

# Re-Enumerating the Rationals

BRYAN LU

2 May 2022

Recall the standard proof of showing that  $|\mathbb{Q}^+| = |\mathbb{N}|$ :

*Proof.* Consider  $\mathbb{N} \times \mathbb{N}$ , and consider the mapping  $d : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{Q}$  that sends  $(a, b) \mapsto \frac{a}{b}$ , as represented by this table:

$d$	1	2	3	4	5
1	$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$
2	$\frac{2}{1}$	$\frac{2}{2}$	$\frac{2}{3}$	$\frac{2}{4}$	$\frac{2}{5}$
3	$\frac{3}{1}$	$\frac{3}{2}$	$\frac{3}{3}$	$\frac{3}{4}$	$\frac{3}{5}$
4	$\frac{4}{1}$	$\frac{4}{2}$	$\frac{4}{3}$	$\frac{4}{4}$	$\frac{4}{5}$
5	$\frac{5}{1}$	$\frac{5}{2}$	$\frac{5}{3}$	$\frac{5}{4}$	$\frac{5}{5}$

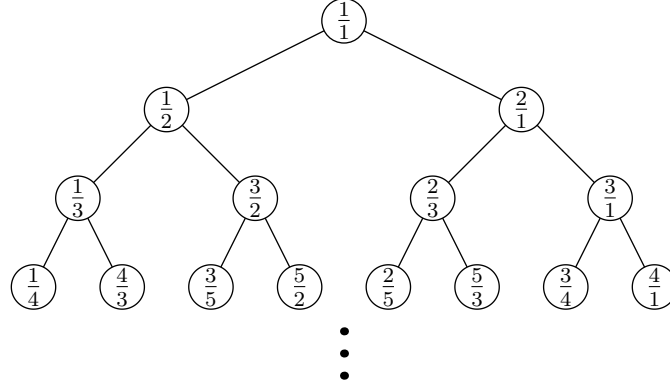
Note obviously that  $d$  is surjective. Now, consider the enumeration of  $\mathbb{N} \times \mathbb{N}$  by going along diagonals, i.e. in the ordering  $(1, 1), (1, 2), (2, 1), (3, 1), (2, 2), (1, 3), (1, 4), \dots$  etc. (This ordering defines some map  $f : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$ .) Since  $\mathbb{N} \times \mathbb{N}$  is countable and  $\mathbb{Q}$  embeds into  $\mathbb{N} \times \mathbb{N}$  through  $d$ , we see that  $\mathbb{Q}$  must also be countable.  $\square$

However, what if we wanted to compute this enumeration for  $\mathbb{Q}^+$ , i.e. compute an explicit bijection between  $\mathbb{N}$  and  $\mathbb{Q}^+$ . Doing it this way is hard, as we would have to compute  $f(n)$  for arbitrary  $n$ , and ignore the ones that we've already seen. The naive way takes exponential time in the amount of data needed express  $n$  – linearly traversing this grid is obviously linear in  $n$ , but if we only need  $\log_2 n$  bits to express  $n$ , then we see this algorithm actually runs in exponential time. This is especially because we have to consider duplicates of rationals already seen, which are annoying to deal with. As a CS major, we should really be upset by this and want to find a different enumeration of  $\mathbb{Q}^+$ . This lecture shows a different approach to showing  $|\mathbb{N}| = |\mathbb{Q}^+|$  using the Calkin-Wilf Tree (2000).

As an aside, this tree has been mentioned in the 20th century as the *Raney Tree* by Berstel and de Luca, and a similar tree has shown up in Kepler's *Harmonices Mundi*. The enumeration of rationals we will end up getting is also well-known, called *Stern's diatomic series*.

## 1 The Calkin-Wilf Tree

Consider the following complete binary tree, generated as follows. Start with a root labeled with the rational number  $\frac{1}{1}$ . Then, for every vertex in this tree labeled with the rational  $\frac{p}{q}$ , create a left child labeled with the rational  $\frac{p}{p+q}$ , and a right child labeled with the rational  $\frac{p+q}{q}$ . Let's see what happens when we grow this tree for the first few levels:



Note that left children are always less than 1 and right children are always greater than 1. Moreover, for a node labeled  $\frac{x}{y}$ , its parent is  $\frac{x}{y-x}$  if  $y > x$  or  $\frac{x-y}{y}$  if  $x > y$ .

First, we claim that this tree generates  $\mathbb{Q}^+$  exactly:

**Theorem 1 (New Enumeration of  $\mathbb{Q}^+$ )**

The following is true about the Calkin-Wilf Tree:

1. Every rational number appears once and exactly once in the tree.
2. Moreover, every rational appears in lowest terms, i.e. if  $\frac{p}{q}$  is generated by this tree, then  $\gcd(p, q) = 1$ .

*Proof.* We prove both statements:

1. To prove this, we impose a total ordering on  $\mathbb{Q}^+$ . We will say  $\frac{p}{q} <' \frac{r}{s}$  in our ordering if  $q < s$ , or  $q = s$  and  $p < r$ .

We first prove that every rational number appears at least once. Assume for the sake of contradiction the set  $N$  of all positive rationals not in the tree is non-empty. We can totally order the set  $N$  using the inherited total order  $<'$ , and since  $1 = \frac{1}{1}$  is a minimal element, we know there must be a minimal element  $\frac{x}{y}$  in  $N$  with respect to this order by Zorn's Lemma.

First, if  $x > y$ , then consider  $\frac{x-y}{y}$ . This can't be in the tree because otherwise, its child  $\frac{x}{y}$  would be in the tree, which it isn't. However, it is smaller than  $\frac{x}{y}$  with respect to  $<'$ , contradiction, so there could not have been a minimal element of  $N$ . A similar analysis holds if  $x < y$  – consider then  $\frac{x}{y-x}$ , the parent of  $\frac{x}{y}$  in the tree. This is also less than  $\frac{x}{y}$  in this order, so again we get a contradiction.

Now, we show that each element appears at most once. Take  $\frac{x}{y}$  as above, but instead let  $T$  be the set of all rationals that appear at least twice in the tree is non-empty. By the same ordering that we endowed  $T$  in the last proof, we take the minimal element of  $T$ . Again, when considering the parent of this number in the tree, we note that this can either be  $\frac{x}{y-x}$  or  $\frac{x-y}{y}$ , which must also appear at least twice, so either must appear in  $T$ . However,  $x - y < x$  and  $y - x < y$ , so these parents would have appeared before  $\frac{x}{y}$  in  $T$ , which is a contradiction.

2. Same construction as above: take  $G$  to be the set of  $\frac{x}{y}$  that appear in the tree but not in lowest terms, i.e. the fraction is generated as some  $\frac{x}{y}$  such that  $\gcd(x, y) = k > 1$ , so we may write  $\frac{x}{y} = \frac{kx'}{ky'}$  where  $\gcd(x', y') = 1$ . The parent of this node either must be  $\frac{kx'}{k(y'-x')}$  or  $\frac{k(x'-y')}{ky'}$ , but then the numerator and denominator must also have a greatest common divisor of  $k$ , which again is a contradiction. Thus, all fractions  $\frac{x}{y}$  in the tree are in lowest terms.

□

This has an immediate corollary:

### Corollary 2

$\mathbb{Q}^+$  is countable, ie. the set of all positive rationals has the same size as  $\mathbb{N}$ .

*Proof.* The function  $f : \mathbb{Z}^+ \rightarrow \mathbb{Q}^+, \forall n \in \mathbb{Z}, f(n) = q_n$  is a bijection by Our Sequence Has It All (Theorem 1), meaning that every positive integer is mapped to exactly one positive rational, and vice versa.  $\square$

Note that we could have not used the tree at all in our analysis, but the recurrence relation of  $b(n)$  made the tree a very natural way to present the sequence  $\{q_n\}_{n=1}^\infty$ , which I find rather neat. Furthermore, the tree sort of gives us a different explicit ordering of the rationals other than that used in the diagonalization argument, which I think is really cool.

We still haven't specifically shown how to construct the  $n$ th element in the sequence  $\{q_n\}_{n=1}^\infty$ , so I'll state a method here for others to verify/prove. To get the  $n$ th number in the sequence, consider writing  $n$  in regular binary, running a run-length encoding of the binary string backwards, meaning that starting from the back, we write down the length of strings of consecutive digits. The list of numbers that you write down, when written as the sequence for a continued fraction, gives the  $n$ th element in the sequence. For example, here is the 17th element:

$17 = 10001_2 \rightarrow \text{reverse run-length encoding: } 131$

$$[1; 3, 1] = 1 + \frac{1}{3 + \frac{1}{1}} = 1 + \frac{1}{4} = \frac{5}{4}$$

The *HRT* has a real name in the literature – it is called the *Calkin-Wilf Tree*. The *HRT* is a fairly new tree, as it comes from a paper published in 2000, so there's still a lot to investigate about the *HRT*! See what else you can find about this interesting tree. :)

## 2 Binary Numbers, But Better

First, we look at a completely unrelated topic, which centers on representing integers in binary, while allowing some rules to be relaxed. Traditional binary requires every digit can only be 1 or 0, and the  $i$ th digit from the right of the number represents a value of  $2^{i-1}$ . This is the normal binary representation that we will henceforth abandon.

Here is the question that we will tackle first:

**Question.** Suppose now we write integers in the same style as binary, with each place value representing the same amount per place, but we are now allowed to use the digit 2. Let's call this system **hyperbinary**. How many ways are there to write  $n$  in hyperbinary?

Notice that hyperbinary makes the representation of an integer  $n$  non-unique, unfortunately, as one can see by the case of  $n = 4$ :

$$4 = 1 \cdot 2^1 + 2 \cdot 2^0 = 1 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0 = 2 \cdot 2^1 + 0 \cdot 2^0$$

To make some notation, let's put the subscript *hb* below a number to indicate that it is written in hyperbinary. Here is 4 again, in all of its forms:

$$4 = 12_{hb} = 100_{hb} = 20_{hb}$$

Let's suppose that  $b(n)$  is the number of ways we can represent  $n$  in hyperbinary. It's not obvious how  $b(n)$  depends on  $n$ , but clearly  $b(n) \geq 1$  for all  $n$ , as at least  $n$  has a unique regular-binary representation.

First, we can compute some of the easier cases for small  $n$ :

$n$	$b(n)$	Hyperbinary Representations	$n$	$b(n)$	Hyperbinary Representations
0	1	$0_{hb}$	5	2	$101_{hb}, 21_{hb}$
1	1	$1_{hb}$	6	3	$110_{hb}, 102_{hb}, 22_{hb}$
2	2	$10_{hb}, 2_{hb}$	7	1	$111_{hb}$
3	1	$11_{hb}$	8	4	$1000_{hb}, 200_{hb}, 120_{hb}, 112_{hb}$
4	3	$100_{hb}, 20_{hb}, 12_{hb}$	9	3	$1001_{hb}, 201_{hb}, 121_{hb}$

*Solution.* Notice that all odd integers of the form  $2k + 1$  must still always have hyperbinary representations that always end in the digit 1, by parity. If we were to remove the last digit, we would have exactly one hyperbinary representation for the integer  $k$ . This implies that

$$b(2k + 1) = b(k).$$

We can also see that for an even integer  $2k$ , the last digit can either be 0 or 2. If the last digit was 0, removing it gives a hyperbinary representation for  $k$ , and similarly removing a 2 gives a hyperbinary representation for  $k - 1$ . Thus,

$$b(2k) = b(k) + b(k - 1).$$

□

This seems to describe very little about the asymptotic behavior of the sequence, but it tells us enough for us to proceed.

Our first method of showing that the rationals are countable is by considering the sequence  $\{q_n\}_{n=1}^{\infty} = \left\{ \frac{b(n)}{b(n+1)} \right\}_{n=1}^{\infty}$ . The first few terms of this sequence look like this:

$$\{q_n\}_{n=1}^{\infty} = \frac{1}{1}, \frac{1}{2}, \frac{2}{1}, \frac{3}{3}, \frac{2}{2}, \frac{3}{3}, \frac{1}{1}, \frac{4}{4}, \dots$$

Wait, if we run a level-order traversal on this tree, going from the left to the right, we appear to listing out the sequence  $\{q_n\}_{n=1}^{\infty}$  in order. That's rather unexpected! Here is our proposition/conjecture:

**Theorem 3 (This Tree is Hyperbinary)**

The sequence  $\{q_n\}_{n=1}^{\infty} = \left\{ \frac{b(n)}{b(n+1)} \right\}_{n=1}^{\infty}$  is created by a level-order traversal on the vertices of this tree.

To show that we should have expected this result, we'll look at how our tree was grown, and argue inductively (and strongly).

*Proof.* Our base case is simply the fraction at the root,  $\frac{1}{1}$ , which is also  $\frac{b(0)}{b(1)}$ . This is the first element of the sequence,  $q_1$ , which appears first in the level-order traversal.

Now, suppose we consider a vertex of the tree that has the label  $\frac{b(n)}{b(n+1)}$ . By this construction, the left child has label

$$\frac{b(n)}{b(n) + b(n+1)} = \frac{b(n)}{b(2n+2)} = \frac{b(2n+1)}{b(2n+2)},$$

so we see this is an element of the sequence  $\{q_n\}_{n=1}^{\infty}$ . We can show that this node will be reached as the  $(2n+2)$ th element in our level-order traversal, by a direct counting argument. Notice that since the tree is complete, at level  $n$  (where level 0 only includes the root) there are  $2^n$  vertices, so the left-most vertex of level  $n$  is the  $2^n$ th vertex to be traversed. For the  $n$ th vertex traversed, which happens to be in level  $k$ , there are  $n - 2^k$  nodes in level  $k$  that are traversed before it, so in level  $k + 1$ , the left child of  $n$  will be the  $[2^{k+1} - 1 + 2(n - 2^k) + 1]$ th  $= (2n)$ th

vertex traversed. This is sufficient to show that for the vertex labeled  $\frac{b(n)}{b(n+1)}$ , its left child will be the  $(2n+2)$ th vertex traversed.

Similarly, the right child of the vertex labeled  $\frac{b(n)}{b(n+1)}$  is

$$\frac{b(n) + b(n+1)}{b(n+1)} = \frac{b(2n+2)}{b(2n+3)},$$

and we know that this will be the  $(2n+3)$ th vertex traversed, as it is traversed right after the left child. This shows that the sequence  $\{q_n\}_{n=1}^{\infty} = \left\{ \frac{b(n)}{b(n+1)} \right\}_{n=1}^{\infty}$  appears in this order on the vertices of the tree.  $\square$

To finish, we appeal to the proposition This Tree is Hyperbinary, so as every element in the sequence  $\{q_n\}_{n=1}^{\infty}$  corresponds to one vertex in  $HRT$ , and  $HRT$  contains all of the rational numbers in lowest form exactly once, so does the sequence  $\{q_n\}_{n=1}^{\infty}$ .

in lowest terms, ie. the set  $S = \left\{ \frac{p}{q} \mid p, q \in \mathbb{Z}^+, \gcd(p, q) = 1 \right\}$ , is equal to the set of elements in  $\{q_n\}_{n=1}^{\infty}$ .

### 3 Continued Fractions