KAPREKAR'S CONSTANT(S)

LIAM DONOVAN

Abstract. D.R. Kaprekar was an Indian 'mathematician'. The only reason that's in quotes is because he wasn't really a mathematician by trade, he was a school teacher with no postgraduate mathematical training. However, he made many striking discoveries, particularly with classes of natural (counting) numbers. This paper is going to talk about one such result, call Kaprekar's number, or Kaprekar's constant. To be more exact, we're only going to talk about one of them, even though there are many others. We're specifically going to talk about the four digit example, in base 10, since that's the one that Kaprekar found. Truthfully, it doesn't have many uses, *but* it is very interesting and exemplifies a quality often lost in modern mathematics: just doing math for fun and not because you seek to be overwhelmingly rigorous and exact.

Contents

Date: May 2023.

1. A Quick Exercise

Alright so let's play a game: give me any four digit number. I'm going to ask you to *not* make all its digits the same (since then the game would end quite quickly, with a not so satisfying result). Let's say you gave me 2001. What I'm going to do is rewrite this, from smallest digit to largest digit, which give us 0012. And then, I will do the opposite, rewrite the number from largest to smallest digit, which give us 2100. Okay, now I'm going to subtract the smaller one from the bigger one:

$$2100 - 0012 = 2088$$

Okay, so now I'm going to continue this process with the new value, 2088. Ordering from greatest to least and least to greatest, then subtracting.

$$8820 - 0288 = 8532$$
$$8532 - 2358 = 6174$$
$$7641 - 1467 = 6174$$

Alright, so now we've gotten ourselves into an interesting situation. Of course, if I reorder those digits, the last two rows are going to repeat, forever. Let's try again, this time with maybe 4950:

$$9540 - 0459 = 9081$$
$$9810 - 0189 = 9621$$
$$9621 - 1269 = 8352$$
$$8532 - 2358 = 6174$$
$$7641 - 1467 = 6174$$

Well, that's weird we stopped at the same place again, we end up in that loop of 6174. Okay, one more time, this time with 2697:

$$9762 - 2679 = 7083$$
$$8730 - 0378 = 8352$$
$$8532 - 2358 = 6174$$

And it happened again! So there *is* something going on with this number. If we reorder its digits, it's difference has the exact same digits!

Now let's think about why I asked you *not* to give me a number with all the same digit. Let's say you gave me any number, with some digit x, so then the four digit number would be $xxxx$, and, of course since x is a digit, it is between 0 and 9, inclusive. Well, if I reorder this, nothing happens, so I get:

$$xxxx - xxxx = 0000$$

So, if all the digits are the same, of course, I just get 0. This whole algorithm is called the *Kaprekar's routine* and numbers that repeat infinitely are called *Kaprekar's constants*. We can think of this as a function $K_b(n)$ where b is the base that's we're working in and n is the number that I order its digits. We're usually working in base 10 (normal numbers), so we'll just go ahead and assume that $b = 10$ for this, and not write it out. But there's a actually Kaprekar's numbers in many other bases, like binary (base 2) and hexadecimal (base 16).[1] 0 is a bit of a boring case, so we call it the *trivial Kaprekar constant.*

But how do we know this number is the only number, except 0, that has this property, maybe there's a whole bunch of number's with this property, we just haven't found them yet? Well, what property did we say this number had? We said that its digits are the same, when we perform the routine. We said that $K(n)$ is the 'function' that describes our routine. So what we're really saying is that $K(6174) = 6174$, that is, it's input is the same as its output (it maps to itself); if I put it into the routine, the number that pops out is the exact same. So now we're out to figure out if this number is unique or not.

[1]It's actually the same number, it's just represented differently, if it converges in one base system, it converges in another, also.

2. Starting Our Search

Let's say we have some natural, four digit, number (a positive, whole number), n, which returns itself when put into Kaprekar's routine, that is $K(n) = n$. And we said that this number is four digits, so we write it as $\overline{ABCD}$. I know this 'bar' notation means a lot of different things in different fields of math (mean, complex conjugate, etc.), but here it just means that the numbers under it are digits, so we don't get confused with multiplication. Now, let's call our number, from greatest digit to least $a_U = \overline{abcd}$. Since we ordered it from greatest to least, and the fact that they are digits, we know $9 \geq a \geq b \geq c \geq d \geq 0$. Then we can call our number with the digits from least to greatest $a_L = \overline{dcba}$. So, let's try and figure out some properties that can help us determine what a, b, c, d have to be so that $a_U - a_L$ has the same digits as $ABCD$.

3. Subtracting Our Terms

Before we go to the next step, let's remember what is means to have a number written as $\overline{abcd}$. As an example, think about the number 1234, what are we actually saying here? Well, we are saying the number has 1 one thousands 10^3, 2 hundreds 10^2, 3 tens 10^1, and 4 ones 10^0. This is called positional notation. So what we're really saying when we write 1234 is actually: $1 \cdot 10^3 + 2 \cdot 10^2 + 3 \cdot 10^1 + 4 \cdot 10^0$.

So this means we can write a_U as $a \cdot 10^3 + b \cdot 10^2 + c \cdot 10^1 + d \cdot 10^0$ and a_L as $d \cdot 10^3 + c \cdot 10^2 + b \cdot 10^1 + a \cdot 10^0$. So their difference is $10^3(a-d) + 10^2(b-c) + 10^1(c-b) + 10^0(d-a)$, right? Well, that *would* be the case if $d-a$, $c-b$, $b-c$, $a-d$ were all positive, or zero. However, we know that $9 \geq a \geq b \geq c \geq d \geq 0$. So, what does this mean for $d-a$ and $a-d$? Well, they can't both be positive at the same time, right? Either one of them is positive, and the other is negative, or they're both 0, when $a = d$. But if $a = d$, then going back to the inequality: $9 \geq a \geq b \geq c \geq d \geq 0$, they all have to be equal, right, since we need a to be greater than or equal to b, which is greater than or equal to c, which is then greater than or equal to d, but we said $d = a$, so they all must be equal. But then what happens? Well, $a = b = c = d$, so we're back to that $xxxx$ case again, where $K(xxxx) = 0$. So this is just the trivial case.

So what do we do? Well, what do we normal do when subtraction gives us a negative digit? As an example, try subtracting 7785 from 9563. In elementary school (at least in the US) it was taught to us like this:

$$\begin{array}{r} 9\,{}_{1}5\,{}_{1}6\,{}_{1}3 \\ -\;{}_{1}7\,{}_{1}7\,{}_{1}8\;5 \\ \hline 1\;7\;7\;8 \end{array}$$

We are taught if the columns bottom number is greater then the top number, we need to 'carry' from the next column over. Let's think about what *actually* happening here. Let's rewrite what we're doing in the positional notation again, since that's what we're gonna be using for the actual problem anyway; it's probably easier to see what's going on. Rewriting our subtraction:

$$10^3(9-7) + 10^2(5-7) + 10^1(6-8) + 10^0(3-5)$$

The problem is that $3 - 5 < 0$, so what we do is add and subtract 10, in that term. Since the difference of two digits is, at the lowest -9, when the top number is 0 and the bottom one is 9, if we do this, there will always be a positive number in that spot. So then, we get:

$$10^3(9-7) + 10^2(5-7) + 10^1(6-8) + 10^0(3-5+10) - 10$$

I can just put that extra 10 along with the rest of the 10^1 terms, then factor our the 10^1 to put an extra 1 inside the parentheses.

$$10^3(9-7) + 10^2(5-7) + 10^1(6-8-1) + 10^0(3-5+10)$$
$$10^3(9-7) + 10^2(5-7) + 10^1(6-8-1) + 10^0(8)$$

Moving over to the next column, we got the same problem again, but this time, we're gonna need to add and subtract $10^2 = 100$. Why? Efficiency. The same way that minimum value of the ones column was -9, the digit would still be -9, but we're now in the 10^1 column, so this is basically $10^1 \cdot -9 = -90$. So, it's more convenient to just add and subtract 10^2. I can do that same factoring with the leftover -10^2, like we did in the 10^1 column.

$$10^3(9-7) + 10^2(5-7-1) + 10^1(6-8-1+10) + 10^0(8)$$
$$10^3(9-7) + 10^2(5-7-1) + 10^1(7) + 10^0(8)$$

Doing the same thing one more time, we get.

$$10^3(9-7-1)+10^2(5-7-1+10)+10^1(7)+10^0(8)$$
$$10^3(9-7)+10^2(7)+10^1(7)+10^0(8)$$

Now, the last row is greater than zero, so we can just subtract.

$$\begin{aligned}&10^3(1)+10^2(7)+10^1(7)+10^0(8)\\&=1778\end{aligned}$$

So, that's how subtraction *really* works. Actually this strategy is called *subtraction by regrouping.*

Alright, so we have these $a-d$ and $d-a$ terms. How do we know which one should be negative. Well, if $a-d$ is negative, it has no place to borrow from. Let's do an example where this happens, like 5000 minus 3000. Well that's just -2000. This is pretty obvious, but if the farthest left term is greater, in the bottom number, then that whole number is bigger than the top number, meaning that if we subtract we get a negative number. But, of course, Kaprekar's constant can't be negative, since then we would be adding infinitely and would never reach a loop. So $a>d$.

But then, we know that $d-a$ will be negative, so we're gonna have to carry from the next row over:

$$\begin{aligned}\overline{abcd}-\overline{dcba}&=10^3(a-d)+10^2(b-c)+10^1(c-b)+10^0(d-a)\\&=10^3(a-d)+10^2(b-c)+10^1(c-b-1)+10^0(d-a+10)\end{aligned}$$

Now, let's think about the next column, $c-b-1$. We know that $b\geq c$, so $c-b-1$ is *always* gonna be negative. So, we're gonna need to borrow again.

$$\begin{aligned}\overline{abcd}-\overline{dcba}&=10^3(a-d)+10^2(b-c)+10^1(c-b-1)+10^0(d-a+10)\\&=10^3(a-d)+10^2(b-c-1)+10^1(c-b-1+10)+10^0(d-a+10)\end{aligned}$$

Like we said before $b\geq c$, so, if $b\neq c$, then they're difference is at least 1, so $b-c-1\geq 0$, so we won't need to carry.

3.1. **Dealing with c=d.** Let's see about this. If $c=d$, then we would have

$$n=10^3(a-d)+10^2(-1)+10^1(9)+10^0(d-a+10)$$

So, we need to borrow from the 10^3 spot:

$$\begin{aligned}&=10^3(a-d-1)+10^2(-1+10)+10^1(9)+10^0(d-a+10)\\&=10^3(a-d-1)+10^2(9)+10^1(9)+10^0(d-a+10)\\&=\overline{(a-d-1)99(d-a+10)}\end{aligned}$$

remember that this *has* to correspond to the subtraction of the digits ordered. Let's assume that $a-d-1$ is less than or equal to $d-a+10$, all this does is switch the digits:

$$\overline{(a-d-1)99(d-a+10)}=\overline{99(d-a+10)(a-d-1)}-\overline{(a-d-1)(d-a+10)99}$$

For notation, let's just say $d-a+10=x$ and $a-d-1=y$. Now, we get:

$$\overline{y99x}=\overline{99xy}-\overline{yx99}$$

Can x ever be equal to $y-9$? Let's plug our equalities back in

$$\begin{aligned}x&=y-9\\d-a+10&=a-d-1-9\\20&=2a-2d\\10&=a-d\end{aligned}$$

Can the difference of two digits *ever be 10*? Of course not, since a and d are at most 9. So, the max of $a-d$ is 9. The alternate case, where we assume that $d-a+10$ is less than or equal to $a-d-1$ gives:

$$\overline{y99x}=\overline{99yx}-\overline{xy99}$$

This time $x=x-9$, which is just plain false.

Thus, we know that $c\neq d$.

4. The Final System

So, now that we know don't need to carry, since $c > d$, we have our final number's form (if it exists).

$$n = \overline{ABCD} = \overline{abcd} - \overline{dcba} = \overline{(a-d)(b-c-1)(c-b+9)(d-a+10)}$$

Or if we like positional notation:

$$n = \overline{ABCD} = 10^3(a-d) + 10^2(b-c-1) + 10^1(c-b+9) + 10^0(d-a+10)$$

Now, remember that the digits of our original subtraction, $\overline{abcd}$ are the same as the final number, $\overline{ABCD}$, so each small letter corresponds to a big letter, we just gotta figure our which one matches to which. As our analysis concluded, they have to satisfy each of those equations that we derived, in order for this to work. I'm going to rearrange this into a system of equations, to help us figure this out.

5. Solving the System

Let's just assign the equations to the big letters, then we can match the small letters to them. It's fair to point out that we really only want to try plug in letters into equations *with* that letter in it, otherwise we just end up doing a lot of pointless algebra.

$$\begin{aligned} A &= a - d \\ B &= b - c - 1 \\ C &= c - b + 9 \\ D &= d - a + 10 \end{aligned}$$

We can also make some nice observations by adding (4) and (1), since $a - d = -(d - a)$. We can also do this with (2) and (3).

$$\begin{aligned} A + D &= a - d + d - a + 10 = 10 \\ B + C &= b - c - 1 + c - b + 9 = 8 \end{aligned}$$

$$\begin{aligned} &\implies A + D = 10 \\ &\implies B + C = 8 \end{aligned}$$

Alright, so there's a bunch of different combinations to try out. However, let's go ahead and see if we can't organize these terms a little, to reduce the amount of cases we have to try.

In the end, here's the whole system we have to solve:

$$\boxed{\begin{array}{ll} (1) & 9 \le a \le b < c \le d \le 0 \\ (2) & A = a - d \\ (3) & B = b - c - 1 \\ (4) & C = c - b + 9 \\ (5) & D = d - a + 10 \\ (6) & A + D = 10 \\ (7) & B + C = 8 \end{array}}$$

First of all, a good question is: can $A, B, C, D = 0$? If one of them is 0, then it must be equal to d, since it's the smallest possible digit. Even, better by equations (6) and (7), it would give us another digit. If a digit was 0 then we would get:

$$n = \overline{abc0} - \overline{0cba}$$

we have to carry on the ones column and the tens column

$$= \overline{a(b-1)(c-b+9)(-a+10)}$$

Now, let's think can any of those digits be 0, because we need it to appear again. a is greater than d, so, of course, a is not 0. $b-1$ can't be zero, since that means that $b = 1$, but $b > c$, which means that $c = 0$. But then we need another 0 somewhere. This would require that either $c - b + 9 = 0 \implies c - b = -9 \implies c = -10$, or $-a + 10 = 0 \implies a = 10$, which is, of course, impossible. So none of the digits can be 0! So, equation (1) really is $9 \le a \le b < c \le d < 0$.

If none of the digits are 0, then, we can see $a =\neq A$, since $a = a - d$, but this is only true if $d = 0$. So this isn't possible. Likewise, $B \neq b$, since $b = b - c - 1$ since this only works if $c = -1$.

Another thing we can notice is that $A = a - d$ is the biggest digit minus the smallest and $b - c$ is the (maybe) smaller digit minus a (maybe) larger digit. So, does this mean that $A > B$? Let's check:
A problem is that equations (1) and (5) are in terms of a, d, whereas equations (2) and (3) are in terms of b, c. So, we needs to somehow figure out a way for them to 'communicate' with each other, we can relate all the small letters using equation (1). So, we're gonna try that.

$$a - d \overset{?}{>} b - c - 1$$

here's the trick: can we find something bigger (in terms of a, b that's larger than B that's less than A, since then $A < B$. Well, since $a \geq b$, we know that $a - c - 1$ can only be greater than or equal to $b - c - 1$, right?

$$\implies a - c - 1 \geq b - c - 1$$

Now, what else do we know? We know that $c \geq d$, so that means that $a - d - 1$ is greater than or equal to $a - c - 1$

$$\implies a - d - 1 \geq a - c - 1 \geq b - c - 1$$

But look! doesn't $a - d$ have to be *greater* than $a - d - 1$, if d is positive? So:

$$\begin{aligned} a - d > a - d - 1 \geq a - c - 1 \geq b - c - 1 \\ \implies A > a - d - 1 \geq a - c - 1 \geq B \end{aligned}$$

So, it's true that $A > B$. Notice we also found out that $a - d - 1 \geq b - c - 1$. So, $a - d \geq b - c$

I wonder if we can do this with any of the other big letters? It's pretty hard to do with B, C, since they have the same small letters, so there will be a lot of algebra, so let's try C and D. Since we found that $a - d \geq b - c$, we can multiply by -1 and get $d - a \leq b - c$. So, is $C > D$? Notice that C is just the left hand side of this equation plus 9. We want to use that inequality to show some relation between C, D. So that 9 should cancel. Well if we just subtract 1 from D: $C = c - b + 9 \overset{?}{>} d - a + 10 - 1 = D - 1$. Which is just $C > D - 1$. Here's the problem, $C = D - 1$ if $c - b = d - a$, which we know is possible. So, we need to consider that, meaning that we're testing $C \geq D - 1$. It's probably a good idea to test if they can be equal, since if they're not then we also know that $C > D$.

$$c - b = d - a$$

Since $c \geq d$ and $a \geq b$, if $c > d$, b can't be large enough to make this equality true. Likewise, if $a > b$, then c can't be small enough for this equality to hold. So for this to work, we need $c = d$ and $a = b$.

$$\begin{aligned} n &= \overline{aacc} - \overline{ccaa} \\ n &= \overline{(a-c)(a-c)(c-a)(c-a)} \end{aligned}$$

But remember, we need a number, where the first digit is greater than the second, so this *doesn't* work. So $C > D - 1$. This means $C - D > -1$, that is $C \geq D$

Okay, so now we have even more conditions!

$$\boxed{\begin{aligned} A > B \quad (8) \\ C \geq D \quad (9) \end{aligned}}$$

We know for our number, that the first digit must be greater than the second and that the third digit must be greater than or equal to the final digit. From this, we can put together some cases. Like we said before $a \neq A$, so c or b must be the first digit, since the first digit exceeds the second digit. If c is the first digit, then d has the be the second digit. If b is the first digit, then c or d can be the second digit. So either c or d must be the second digit. Then, from equation (9), if b is the first digit, then a has to be the third, since it has to be bigger than the final digit. If c is the first digit, then, again a is the third digit. So, regardless, $a = C$. Okay, so our possibilities are $n = \overline{ABCD} = \overline{bdac}$, or $\overline{cdab}$. Let's starting checking.

5.1. cdab.

So, this one's a little rough, we actually need to do some algebra. If I were to sub $c = A$ into B, that $B = d$ would cancel, so that seems like a good plan.

$$c = a - d$$

plug into equation (2)

$$\begin{aligned} d &= b - (a - d) - 1 \\ &= b - a + d - 1 \\ \implies 0 &= b - a - 1 \\ \implies b &= a + 1 \end{aligned}$$

But that means that $b > a$, which is a contraction. So if this alleged number even exists, it *must* fit the following form: $\overline{bdac}$. So, that doesn't work.

5.2. **bdac.** Alright, get ready, big algebra incoming! Here's our system

$$\begin{aligned} b &= a - d \\ d &= b - c - 1 \\ a &= c - b + 9 \\ c &= d - a + 10 \end{aligned}$$

Now, it's time to get into it! Let's plug the first equation into the second and third one:

$$\begin{aligned} d &= a - d - c - 1 \\ a &= c - (a - d) + 9 = c - a + d + 9 \\ c &= d - a + 10 \end{aligned}$$

Now let's plug that last one into the other two:

$$\begin{aligned} d &= a - d - (d - a + 10) - 1 = a - d - d + a - 11 = 2a - 2d - 11 \\ a &= d - a + 10 + c - a + d + 9 = 2d - 2a + 19 \end{aligned}$$

Now, I have two equations and two unknowns, so this should be immediately solvable. First, I'll rearrange the terms:

$$\begin{aligned} 3d &= 2a - 11 \\ -2d &= -3a + 19 \end{aligned}$$

multiply the first equation by 2 and the second by 3:

$$\begin{aligned} 6d &= 4a - 22 \\ -6d &= -9a + 57 \\ \implies 0 &= -5a + 35 \\ \implies a &= 7 \\ 6d &= 4a - 22 = 4(7) - 22 \\ \implies d &= 1 \\ c &= d - a + 10 = 1 - 7 + 10 = 4 \\ b &= a - d = 7 - 1 = 6 \end{aligned}$$

So $a = 7$, $b = 6$, $c = 4$, $d = 1$. The orientation is $\overline{ABCD} = \overline{bdac} = 6174$. This is Kaprekar's constant. $\square$