representing the set of prime numbers in bases two through ten are derived and tabulated. In each case it is found that the number of kernel entries and the maximum length of an entry are of roughly the same size as the number base.

The kernel of a formal language L is the set of words $w \in L$ such that no proper subword (i.e. no proper subsequence) of w is a word of L. It has been shown [2] that every language has a finite kernel. Here we derive and tabulate the kernels of the languages consisting of the representations of the set of prime numbers in number bases two through ten. Results are shown in table 1. w_n^m denotes the nth word, ordered by increasing numerical value, of the m-kernel, i.e., the kernel for base m. v_n^m is the numerical value of w_n^m, here expressed in base ten. S_m and L_m are respectively the number of words and the maximum length of a word, in the m-kernel.

Inspection of table 1 shows that, while the values of the v_n^m are rather irregularly distributed over several orders of magnitude, the values of S_m and L_m are reasonably well-behaved, being of the same order of size as m. It is evident that $S_m \to \infty$ as $m \to \infty$ since S_m must be at least as large as the number of primes less than m. Other properties which seem to hold but whose proofs (if they exist) do not appear simple are that $L_m \to \infty$ as $m \to \infty$ and that L_m is always $\leq S_m$.

It is evident that for given integers m, k > 1, we can decide if k is the value of an element in the m-kernel, so that the set of all pairs (m, w_n^m) for $m \geq 2$, $w_n^m \in$ m-kernel, is recursively enumerable (where, for formal purposes, we may regard m as a word in base ten notation). To establish a given m-kernel, however, we need to be able to decide whether a given set of w_n^m values includes the entire kernel or not. It is reasonable to ask whether such a property is decidable. Although we do not answer this question here, for the values $2 \leq m \leq 10$ we are able, using *ad hoc* considerations, to establish the m-kernel in each case.

We write $v \to w$ to mean that w is a proper subword of v (thus if $w \in L$ then v cannot be in the kernel of L). When v is understood it may be omitted. We use the standard subscript notation for values in different number bases; e.g. $111011_2 = 73_8 = 59_{10}$. All subscripts are in base 10 and are omitted when the meaning is clear. Superscripts are normally used to indicate repetitions of digits; e.g. $10^31 = 10001$. However, in arithmetic expressions involving the operations +, -, or /, the superscripts indicate exponentiation and all numbers are in base ten. An expression x(y) indicates that x and y are in base ten and that y is the smallest prime factor; thus 81(3).

Table 1: The M-kernel for $2 \le m \le 10$

Key: S_m = size or number of entries in the kernel, L_m = maximum length of an entry, w_n^m = an individual entry ($1 \le n \le S_m$), v_n^m = numerical value of w_n^m in base ten.

$m = 2$	n	w_n^m	v_n^m
$S_m = 2$			
$L_m = 2$	1	10	2
	2	11	3

$m = 3$	n	w_n^m	v_n^m
$S_m = 3$			
$L_m = 3$	1	2	2
	2	10	3
	3	111	13

$m = 4$	n	w_n^m	v_n^m
$S_m = 3$			
$L_m = 2$	1	2	2
	2	3	3
	3	11	5

$m = 5$	n	w_n^m	v_n^m
$S_m = 8$			
$L_m = 5$	1	2	2
$m = 5$ (cont'd)	2	3	3
	3	10	5
	4	111	31
	5	401	101
	6	414	109
	7	14444	1,249
	8	44441	3,121

$m = 6$	n	w_n^m	v_n^m
$S_m = 7$			
$L_m = 5$	1	2	2
	2	3	3
	3	5	5
	4	11	7
	5	4401	1,009
	6	4441	1,033
	7	40041	5,209

m = 7	n	w_n^m	v_n^m
S_m = 9			
L_m = 5	1	2	2
	2	3	3
	3	5	5
	4	10	7
	5	14	11
	6	16	13
	7	41	29
	8	61	43
	9	11111	2,801

m = 8	n	w_n^m	v_n^m
S_m = 14			
L_m = 9	1	2	2
	2	3	3
	3	5	5
	4	7	7
	5	111	73
	6	141	97
	7	161	113
	8	401	257
	9	661	433
	10	4611	2,441
	11	6441	3,361
	12	60411	24,841

m = 8 (cont'd)	n	w_n^m	v_n^m
	13	444641	149,921
	14	444444441	76,695,841

m = 9	n	w_n^m	v_n^m
S_m = 12			
L_m = 4	1	2	2
	2	3	3
	3	5	5
	4	7	7
	5	14	13
	6	18	17
	7	41	37
	8	81	73
	9	601	487
	10	661	541
	11	1011	739
	12	1101	811

m = 10	n	$w_n^m = v_n^m$ (commas inserted)
S_m = 27		
L_m = 8	1	2
	2	3
	3	5
	4	7

m = 10 (cont'd)	n	$w_n^m = v_n^m$ (commas inserted)
	5	11
	6	19
	7	41
	8	61
	9	89
	10	409
	11	449
	12	499
	13	881
	14	991
	15	6,469
	16	6,949
	17	9,649
	18	9,949
	19	60,649
	20	90,001
	21	90,469
	22	666,649
	23	946,669
	24	9,000,049
	25	60,000,049
	26	66,000,049
	27	66,600,049

Evaluation of the m-kernel for $2 \leq m \leq 10$.

$m = 2$.

For $m = 2$ we obtain $w_1^m = 10 = 2_{10}$, $w_2^m = 11 = 3_{10}$. For any prime $p > 3_{10}$ we must have $p \nrightarrow 11$ since p must start and end with 1 in base 2; thus the 2-kernel is complete.

$m = 3$.

For this case we obtain $w_1^m = 2$, $w_2^m = 10$, $w_3^m = 111$. To show that these entries exhaust the kernel we note that any larger value must have the form 10^n1 because (1) the digit 2 is excluded, (2) no more than 2 1's may be present to avoid $\rightarrow 111$, and (3) the final digit must not be 0 to avoid divisibility by 3. However $10^n1 = 3^{n+1}+1$ is always even, excluding this case.

$m = 4$.

For this case $w_1^m = 2$, $w_2^m = 3$, $w_3^m = 11$. For any larger value in the kernel, digits 2 and 3 are excluded and no more than one 1 may be present (resulting in the form 10^n); hence no such prime can exist.

$m = 5$.

For $m = 5$ we have $w_1^m = 2$, $w_2^m = 3$, $w_3^m = 10$, $w_4^m = 111$. From this point on, digits 2 and 3 are excluded and no more than two 1's are permitted. The remaining allowable digits then, are 0, 1, and 4. Every remaining entry must have an odd number of 1's to exclude the number being even, since all other digits would be even. Thus each remaining entry must have exactly one 1, which must not be followed by 0 (to avoid $\rightarrow 10$).

The 3 digit candidates satisfying these restrictions are 144, 401, 414, and 441, of which 401 and 414 are prime and are thus the next two words in the kernel.

To determine the remaining kernel entries we first show that all such entries must have the form 14^n or 4^n1. To show this we first note that 0 is excluded as a digit because 1 would have to follow every occurrence of 0 (to avoid $\rightarrow$ 10); thus 4 would have to precede every 0, but then $\rightarrow$ 401. Thus the only permissible digits are 1 and 4. Since only one 1 is allowed, a number beginning with 1 must have the form 14^n. Similarly, a number beginning with 4 must have the form 4^n1 to avoid $\rightarrow$ 414. The existence of an entry 14^n in the kernel clearly precludes the existence of any other word in this form, and similarly for 4^n1. A check shows that the smallest primes having these forms are $14444 = 1{,}249_{10}$ and $44441 = 3{,}121_{10}$, thus completing the 5-kernel.

$m = 6.$

Enumerating the initial elements for this case we obtain $w_1^m = 2$, $w_2^m = 3$, $w_3^m = 5$, $w_4^m = 11$. All other entries must contain only digits 0, 1, and 4, with exactly one 1 present as the last digit to prevent the number from being even. Additional restrictions are that numbers of the form 40^n1 and $40^{2n+1}41$ are excluded since they are divisible by 5 and 7 respectively. The 3- and 4-digit numbers satisfying these restrictions are $441 = 169_{10}$, $4401 = 1{,}009_{10}$, and $4441 = 1{,}033_{10}$, of which only the latter two are prime. Any additional elements of the kernel must have the form $40^{2n}41$ for $n > 0$, because (1) at most two 4's are permitted (since 40^n1 is excluded this means that exactly two 4's must be present), and (2) if a 0 occurred between the second 4 and the final digit 1 then → 4401. The smallest number in this form is 40041 = $5{,}209_{10}$ which is prime and therefore in the kernel. Since no larger number of this form is allowed, the 6-kernel is complete.

m = 7.

The 1- and 2-digit entries for this case are 2, 3, 5, 10, 14, 16, 41 and 61. To complete the kernel we first show that any remaining entry must have the form 1^{2n+1} for $n \geq 1$ (and thus there can be at most one additional entry). To show this we first note that a number which contains only even digits would be even, hence any kernel entry must have an odd digit. The only allowable odd digit is 1 since 3 and 5 are excluded. Thus, if an entry contained 0, 4, or 6 we would have → one of the following: 10, 14, 16, 41 or 61, all of which are in the kernel. This shows that 1 is the only allowable digit. The smallest prime in this form is 11111 = $2{,}801_{10}$, which completes the kernel.

m = 8.

The 1-digit entries for this case, namely 2, 3, 5, 7, restrict the remaining allowable digits to 0, 1, 4, and 6. Since 1 is the only allowable odd digit, every additional entry must end in 1. None of the 2-digit candidates 11, 41, 61 satisfying these restrictions is prime. However, of the 3-digit candidates, those found to be prime are 111 = 73_{10}, 141 = 97_{10}, 161 = 113_{10}, 401 = 257_{10}, and 661 = 433_{10}.

Additional restrictions from this point on are that (1) no more than two 1's can appear, (2) no more than one 6 can appear (to avoid → 661 which would follow since every entry must end in 1), and (3) an entry of the form 10^n1 is excluded because its value would be $8^{n+1}+1 = (2^{n+1}+1)(4^{n+1} - 2^{n+1}+1)$, i.e., it could be factored. If 1 occurs then it must not be followed by 4 or 6 to avoid $\rightarrow$ 141 and $\rightarrow$ 161. Thus we conclude that (4) all 1's must be consecutive and right-adjusted, and no entry can begin with 1. If the number starts with 6 and contains 0 it must have either the form 60^n4^m1 or 60^n4^m11. However, 60^n1 is divisible by 7, while 60^n11 and 60^n41 are divisible by 3, excluding these cases. Finally, since $8 \equiv -1 \bmod 3$, a number containing only 4's and 1's must have an odd number of digits to avoid being divisible by 3.

The 4-digit numbers satisfying all the above restrictions are $4461 = 2{,}353_{10}$ (13), $4611 = 2{,}441_{10}$ (prime), $4{,}641 = 2465_{10}$ (5), $6411 = 3{,}337_{10}$ (47), and $6441 = 3{,}361_{10}$ (prime). From this point the admissible forms are $4^{2n+1}11$, $4^{2n}1$, 4^n61, 4^n641, and 60^n411, and there is at most one entry for each. The 5-digit numbers in these forms are $44411 = 18{,}697_{10}$ (7), $44441 = 18{,}721_{10}$ (97), $44461 = 18{,}737_{10}$ (41), $44641 = 18{,}849_{10}$ (3), and $60411 = 24{,}841_{10}$ (prime).

No further kernel entries can begin with 6. For the 6-digit candidates, the only possibilities are $444461 = 149{,}809_{10}$ (11), and $444641 = 149{,}921_{10}$ (prime). From this point the only allowable forms are $4^{2n+1}11$, $4^{2n}1$, and 4^n61. The 7-digit candidates in these forms are $4444411 = 1{,}198{,}345_{10}$ (5), $4444441 = 1{,}198{,}369_{10}$ (23), and $4444461 = 1{,}198{,}385_{10}$ (5). The only 8-digit possibility is $44444461 = 9{,}586{,}993_{10}$ (13). The 9-digit possibilities are $444444411 = 76{,}695{,}817_{10}$ (11), $444444441 = 76{,}695{,}841_{10}$ (prime), and $444444461 = 76{,}695{,}857_{10}$ (7). This last entry precludes any others since it would be a subword; thus the 8-kernel is complete.

$m = 9.$

The 1-digit entries are 2, 3, 5, and 7, restricting the digits in the remaining entries to 0, 1, 4, 6, and 8. The 2-digit primes satisfying these restrictions are 14, 18, 41, and 81. From this point every prime must end in 1 to avoid being even, and for the same reason, must contain an odd number of 1's. Also, 4 and 8 cannot occur, so the only allowable digits are 0, 1, and 6. Finally we can exclude a number of the form 1^n, $n \geq 2$, because its value is $(9^{n+1}-1)/8 = (3^{n+1}+1)(3^{n+1}-1)/8$, i.e. it can be factored.

The 3-digit numbers satisfying all of the above restrictions are $601 = 487_{10}$ and $661 = 541_{10}$, both of which are prime. From this point at most one 6 is allowed. The 4-digit numbers satisfying this additional restriction are $1011 = 739_{10}$ (prime), $1101 = 811_{10}$ (prime), $6001 = 4{,}375_{10}$ (5), and $6111 = 4{,}465_{10}$ (5). We now show that no further kernel entries are possible. First we show that no further entry can contain 0. If an entry contains 0 then 1 must precede it (otherwise → 601). Since the number of 1's must be odd, there would then have to be at least three of them. But then → 1011 or → 1101. Thus the only allowable digits are 1 and 6. However in base 9 the numbers 11, 16, 61, and 66 are all divisible by 5; thus every entry must have an odd number of digits. Since each entry must also have an odd number of 1's the number of 6's must be even. However, no more than one 6 may be present (to avoid → 661); thus no 6 can be present, and the entry must then have the form 1^n. Since this latter case was ruled out earlier, we conclude that the 9-kernel has no further entries.

$m = 10.$

The 1- and 2-digit entries are 2, 3, 5, 7, 11, 19, 41, 61, and 89. From this point the allowable digits are 0, 1, 4, 6, and 9, and the last two digits in every entry must be either 01, 09, 49, 69, 81, 91, or 99. In each entry, 1 may appear at most once and must not precede 9, nor can it be preceded by 4 or 6. Since every entry must end with 1 or 9 (to avoid being even) no number can begin with 1. 8 must not precede 9. We can rule out all numbers of the form 46^n9 since they are divisible by 7 (giving 6^n7 as quotient) and all numbers the sum of whose digits is divisible by 3 since these too are divisible by 3. The 3-digit primes satisfying these restrictions are 409, 449, 499, 881, and 991.

We now show that from this point (1) 8 is excluded as a digit and (2) no entry can begin with 4 so that all further entries must begin with 6 or 9. To show that 8 is excluded we note that any entry containing 8 must end in 1 (to avoid → 89); thus at most one 8 must appear to avoid → 881. However 4 cannot appear (otherwise → 41) so the only other nonzero digits allowed are 6 and 9. But only one 1 is allowed, so the sum of the digits would be a multiple of 3 and the number could not be prime. To show that no entry can begin with 4, we note that such an entry would have to end in 9 to avoid → 41 and thus all interior digits must be 6 to avoid → 409, → 449, or → 499. However a number of the form 46^n9 was ruled out earlier; thus no number can begin with 4.

To further limit the allowable choices of entries we show that all such entries must end with either 001, 049, 469, 649, 669, or 949. First we show that the allowable 2-digit endings are 01, 49, and 69. 81 is clearly eliminated, leaving the possibilities 01, 09, 49, 69, 91 and 99 from our earlier list. To eliminate 09 and 99 we note that any entry with one of these endings would have to contain 4 (otherwise all non-zero digits would be 6 or 9) but then → 409 or → 499. To eliminate 91 we note that 9 must precede in any entry ending this way to avoid → 41 or → 61, but then → 991.

By considering the allowable 2-digit endings and previous kernel entries we obtain the following list including the allowable 3-digit endings: 001, 049, 069, 469, 649, 669, 901, 949, and 969. We can eliminate 069 and 969 because a 4 must precede these groups, implying → 409 or → 499. Similarly, 901 is excluded because otherwise → 991. Thus we obtain the earlier list of 3-digit endings as the only permissible ones.

The 4-digit primes having one of these six endings and satisfying the other restrictions established earlier, are 6,469, 6,949, 9,649, and 9,949. From this point we can eliminate 949 as an ending because it would have to be preceded by 6 or 9.

To determine the remaining entries in the kernel we first recall that the allowable 3-digit endings at this point are 001, 049, 469, 649, and 669. It is not difficult to show that the remaining entries must have one of the forms $6\{0,6\}^n 049$, $6\{0,6\}^n 649$, $90^n 001$, $90^n 049$, $90^n 469$, or $946^n 669$. The 5-digit candidates in these forms are 60,049 (11), 60,649 (prime), 66,049 (257), 66,649 (11), 90,001 (prime), 90,049 (17), 90,469 (prime), and 94,669 (41).

From this point the entries must have one of the four forms $6^{m+1} 0^n 049$, $6^n 649$, $90^n 049$, or $946^n 669$. The 6-digit candidates in these forms are 600,049 (17), 660,049 (13), 666,049 (79), 666,649 (prime), 900,049 (19), and 946,669 (prime). The remaining entries are limited to the forms $6^{m+1} 0^n 049$ with $0 \le m \le 2$, and $90^n 049$. The 7-digit candidates in these forms are 6,000,049 (11), 6,600,049 (19), 6,660,049 (11), and 9,000,049 (prime). From this point the only allowable form is $6^{m+1} 0^n 049$ with $0 \le m \le 2$. The 8-digit entries in this form all turn out to be prime (surprisingly) and thus complete the 10-kernel.

References:

1. Goldberg, K., Newman, M., and Haynsworth, E., Combinatorial analysis, in "Handbook of Mathematical Functions," (M. Abramowitz and I.A. Stegun, Eds.) Chap. 24, pp. 844-873, National Bureau of Standards Applied Mathematics Series 55, 1964.

2. Kruskal, J.B., Well quasi-ordering, the Tree Theorem, and Vasonyi's conjecture, Trans. Amer. Math. Soc. 95 (1960), 210-225.

3. Lehmer, D.N., List of prime numbers from 1 to 10,006,721, Carnegie Institution of Washington, Publication No. 165, Washington, D.C. (1914).