Analysis: Foregone Solution

There are at least three different approaches to this problem, and they will have different degrees of success!

Brute force

In test set 1, **N** is less than 100000, so if we want to find two numbers A and B such that A + B = **N**, there are fewer than 100000 possible choices for A. (Once we choose a candidate value for A, we know that B = **N** - A.) We can try all possible choices of A until we find one for which neither A nor B contains a 4. However, we cannot apply this strategy to test set 2, in which we might have to check almost 10^9 values!

Randomized

We can repeatedly choose a candidate value of A uniformly at random from the inclusive range [1, **N**-1], and check whether both it and B = **N** - A are free of the dreaded 4s. But will this randomized approach find a solution fast enough?

How many numbers less than 10^9 do not contain a 4? Consider the set of all 9-digit numbers, with leading zeroes allowed; notice that this set therefore includes all 8-digit, etc. numbers. If we choose a number at random from this set, the probability that it lacks a 4 is the probability that its first digit is not a 4, times the probability that its second digit is not a 4, and so on. Since the probability is 9/10 for any one digit, and the probabilities for different digits are independent, we have an overall probability of $(9/10)^9$, which is about 0.387. In practice, depending on the actual value of **N**, we might have fewer valid choices, but the size of this answer should give us some hope that we can find our A and B quickly. If we presume that the probability that A has no 4s is not too much smaller than our estimate, and that A not having a 4 does not substantially increase the odds that B has a 4, there should be many working values of A out there to find.

We can argue more rigorously. Let D be the number of digits in **N**; we will consider the set of all D-digit numbers less than **N**, with leading zeroes allowed. Suppose that we choose A randomly from that set, and let B = **N** - A, padding with leading zeroes as needed to bring A and B up to D digits.

Consider the last digit of A. If it is a 4, or if it is (**N** - 4) mod 10 (which means that the last digit of B is a 4), then we have failed. But at least 8 of the 10 possible values for the last digit are OK. Then, assuming we have not yet failed, consider the next-to-last digit of A. There is at most one value for that digit that would imply that the next-to-last digit of B is a 4, so at least 8 of the 10 possible values for the next-to-last digit of A are OK. Notice that even though the value of the last digit of A might determine which *particular* value for the next-to-last digit of A would imply that the next-to-last digit of B is a 4, we can still be sure that there exist at least 8 OK values.

And so on, for all D-1 digits up to the first. We only need to worry about the first digit of A or B being a 4 if the first digit of **N** is at least 4, in which case there are certainly at least five possible values for the first digit of A, of which (just as before) at most two can be bad. So the overall probability that a randomly chosen A is OK is no smaller than $(8/10)^{D-1} \times (3/5)$. For D = 9, this is around 0.1, so our randomized approach should only need an expected 10 or so tries to find an answer, and the probability that it will fail to find one after e.g. 1000 tries is vanishingly small (around 10^{-46}).

If you check all values with D digits for some small D, you will find that the worst case (i.e. the one with the fewest solutions) is a 4 followed by D-1 9s. For **N** = 49999999, the fraction of possible choices that will work is around 0.1258, so our bound of 0.1 was not particularly conservative!

However, this approach is highly unlikely to work for test set 3, in which **N** can be on the order of 10^{100}. Our lower bound on the probability that a randomly chosen solution is correct becomes $(8/10)^{99} \times (3/5)$, which is about 1.5×10^{-10}. Even if we try a more clever randomized approach in which we build a random A (less than **N**) out of non-4 digits and then check the corresponding B, which would change those 8/10 factors to 9/10, that bound would be on the order of 10^{-5}, which seems too low to work well on 100 test cases. Can we find a deterministic solution?

Constructive

Observe that we can write **N** in terms of a sum of powers of 10 times its digits. For example, 4837 = 4 × 1000 + 8 × 100 + 3 × 10 + 7 × 1.

Notice that every digit can be written as a sum of two non-4 digits. In particular, we can write 4 = 2 + 2, and for any other digit X, we can write X = 0 + X.

We can apply these observations to solve our problem digit by digit! In the above example, we can decompose 4 × 1000 into 2 × 1000; and 2 × 1000; 8 × 100 into 0 × 100 and 8 × 100, and so on. In this way, we are building up the digits of our desired A and B; A is 2 × 1000 + 0 × 100... = 2000, and B is 2 × 1000 + 8 × 100... = 2837.

In summary, we can set A to be the same as **N** with every 4 replaced with a 2, and every non-4 replaced with a 0. B is simply **N** with every 4 replaced with a 2. Even though we would normally need a "big integer" implementation to handle numbers as large as a googol, in this case we can implement the solution using only strings!