Uncountability of $\mathbb{R}$

S. Kumaresan
School of Math. and Stat.
University of Hyderabad
Hyderabad 500046
kumaresa@gmail.com

The fact that we need the LUB property of the reals to show that the set of real numbers is uncountable is hardly appreciated by many students. Even though most of the students have seen a proof of the uncountability of the reals many are not aware of a subtle point of the proof. Below we give three proofs of this result—a not-so-well-known proof using the LUB property of $\mathbb{R}$ directly, the second one using the nested interval theorem and the third (so-called Cantor's diagonal trick).

Theorem 1. *The set* $[0, 1]$ *is uncountable.*

Proof 1. If $[0, 1]$ is countable, since $[0, 1]$ is infinite, there exists a bijection $f : \mathbb{N} \to [0, 1]$. Let $z_n := f(n)$. We define two sequences (x_n) and (y_n) whose elements are defined recursively. Let x_1 be the z_r where r is the first integer such that $0 < z_r < 1$. Let y_1 be z_s where s is the first integer such that $x_1 < z_s < 1$. Assume that we have chosen $(x_i)_{i=1}^n$ and $(y_i)_{i=1}^n$ with the property

$$0 = x_0 < x_1 < x_2 < \cdots < x_n < y_n < y_{n-1} < \cdots < y_2 < y_1 < y_0 = 1.$$

We choose x_{n+1} to be the z_r where r is thefirst integer such that $x_n < z_r < y_n$. Let y_{n+1} be z_s where s is the first integer such that $x_{n+1} < z_s < y_n$. Clearly the set $\{x_n\} \subset [0, 1]$ is nonempty and bounded above. Let $x := \sup\{x_n\}$. Then it is easily seen that $x \in [0, 1]$ and that $x \neq z_n$ for $n \in \mathbb{N}$. Details! □

Proof 2. We wish to show that the set of reals in $[0, 1]$ is uncountable. Assume the contrary that $[0, 1]$ is countable. That is, there exists a map $f : \mathbb{N} \to [0, 1]$ which is onto. Let $x_n := f(n) \in [0, 1]$. To arrive at a contradiction, we shall employ the Nested interval theorem. Let us subdivide the interval $[0, 1]$ into three closed subintervals of equal length. Then there exists at least one subinterval which does not contain $f(1) = x_1$. (Draw a picture. The worst possible case is when $f(1) = x_1$ happens to be either $1/3$ or $2/3$.) Call this subinterval as J_1. Let us subdivide J_1 into three equal closed subintervals. Either $x_2 = f(2) \notin J_1$ or it lies in at most two of the subintervals of J_1. In any case there exists a subinterval of J_1 which does not contain $f(2)$. Choose one such and call it J_2. We proceed along this line to construct a subinterval J_n of J_{n-1} which does not contain $f(n)$. Thus we would have obtained a nested sequence (J_n) of closed and bounded intervals. By the nested interval theorem their intersection is nonempty: $\cap_n J_n \neq \emptyset$. Let $x \in \cap J_n$. Then $x \in [0, 1]$. So there exists an $n \in \mathbb{N}$ such that $x = f(n)$. Thus $f(n) = x \in \cap_n J_n$. This contradicts our choice of J_n. Hence our assumption that $[0, 1]$ is countable is wrong. □

Proof 3. This is the "standard" proof found in almost all text-books. The idea behind this proof is simple: If $[0,1]$ is written as a sequence (x_n) we write x_n in decimal form and construct a number $y = 0.y_1y_2\cdots y_n\cdots$ such that y and x_n differ at their n-th decimal digit. But however to make this idea work one should exercise care.

Let $x_n = x_{n1}x_{n2}\cdots x_{nn}\cdots$ be a decimal expansion of x_n. We choose $y_n \in \{0,1,\ldots,9\}$ but $y_n \neq x_{nn}$. We define $y = 0.y_1y_2\cdots y_n\cdots$. By LUB there exists a real number $y \in [0,1]$. For, if we denote $s_n := \sum_{k=1}^n y_k/10^k$ then $\{s_n\}$ is a subset of nonnegative real numbers bounded above by 1: $s_n \leq \sum_k 9/10^k = 1$. Thus by LUB property there exists a real number y. From the theory of infinite series this y is the sum of the infinite series $\sum_{k=1}^\infty y_k/10^k$.

Now the usual argument goes as follows: This number y must be x_n for some n. But then this is impossible as y and x_n differ at the n-th decimal digit. This argument is not completely correct for the simple reason that some real numbers have more than one decimal expansions. For instance the number y under consideration could have been $0.499\cdots9\cdots$ which represents $1/2$. But $1/2$ might appear in our sequence as $0.500\cdots0\cdots$. Thus we cannot really conclude the number y we constructed is not in our list!

The way out is to allow only special kind of decimal expansions. One knows that the decimal expansion is unique for any $x \in [0,1]$ unless it is of the form $k/10^n$ for some $k \in \mathbb{N}$ and $n \in \mathbb{N}$ with $k \leq 10^n$. In this exceptional case there are exactly two expansions—one which ends in zeros and the other which ends in 9's. The first one is called the terminating decimal expansion and the second the nonterminating one. Thus we observe the following:

1. The correspondence between the real numbers and the decimal expansions in which we choose the terminating one if there are two expansions is one-to-one and onto.

2. The correspondence between the real numbers and the decimal expansions in which we choose the non-terminating one if there are two expansions is one-to-one and onto.

Now we return to the proof. Assume we adopt the first convention above, i.e., we choose the terminating decimal expansions if there are more than one decimal expansions. In this case we take $y_n = 2$ if $x_{nn} \neq 2$ and $y_n = 5$ if $x_{nn} = 2$. Then $0.y_1y_2\cdots y_n\cdots$ is a decimal expansion which does not end in an infinite sequence of 9's. Hence $y := 0.y_1y_2\cdots y_n\cdots$ is a real number in $(0,1)$ which is not equal to any of the x_n's.

If we decide to adopt the second convention of representing reals in $(0,1]$ by nonterminating decimal expansion, then we choose $z_n = x_{nn}+1$ if $x_{nn} < 5$ and $z_n = x_{nn}-1$ if $x_{nn} \geq 5$. Then $0.z_1z_2\cdots z_n\cdots$ is a decimal expansion of the form we agreed up on representing a real number $z \in [0,1]$. Clearly $z \neq x_n$ for any n. $\square$