Katedralskolan, Uppsala

Finding the n^{th} digit in a sequence of positive integers placed in a row in ascending order.

By Artem Los

12345 . . . ?

Candidate number: 1291-032
Supervisor: Erik Östergren
Page count: 12

Abstract

In this essay the sequence of positive integers placed in a row in ascending order is going to be examined, i.e.

$$12345678910111213\ldots99100101102\ldots$$

First, it is going to be found that the number of digits in an interval of all x digit numbers is:

$$f(x) = 9 \times x \times 10^{x-1} \qquad x \in \mathbb{Z}^+$$

Secondly, it is going to be found using Perturbation method that the sum of $f(x)$ has a closed form, which is:

$$g(n) = \sum_{1 \leqslant k \leqslant n} f(k) = \frac{9(n+1)10^n - 10^{n+1} + 1}{9} \qquad k, n \in \mathbb{Z}^+$$

A pattern will be observed in the values produced by the sum above, which will make it possible to construct the number even quicker. This is proved using the derived closed form above. Later on, the problem is going to be solved and generalised. It is going to be shown that by calculating,

$$p = 10^{\lceil a \rceil} - 1 - \left\lfloor \frac{g(\lceil a \rceil) - g(a)}{\lceil a \rceil} \right\rfloor, g(a) = n \qquad a \in \mathbb{R}^+$$

and

$$(g(\lceil a \rceil) - g(a)) \mod \lceil a \rceil$$

the n^{th} digit can be found. Through simplification of these formulæ, it is going to be found how this can be performed much quicker.

Contents

List of Tables

Introduction

Throughout this work, my main objective is to take this very hard looking problem, solve it, and, construct a very quick way of solving it that does not require that much thinking. The end product should include as few calculation steps as possible.

In the beginning, intervals are going to observed in order to construct an expression for the number of digits in an interval of all n digit numbers. Later on, a closed form for the sum of all intervals is going to be constructed, which will make it possible to find a quick way to solve this problem. The final part of the essay will clarify how the n^{th} digit can be found in a quick way and later, in the conclusion, some suggestions are going to be made of how the solution can be implemented into a computer program.

Rationale

The idea of trying to solve this problem arose when I discovered a book on my bookshelf, called *Moscow Mathematics Olympiads*[1]. It was quite a boring book according to the current standards (black and white, few pictures, old paper), however, the problems there spoke for themselves. They were from the highest mathematics Olympiad in the Soviet Union. Particularly this problem was for 10-11 graders, which is for similar pupils as me.

Some years ago, when I was very excited about modular arithmetic, which, by the way, received attention quite recently (sometime in the '70s), I tried to solve the problem by investigating the frequency of digits $1 \to 9$ and their place. This did not lead anywhere, since I so forcefully wanted to use modular arithmetic, and did not consider other options.

Once I became mature for this problem, I started to consider intervals more and more, and after a week of intense investigation (and some more weeks after that), I was able to produce this essay, where I went into depth of specific concepts.

The question Find a way to calculate the n^{th} digit in the sequence of ascending integers, i.e.

$$123456789101112\ldots99100101\ldots$$

Notation Some of these might look familiar to you, however, in case some concepts are new, it is good to clarify them at this stage, rather than doing it in the actual essay.

Table 1: Definition of specific notation used in this essay

$a_n \ldots a_m$	An interval between a_n and a_m, where n, m denote the position of a number in the positive integer set. It is referred to as *interval form.*
n_d	A function that takes in one parameter (in an interval form) and returns the number of digits in that interval.
$10^{n-1} \ldots 10^n - 1$	An interval of all x digit numbers.
$\lfloor x \rfloor$	Floor function: max $\{n \mid n \leqslant x, n \in \mathbb{Z}\}$
$\lceil x \rceil$	Ceiling function: min $\{n \mid n \geqslant x, n \in \mathbb{Z}\}$
x mod y	Binary operator: $x - y\lfloor x/y \rfloor$, $y \neq 0$

[1] G.A. Galperin, A.K. Tolpygo (1986). Moskovskie matematicheskie olimpiady. Moscow, USSR: Prosveshenie. p.24, p.155.

1 Analysis of the sequence

In order to be able to solve this problem, it is a good idea to specify intervals, where it is known at what place a specific digit is. Throughout this essay, intervals of a great importance are those that include numbers with the same amount of digits. For example, the first objective is to find the number of digits in a one digit number interval. The second objective is to find the number digits in a two digit number interval, and so on.

Let us start with the interval of all one digit numbers:

$$1 \dots 9 \tag{1.1}$$

Sure enough, there are 9 digits in total, so we have solved the first interval. Now we are to investigate the second interval that consists of only two digit numbers, i.e.

$$10 \dots 99 \tag{1.2}$$

To make this simpler, let us split this interval into smaller parts, i.e.

$$\left.\begin{array}{l} 10 \dots 19 \\ \vdots \\ 90 \dots 99 \end{array}\right\} \times 9$$

By adding the number of digits in each of these intervals, we get the number of digits in the interval (1.2). It is easy to see that there are 2×10 digits in the interval $10 \dots 19$, because it is actually two $0 \dots 9$ intervals. The number of digits in $10 \dots 19$ is the same for the remaining parts of the interval (1.2), and since we know that the total number of these small intervals is nine, we can now figure out the number of digits in the interval (1.2). It is,

$$n_d(10 \dots 99) = 9 \times (2 \times 10) = 180$$
$$\therefore n_d(1 \dots 99) = 9 + 180 = 189 \tag{1.3}$$

This also means that 9 is the 189th digit in the interval (1.3).

Let us now investigate the third interval – an interval with only three digit numbers, i.e.

$$100 \dots 999 \tag{1.4}$$

As before, when the interval (1.2) was split up into smaller parts, the same has to be done for the interval (1.4).

$$\left.\begin{array}{l} \left.\begin{array}{l} 100 \dots 109 \\ \vdots \\ 190 \dots 199 \end{array}\right\} \times 10 \\ \vdots \\ \left.\begin{array}{l} 900 \dots 909 \\ \vdots \\ 990 \dots 999 \end{array}\right\} \times 10 \end{array}\right\} \times 9$$

In the interval $100\dots109$, there are 3×10 digits, and this is the same for the remaining small intervals, e.g. for $190\dots199$. So, the total number of digits in the interval (1.4) is,

$$n_d(100\dots999) = (3 \times 10) \times 10 \times 9 = 9 \times 3 \times 10^2 \tag{1.5}$$

We might start to see a pattern that for an inteval of all n digit number, i.e.,

$$n_d(10^{n-1}\dots10^n - 1) = 9 \times n \times 10^{n-1} \qquad n \in \mathbb{Z}^+ \tag{1.6}$$

The proof for (1.6) is very simple, however, it requires an entirely different approach than what we have used in previous examples. Before we prove this in general, let us look at an interval of all 4 digit numbers, i.e.

$$1000\dots9999 \tag{1.7}$$

Instead of splitting this interval into smaller pieces, we can express (1.7) so that it will allow us to use combinatorics. So,

$$n_d(1000\dots9999) = n_d(0000\dots9999) - n_d(0000\dots0999)$$

The first expression on the right hand side is basically asking us to find the number ways we can arrange 10 objects i.e. $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ in a row with 4 places. The second expression on the right hand side is similar, but there are only three places in a row. Therefore,

$$n_d(0000\dots9999) - n_d(0000\dots0999) = 4 \times 10^4 - 4 \times 10^3 = 36000$$

Notice that we take the number of permutations times 4 to get the number of digits (1 permutation is 4 digits).

In general, for an interval of all n digit numbers, i.e.

$$n_d(10^{n-1}\dots10^n - 1) = n_d(\overbrace{0\dots0}^{n}\dots10^n - 1) - n_d(\overbrace{0\dots0}^{n}\dots10^{n-1} - 1) \tag{1.8}$$

Here, the problem reduces to permutations with repetitions. In the first expression on the right side in (1.8), there are 10 objects, $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, that are to be placed in a row with n places. In the second expression on the right side, there are 10 objects that are to be placed in a row with $n-1$ places. Since we need the number of digits, we need to take each permutation times the number of digits a number consists of, which is n (1 permutation is n digits). Therefore,

$$\begin{aligned} n_d(\overbrace{0\dots0}^{n}\dots10^n - 1) - n_d(\overbrace{0\dots0}^{n}\dots10^{n-1} - 1) &= n \times 10^n - n \times 10^{n-1} \\ &= n(10 \times 10^{n-1} - 10^{n-1}) \\ &= n \times 9 \times 10^{n-1} \end{aligned}$$

$\square$

2 Finding the sum

In this section we are to find a way to generalize (1.6) even further, by constructing a function for the number of digits in a specific interval. We can rewrite (1.6) to a more convenient form:

$$f(x) = 9 \times x \times 10^{x-1} \qquad x \in \mathbb{Z}^+ \tag{2.1}$$

Here, $f(x)$ will give us the number of digits in an interval of all x digit numbers. The benefit of being able to express this as a function is the ability to define certain intervals with an expression. Since we know the expression, it is easier to construct a sum, and later find a closed form of that sum.

It can be seen that the number of digits in the interval $1 \dots 999$ is the sum of one digit numbers, two digit numbers and three digit numbers. Hence, it can be expressed as:

$$\sum_{x=1}^{3} f(x) = 2889$$

Generally, given that the sequence contains numbers from 1 to a specific n digit number, we have:

$$g(n) = 9 \sum_{x=1}^{n} x 10^{x-1} \qquad x, n \in \mathbb{Z}^+ \tag{2.2}$$

Now, we have a sum, which means we are halfway through the problem. This function, $g(n)$, will give us the number of digits in the interval $1 \dots 10^n - 1$. The only thing that remains to find is a closed form for the sum (2.2).

3 Finding the closed form

This section might not be important for small values of n in the sum (2.2), however, since we want to have a solution that covers all positive integers, we need to find a closed form. We do not really want to evaluate the sum manually, for large values of n. So the question we now face is how this is going to be achieved.

3.1 Perturbation method

In this method[2], the idea is to express the sum as a letter, i.e.,

$$S_n = \sum_{0 \leqslant k \leqslant n} a_k \qquad k, n \in \mathbb{Z}$$

and later express S_{n+1} in two different ways, i.e,

$$S_n + a_{n+1} = a_0 + \sum_{0 \leqslant k \leqslant n} a_{k+1} \tag{3.1}$$

[2]Graham R.L., Knuth D. E., Patashnik O. (1994). *Concrete Mathematics - A Foundation for Computer Science.* 2nd ed. Courier Westford, Massachusetts, USA: Addison Wesley. p. 32 - p. 33.

We should now try to express the sum on the right hand side in terms of S_n, and hope that it will lead to a closed form.

So, the actual sum (note that we ignored 9) in (2.2) can be expressed as:

$$S_n = \sum_{0 \leqslant k \leqslant n} k10^{k-1} \qquad k, n \in \mathbb{Z}$$

Now, by expressing this as (3.1), we get:

$$S_n + (n+1)10^n = \sum_{0 \leqslant k \leqslant n} (k+1)10^k$$

We can split up the sum on the right hand side into two new sums:

$$S_n + (n+1)10^n = \sum_{0 \leqslant k \leqslant n} k10^k + \sum_{0 \leqslant k \leqslant n} 10^k$$

The first sum on the right hand side is $10S_n$, while the last one is simply a geometric progression. We can now solve for S_n.

$$9S_n = \frac{9(n+1)10^n - 10^{n+1} + 1}{9}$$

$$g(n) = 9 \sum_{0 \leqslant k \leqslant n} k10^{k-1} = \frac{9(n+1)10^n - 10^{n+1} + 1}{9} \qquad k, n \in \mathbb{Z}^+ \tag{3.2}$$

Note, since we did not include the constant 9 from (2.2) into the definition of S_n, $9S_n$ is the answer. Also, as $S_0 = 0$ the boundaries for the sum can be changed to $1 \leqslant k \leqslant n$. Hence, (3.2) is our closed form for (2.2). $\square$

3.2 Observing a pattern - Artem's method

Closed form derived using Perturbation method is good, since it reduces the computational time for large values of n. However, it is much better if we manage to find a closed form that is based on a specific pattern in the actual number, i.e. in the way digits are arranged. For this, let us investigate the values in *Table 2* (on next page). Clearly, we can see that the first digit is $n-1$, and the last digit tends to be 9. Then, we have $n-1$ 8s in the middle. Mathematically, we can express following as:

- First digit is $n-1 \implies (n-1)10^n$
- There are $n-1$ 8s in the middle $\implies \frac{10^{n-1}-1}{9} \times 8 \times 10$
- The last digit is 9 $\implies 9$

Adding these up gives us:

$$n_d(1 \ldots 10^n - 1) = (n-1)10^n + \frac{10^{n-1} - 1}{9} \times 8 \times 10 + 9 \tag{3.3}$$

Table 2: The number of digits in an interval of $1 \dots n$ digit numbers

n	$n_d(1 \dots 10^n - 1)$
1	9
2	189
3	2889
4	38889
5	488889
6	5888889
7	68888889
8	788888889
9	8888888889
10	98888888889

Sure enough, this looks quite likely to be true, in fact, it is true, however, at this stage, we can only claim that it is true for $n \leqslant 10$, because this pattern was observed in values generated with (3.2) up to $n = 10$. Since we know that the closed form (3.2) is true, we can try to rearrange (3.3) in such a way that will be equal to (3.2).

$$
\begin{aligned}
(n-1)10^n + \frac{10^{n-1}-1}{9} \times 8 \times 10 + 9 &= \\
&= n10^n - 10^n + \frac{80}{9}(10^{n-1} - 1) + 9 \\
&= n10^n - 10^n + \frac{8}{9}10^n - \frac{80}{9} + 9 \\
&= 10^n \left(n - 1 + \frac{8}{9}\right) + \frac{1}{9} \\
&= \frac{9\left(n10^n - 10^n + \frac{8}{9}10^n\right) + 1}{9} \\
&= \frac{9\left(n10^n - 10^n + 10^n - \frac{1}{9}10^n\right) + 1}{9} \\
&= \frac{9(10^n(n+1) - 10^n - \frac{1}{9}10^n) + 1}{9} \\
&= \frac{9(10^n(n+1)) - 9 \times 10^n - 10^n + 1}{9} \\
&= \frac{9(10^n(n+1)) - 10^{n+1} + 1}{9}
\end{aligned}
$$

We have now showed that (3.3) is equal to (3.2), which means that using this hyper closed form, we can find the number of digits in $1 \dots 10^n - 1$ with very little computational power. $\square$

4 Solving the problem

Now, after that we have defined concrete intervals, discovered a sum and figured out its closed form we can solve the actual problem. For simplicity, let us work with a number rather than go too much into a general method, where the number is n.

Now, consider the following problem instead: *Find the 2345th digit in the sequence of positive integers placed in a row in ascending order.* Here is what we know:

$$f(x) = 9 \times x \times 10^{x-1} \qquad x \in \mathbb{Z}^+ \tag{4.1}$$

$$g(n) = 9 \sum_{1 \leqslant k \leqslant n} k10^{k-1} \qquad k, n \in \mathbb{Z}^+ \tag{4.2}$$

The first step is to figure out between which intervals this digit is situated, i.e.,

$$g(a) = 2345$$

By using the closed form in (3.2), and an equation solver, $a \approx 2.92$. This means we should find values: $g(\lceil a \rceil)$ and $g(\lfloor a \rfloor)$.

$$g(3) = 2889$$
$$g(2) = 189$$

Since $g(2) < 2345 < g(3)$, the digit must be surrounded by three digit numbers. There is no point of keeping other digits, before the 2345th digit in the interval, so we can subtract them, i.e.,

$$g(3) - g(a) = 544$$

This means that the 2345th digit is 544 digits away from the last digit in the number 999, i.e.,

$$x \dots 999$$

We know that there are only three digit numbers in that interval, so we can divide the number of digits in the interval with how many digits each number contains. This tells us how many numbers away the digit is from 999. Since it is not possible to divide 544 by 3, we need to consider the remainder, i.e. $544 \bmod 3 = 1$, so it is 1. Now, by subtracting 1 from 544 we get 543, which is divisible by 3. This means that we are $543/3 = 181$ numbers away. Now, using simple subtraction, $999 - 181 = 818$. Hence, the the 543th number is the last digit in the number 818. However since we subtracted one, we have to go 1 step back, so the 544th digit is 1. In conclusion, the 2345th digit is 1. $\square$

Let us now generalise the previous steps into a more clear algorithm. Given that we are to find the n^{th} digit:

- Find a such that $g(a) = n$. If a is an integer, the digit is found.
- Let $d = g(\lceil a \rceil) - g(a)$
- Let $r = d \bmod \lceil a \rceil$

- Let $s = d - r$
- Let $q = s/\lceil a \rceil$
- Let $p = 10^{\lceil a \rceil} - 1 - q$
- In $p = (a_r \dots a_1 a_0)$, choose the right digit.

In fact, we can simplify q by substituting previous variables:

$$q = \frac{s}{\lceil a \rceil} = \frac{d-r}{\lceil a \rceil} = \frac{d - d \mod \lceil a \rceil}{\lceil a \rceil} = \frac{d - \left(d - \lceil a \rceil \left\lfloor \frac{d}{\lceil a \rceil} \right\rfloor\right)}{\lceil a \rceil} = \left\lfloor \frac{d}{\lceil a \rceil} \right\rfloor$$
$$= \left\lfloor \frac{g(\lceil a \rceil) - g(a)}{\lceil a \rceil} \right\rfloor$$

So, in order to find p, i.e. the number that contains the digit,

$$p = 10^{\lceil a \rceil} - 1 - \left\lfloor \frac{g(\lceil a \rceil) - g(a)}{\lceil a \rceil} \right\rfloor, g(a) = n \tag{4.3}$$

5 Quick solution

We have now gone through the theory behind the solution to the problem, and optimized it. Since the origin of this question is from a Moscow Mathematics Olympiad, it is important to construct a quick way of finding the n^{th} digit.

The problem[3] asks us to find the 206788th digit in the sequence 123456789101112 . . . .

1. In order to solve $g(a) = 206788$ quickly, we can look at intervals for $1, 2, 3, 4, 5, 6$ etc. digit numbers. From (3.3), we can use the pattern we observed in *Table 2* to find these very quickly.

   Table 3: The number of digits in an intervals with 1,2,3,4,5,6, digit numbers

n	$n_d(1 \dots 10^n - 1)$
1	9
2	189
3	2889
4	38889
5	488889
6	5888889

   We see that $488889 > 206788$, which means that the digit is not encircled by numbers where there are more than 5 digits, hence $\lceil a \rceil = 5$.

[3] G.A. Galperin, A.K. Tolpygo (1986). Moskovskie matematicheskie olimpiady. Moscow, USSR: Prosveshenie. p.24, p.155.

2. We know that $g(a) = 206788$, therefore, using (4.3), we get:

$$10^5 - 1 - \left\lfloor \frac{488889 - 206788}{5} \right\rfloor = 99999 - \lfloor 56420.2 \rfloor = 43579 \tag{5.1}$$

   Notice that you do not need to care about the fractional part of anything inside a floor function, i.e. $\lfloor \pi \rfloor = 3$, and also that when $10^5 - 1$, you get 5 nines. So, for $10^n - 1$ you get n nines.

3. The number with our digit is 43579, so we need to work with the remainder, i.e. find $(g(\lceil a \rceil) - g(a)) \mod 5$. This is equivalent to the steps when rewriting an improper fraction as a mixed fraction:

$$\frac{488889 - 206788}{5} = \frac{282101}{5} = 56420 + \frac{1}{5}$$

   The remainder is therefore 1 (from the proper fraction). This means that the 206788th digit is 7. Notice that the digit count is from right to left, i.e. if the remainder would be zero, the digit would be 9, and so on.

The good thing about this solution is that it does not require that much computational power to find, nor do you need to use problem solving skills to actually get the digit.

Conclusion

The aim of this essay was to find a quick way, preferably using a formula, which would help us to deduce the n^{th} digit in a sequence of positive integers in a row in ascending order. Using the analysis of the function that would return the number of digits given an interval of all n digit numbers, a sum was constructed for these intervals, and later, a closed form was determined. These findings contributed to a general solution for the problem, a proof for an observed pattern, and later a quick solution to the problem. Since the main objective was to find a simple way of finding the n^{th} digit, the fact that it requires one main calculation, which can be performed using functions floor, ceiling and exponent, respectively, tells that this essay reached the final objective.

Author's comments and suggestions

The methods that I used in this essay try to generalise the problem of finding the n^{th} digit, in order to make every step computer friendly (this means that it can be generated with several lines of computer code, if not one line). Already in the analysis, where the method of finding the number of digits of an interval of all n digit numbers was presented, an expression was derived and proved to be true for all values of $n \in \mathbb{Z}^+$. The problem was viewed from different angles and finally proved using basic combinatorics.

The derived expression was very helpful when setting up a sum and later generalising it by finding a closed form. Once again, using a closed form we can easily find the value both numerically and using a computer. It could have been a limitation if I would not be able to find it, because for large values of n, the evaluation of the sum will take a lot of time. The fact that it can be expressed as a single expression leaves room for even more possibilities. It can allow us to analyse how quickly the sum grows, for example.

During my work on the first closed form of the sum that I found using Perturbation method, I saw an interesting pattern in the values generated by the sum. Yet again, in order to prove it, I tried to express it using a mathematical expression which needed to be the same as the first one in order to be true. This fact reduced the time of finding the n^{th} digit both numerically and using a computer.

Later on, when the actual problem of finding the n^{th} digit was to be solved, I tried to generalise the final formula as much as possible. Again, knowing the definition of the mod operator made the final equation even simpler. The expression in (4.3) could easily be calculated by a computer and a human being. Even if the equation solver technique was mentioned, given that $g(a) = n$, $\lceil a \rceil$ can be found by looking at the number of digits n consists of (in computer programming, this would be the length of the string). The $\lceil a \rceil$ would then give us a $\lceil a \rceil + 1$ digit number (for $n > 1$). Therefore, if we want to find the n^{th} digit, $\lceil a \rceil = n-1, n > 1, n < g(n-1)$ or $\lceil a \rceil = n, n > 1, g(n-1) < n < g(n)$. This reduces the computational time even further, since we in the beginning can treat numbers as strings. Later on, a quick subtraction can be performed, which does not take that much time.

/Artem
17.12.2013

Bibliography

Written using TEXstudio.

1. G.A. Galperin, A.K. Tolpygo (1986). Moskovskie matematicheskie olimpiady. Moscow, USSR: Prosveshenie. p.24, p.155.

2. Graham R.L., Knuth D. E., Patashnik O. (1994). *Concrete Mathematics - A Foundation for Computer Science*. 2nd ed. Courier Westford, Massachusetts, USA: Addison Wesley. p. 32 - p. 33.