1 Coming up with proofs

"How does one come up with proofs?"

This is a deep question which I have been asked many times, and I have asked myself many times. But, unfortunately, we don't know a satisfactory answer to this. What I can say with some degree of confidence is the final "proof" that one reads in books/notes/papers, is probably not the first thing that the author wrote. Proofs are written in drafts, and the first draft, the one that really "comes up" with the proof is often not written up. And therefore, one doesn't really know the process behind "how to come up with proofs".

There is good reason why it is not written up – it doesn't make exciting reading. Nevertheless, let me illustrate a possible "first draft" (more aptly, a "stream of consciousness dump") Here is a rather non-trivial theorem (which can be found in your UGP1).

Theorem 1. Let $H(n) = \sum_{i=1}^{n} \frac{1}{i}$. Prove $H(n) = \Theta(\log_2 n)$.

A First Draft, aka, a transcript of my thoughts as I think it would have been.

We need to prove $H(n) = \Theta(\log_2 n)$. So we have to prove $H(n) = O(\log_2 n)$ and $H(n) = \Omega(\log_2 n)$. Let us take things one at a time.

To prove $H(n) = O(\log_2 n)$ we need to find two positive constants a, b such that for ***all*** $n \geq b$, we have $H(n) \leq a \cdot \log_2 n$. Why should this be true? Hmm. When is $\log_2 n$ even a "nice number". Ah, when n is a power of 2. So, let's simplify our life and restrict attention to powers of 2. If we can handle this, then we can think about the other ns. So, let's suppose $n = 2^k$ for some number k.

Ok, what do we have to show then? We need to show $H(n) = H(2^k) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{2^k} \leq a \cdot k$. Hmm. What should a really be? Let us try out some concrete numbers to get an idea. Say $k = 2$, that is, $n = 4$. Then, $H(4) = 1 + 1/2 + 1/3 + 1/4 = \frac{25}{12}$ which is close to k but slightly bigger. Say $k = 3$, that is $n = 8$. Then, $H(8) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{8}$. That (I calculate on the calculator) is: 2.717.... Definitely smaller than k. One last try. $k = 4$, that is $n = 16$. Then, $H(16) = 3.3807...$. Smaller than k. Can it be that $a = 1$, when, say, $k \geq 3$?

> **Remark:** *Till now, we have done a bit of manual calculation, and tried to "guess" a pattern (a=1). At this point, we get an **idea**. How do ideas come to us, I don't know, but I believe they do come if one keeps at it. On this topic, see this excellent video after reading this note.*

We see that $H(2^k)$ is a sum of a bunch of fractions of the form $1/i$, while the right hand side that we are comparing $H(n)$ with is k, which is an integer. Which fractions, of the form $1/i$, add up to give an integer? Clearly, if we sum up the fraction $1/i$, i times, then we do get an integer (we get 1). But we don't have i many $1/i$'s, but rather... a ha! (this is where the idea strikes) We can write:

$$H(8) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}$$

and then "group" these fractions up as

$$H(8) = \frac{1}{1} + \left(\frac{1}{2} + \frac{1}{3}\right) + \left(\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right) + \frac{1}{8}$$

and then ***upper bound*** each term in a parenthesis term by the *first* term in that group. We can do this because fractions get smaller as denominators get bigger. More precisely, we observe

$$H(8) \leq \frac{1}{1} + \left(\frac{1}{2} + \frac{1}{2}\right) + \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}\right) + \frac{1}{8}$$

where note that the fractions in each parenthesized term has been replaced by the first fraction. Thus, we get that $H(8) \leq 1+1+1+\frac{1}{8} \leq 4$. And similarly, $H(16) \leq 1+1+1+1+\frac{1}{16}$. More generally, this will give $H(2^k) \leq k + \frac{1}{2^k} < k+1$. So, we haven't quite prove $\leq k$, but have proved $< k+1$ which will be good enough for our Big-Oh result.

In fact, the same idea above (of grouping fractions together and then replacing them by the same fraction) also allows us to *lower bound* $H(k)$. This is because we can ***lower bound*** each term in a parenthesis term by the *first* term in the ***next*** group. For instance, we can us the grouping of $H(8)$ as above, and then lower bound with

$$H(8) \geq \frac{1}{1} + \left(\frac{1}{4}+\frac{1}{4}\right) + \left(\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}\right) + \frac{1}{8}$$

Each parenthesized term is now $1/2$ instead of 1 giving us $H(8) \geq \frac{1}{2}+\frac{1}{2}$ (plus some other stuff which we are just ignoring). If we did this with 16, we would get $H(16) \geq \frac{1}{2}+\frac{1}{2}+\frac{1}{2}$. More generally, $H(2^k) \geq \frac{k-1}{2}$.

Thus, when $n = 2^k$ is a power of 2, then what we get is $\frac{k-1}{2} \leq H(2^k) \leq k+1$. This proves that $H(2^k) = \Theta(k)$, which is what we really wanted.

Remark: *At this point, we feel we are really close to the full solution. Sure there are these non-powers of* 2 *to take care of, but the idea above felt pretty powerful.*

Next, we realize that $n = 2^k$ was not really a big handicap. Indeed, if n is *not* a power of 2 but rather $2^k < n < 2^{k+1}$ (and thus $k = \lfloor \log_2 n \rfloor$). Then (a) $H(n) > H(2^k)$ (since the LHS has more positive terms) and so $H(n) > \frac{k-1}{2} = \frac{\lfloor \log_2 n \rfloor}{2}$, and (b) $H(n) < H(2^{k+1}) < k+2 = \log_2 n + 2$. Thus, $H(n) = \Theta(\log_2 n)$. That's it! We proved it! Yippee!

So that was the first draft. Next, we need to make sure we write it for the eyes of the public. So we edit, pare, re-read, re-write, and finally in the fourth draft, produce something like this[1] which is in the UGP1 solutions.

Proof. To prove $H(n) = \Theta(\log_2 n)$, we need to show three constants b, a_1, a_2, such that

$$\forall n > b, \qquad a_1 \log_2 n \leq H(n) \leq a_2 \log_2 n$$

Let $L := \log_2 n$, and let $\ell = \lfloor L \rfloor$. Thus, we get

$$2^\ell \leq n < 2^{\ell+1}$$

Now note that

$$\begin{aligned} H(n) &= \frac{1}{1} + \left(\frac{1}{2}+\frac{1}{3}\right) + \left(\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}\right) + \frac{1}{8} \cdots + \frac{1}{n} \qquad (1) \\ &< \frac{1}{1} + \left(\frac{1}{2}+\frac{1}{2}\right) + \\ &\quad \left(\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}\right) + \ldots \\ &\quad + \left(\frac{1}{2^\ell} + \cdots + \frac{1}{2^\ell}\right) \quad \text{potentially adding some terms in the final parenthesis} \end{aligned}$$

[1]Full Disclosure: I am not claiming this is the best way to write this proof. But, I believe it is adequate. Also, the stuff written below is not my fourth draft. One gets better with this as time moves along. One really "writes" the first draft "in ones head". But that takes time.

Each term in the parentheses adds up to 1, and there are $\ell+1$ parentheses in all. Therefore, $H(n) \leq \ell+1 = \lfloor \log_2 n \rfloor + 1$. Similarly,

$$\begin{aligned} H(n) &= \frac{1}{1} + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) \cdots + \frac{1}{n} \\ &\geq \frac{1}{1} + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \\ &\quad \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \dots \\ &\quad + \left(\frac{1}{2^\ell} + \cdots + \frac{1}{2^\ell}\right) \quad \text{perhaps dropping some terms.} \end{aligned} \tag{2}$$

where we have used $\frac{1}{i} \geq \frac{1}{2^c}$ where 2^c is the smallest power of 2 which is $\geq i$. Each parenthesis above equates to $1/2$, and there are $(\ell - 1)$ of them. This gives $H(n) \geq \frac{\ell+1}{2} = \frac{\lfloor \log_2 n \rfloor + 1}{2}$.

In sum, we have proven that for any n,

$$\frac{\lfloor \log_2 n \rfloor + 1}{2} \; < H(n) \; < \lfloor \log_2 n \rfloor + 1$$

Both the first and the third term are $\Theta(\log_2 n)$, implying the "sandwiched" $H(n)$ is also $\Theta(\log_2 n)$. $\square$