Calkin-Wilf Trees

Scribe: Sohom Paul

Date: Thursday, April 4, 2019

1 Hyperbinaries

Example 1. What is the 23rd rational positive positive number?

Solution We need an enumeration of the rational numbers!

We introduce the hyperbinary numbers, where we write binary numbers, but allow for multiple representation of numbers by also letting us use the number 2. For example,

$$\begin{aligned} 4 &= 100 \\ &= 20 \\ &= 12 \end{aligned}$$

Let us count the number of ways we can write integer n in hyperbinary, $b(n)$. If we sum $b(n)$, restarting every time we hit a 1, we get powers of 3 because it's just showing the number of strings of 3 different digits. To find a recurrence for $b(n)$, let's casework on the last digit:

- The only way to make an odd number is to take a number and append a 1: $b(2n+1) = b(n)$
- The only ways to make an even number is to take a number and append a 0 (which doubles it) or a 2 (which doubles it at add 2): $b(2n+2) = b(n+1) + b(n)$

Donald Knuth called this recurrence $fusc(n)$ (fusc for "obfuscate"). Let's return to our original problem of finding the 23rd rational positive number. Define the nth rational number to be $b(n)/b(n+1)$. This list contains every rational in reduced form exactly once.
Let's define the *Calkin-Wilf tree* to be a complete binary tree that has as node n the rational b_n/b_{n+1}. Notice that each node that contains b_n/b_{n+1} has as its left child b_{2n+1}/b_{2n+2} and right child b_{2n+2}/b_{2n+3}. Notice that applying the recurrence that we had earlier yields relations

$$\frac{b_{2n_1}}{b_{2n+2}} = \frac{b_n}{b_n + b_{n+1}}$$
$$\frac{b_{2n_2}}{b_{2n+3}} = \frac{b_n + b_{n+1}}{b_{n+1}}$$

Thus the parent of any node containing x/y is either $x/(y-x)$ or $(x-y)/x$ depending on whether it is a left or right child. However, only one of these numbers is positive, so we can deduce the parent of any node.

We know make the following claims:

Theorem 1.1. $\frac{x}{y} \in T, \frac{x}{y} \neq \frac{1}{1} \implies \gcd(x,y) = 1$

Proof. Assume that $\gcd(x,y) = k > 1$ for at least one $\frac{x}{y} \in T$. Let $\frac{kx'}{ky'}$ be the highest node in the tree. However, by our algorithm for finding the parent of any node, we know the parent must have greatest common divisor at least equal to k because it is $\frac{kx'}{k(y'-x')}$ or $\frac{k(x'-y')}{ky'}$. This contradicts our assumption. ∎

Theorem 1.2. $\frac{x}{y} \in \mathbb{Q}^+$, $\gcd(x,y) = 1 \implies \frac{x}{y} \in T$

Proof. Let S be the set of all rationals not in the tree. Let y be the smallest denominator in S and let x be minimal numerator of a number in S with denominator y. If $x > y$, then $\frac{x-y}{y}$ could not be in the tree because it would have child $\frac{x}{y}$ which, by assumption, isn't in the tree. But $\frac{x-y}{y}$ can't be in S because the way we choose our x and y Otherwise, if $x < y$, consider instead $\frac{x}{y-x}$ ∎

Theorem 1.3. *Every rational that appears in the tree appears at most once.*

Proof. Using the same rules as the previous proof, we find the "smallest" x/y that appears twice in the tree. However, it must have its parent appear twice, which would be yet "smaller". ∎

Theorem 1.4. *The Calkin-Wilf tree contains every reduced positive rational exactly once.*

2 A nice theorem

Definition. *A* runlength encoding *takes a number applies the "look-and-say" rule: see the length of each block of repeated digit, then write down its length and the digit.* $444411002 \to 44212012$

Theorem 2.1. *Consider, starting from the least significant digit of the binary representation of any positive integer n, the length of each run. Letting this be the continued fraction representation of a rational, this gives the nth rational under the ordering given by the Calkin-Wilf tree.*