Algorithms
Computer Science 140 & Mathematics 168
Spring 2012
Writing Inductive Proofs
Ran Libeskind-Hadas

A good proof must be both correct and well-written. This short document reviews some important tips on writing clear and correct proofs by induction.

1 First Example

Let's begin with a simple arithmetic result evidently discovered by Gauss when he was a schoolboy.

Theorem 1. $\forall n \geq 1$

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

We'll begin with a correct and clearly-written proof of this fact.

Proof. The proof is by induction on n.

Basis: $n = 1$. In this case, the left-hand side is $\sum_{i=1}^{1} i = 1$ and the right-hand side is $\frac{1(1+1)}{2} = 1$ and thus the claim holds for $n = 1$.

Inductive Hypothesis: Assume that the statement is true for $n = k$. In other words, assume that

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$$

Inductive Step: We must show that $\sum_{i=1}^{k+1} i = \frac{(k+1)(k+2)}{2}$. By the inductive hypothesis,

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$$

Therefore, adding $k+1$ to both sides we get

$$\begin{aligned}\left(\sum_{i=1}^{k} i\right) + (k+1) &= \frac{k(k+1)}{2} + (k+1) \\ \sum_{i=1}^{k+1} i &= \frac{k(k+1)}{2} + \frac{2(k+1)}{2} = \frac{(k+2)(k+1)}{2} = \frac{(k+1)(k+2)}{2}\end{aligned}$$

□

Note that in this proof, we introduced the variable k in the inductive hypothesis. This is a stylistic choice and is not at all necessary. Some people feel that this contributes to the clarity of the proof. Others would choose not to introduce this extra variable and would simply state the inductive hypothesis as "Assume that the statement is true for n. In other words, $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$." Then, in the inductive step they would say, "We must show that $\sum_{i=1}^{n+1} i = \frac{(n+1)(n+2)}{2}$." Either way of writing the proof is considered entirely acceptable.

Next, let's look at a "bad" proof of this same theorem. After the proof, I'll explain what's "bad" about it.

Proof. The proof is by induction on n.

Basis: $n = 1$. In this case, $\sum_{i=1}^{1} i = 1$ and $\frac{1(1+1)}{2} = 1$ and the statement of the theorem holds.

Inductive Hypothesis: Assume that the statement is true for $n = k$. In other words, assume that

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$$

Inductive Step: We must show that $\sum_{i=1}^{k+1} i = \frac{(k+1)(k+2)}{2}$.

$$\begin{aligned}\sum_{i=1}^{k+1} i &\stackrel{?}{=} \frac{(k+1)(k+2)}{2} \\ (\sum_{i=1}^{k} i) + (k+1) &\stackrel{?}{=} \frac{(k+1)(k+2)}{2} \\ \frac{k(k+1)}{2} + (k+1) &\stackrel{?}{=} \frac{(k+1)(k+2)}{2} \quad \text{by the Inductive Hypothesis}\end{aligned}$$

$$\begin{aligned} \frac{k(k+1)}{2} + \frac{2(k+1)}{2} &\stackrel{?}{=} \frac{(k+1)(k+2)}{2} \\ \frac{(k+2)(k+1)}{2} &\stackrel{?}{=} \frac{(k+1)(k+2)}{2} \end{aligned}$$

Since the left-hand side and the right-hand side of the last expression are equal, the proof is complete. □

What's wrong with this proof? First of all, there is no mathematical symbol $\stackrel{?}{=}$. Using that notation is analogous to writing an ungrammatical sentence. Second, we are starting with what we want to show and deducing something true. This is risky, since it relies on an implicit "if and only if" between each line in the proof above. Although it happens to work out here, imagine that we had done an algebraic step that was not "if and only if". For example, if we had multiplied both sides of one of the lines by 0, we would have deduced in the end that $0 = 0$ which is true. However, we cannot infer from this that the original statement was true.

It's totally fine to use this kind of reasoning in the privacy of your own notes as you are developing the proof. However, when you write up the proof for the world to read, the exposition should be formal and precise. This second proof doesn't cut it.

2 More Subtle Induction

Not all inductive proofs will be arithmetic or algebraic in nature. In fact, most proofs in this course will be about data structures or graphs or other mathematical structures. Let's look at another theorem and two proofs of that theorem: this time a bad one first and then a good one.

Recall the notion of binary trees from CS 60. One node is designated as the *root*. A node of degree 0 or 1 is called a *leaf*. A binary tree in which each internal node ("non-leaf node") has exactly two children is called a *regular binary tree*. Recall also that the *depth* of a node is the distance from the root to that node. For example, a regular binary tree is shown in the figure below.

Theorem 2. *In every regular binary tree with n leaves, the number of internal nodes is $n-1$.*

Proof. The proof is by induction on the number of leaves, n, in the tree.

Basis: When $n = 1$ the tree comprises one leaf and no internal nodes. Thus, the theorem holds in this case.

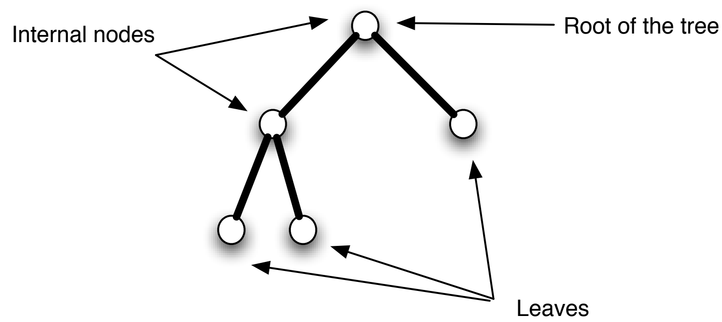

Figure 1: A regular binary tree.

Inductive Hypothesis: Assume that in every regular binary tree with n leaves the number of internal nodes is $n - 1$.

Inductive Step: We must show than in every regular binary tree with $n+1$ leaves there are exactly n internal nodes. Take a tree T with n leaves. By the inductive hypothesis, T has $n - 1$ internal nodes. Now, select some leaf of T, call it v. Give v two new children. This results in a new tree T' with $n+1$ leaves since v is no longer a leaf but two new leaves have been added. Moreover, since v is now an internal node, T' has $(n - 1) + 1 = n$ internal nodes and the proof is complete. □

What's wrong with this proof? The problem here is that in the inductive step we are obligated to show that **every** regular binary tree with $n+1$ leaves has n internal nodes. What we did was to show that **some specific** tree T' with $n + 1$ leaves has n internal nodes. One might counter, "well any regular binary tree with $n+1$ leaves can be constructed by starting with some regular binary tree with n leaves and do what we did in the proof – namely turn some leaf into an internal node by giving it two leaves as children." This is indeed true, but we would need to prove that it in order for this inductive proof to be complete. This is almost always difficult and cumbersome. Moreover, there are cases where this kind of augmentation simply does not work; It does not allow us to construct every structure of the next larger size.

So what's the fix? The secret to all happiness is to begin the inductive step by taking some **arbitrary** regular binary tree T' with $n + 1$ leaves. We'll then appeal to the inductive hypothesis which applies to *every* regular binary tree with n leaves. Here is how this is typically done:

Proof. The proof is by induction on the number of leaves, n, in the tree.

Basis: When $n = 1$ the tree comprises one leaf and no internal nodes. Thus, the theorem holds.

Inductive Hypothesis: Assume that in every regular binary tree with n leaves the number of internal nodes is $n - 1$.

Inductive Step: We must show that every regular binary tree with $n + 1$ leaves has n internal nodes. Let T' be any arbitrary regular binary tree with $n + 1$ leaves. Let u be an internal node of maximum depth. Then the children of u must both be leaves, since otherwise u would have a child which is an internal node of greater depth than u, contradicting the assumption that u was an internal node of maximum depth. Now, construct a new tree T by removing the children of u from T'. Note that u is now a leaf in the tree T and that T is still a regular binary tree since T' was a regular binary tree and no node other than u has changed. Tree T has $(n + 1) - 2 + 1 = n$ leaves since two leaves were removed from T' but u has become a leaf. Therefore, by the induction hypothesis, T has $n - 1$ internal nodes. Since T' has one more internal node than T, T' has $(n - 1) + 1 = n$ internal nodes and the proof is complete. □

Let's make a few observations about this proof. First, we began with an *arbitrary* regular binary tree with $n + 1$ leaves since we want to show that the claim holds for every such tree. In contrast, in the previous proof, there was no guarantee that we had considered every possible regular binary tree with $n + 1$ leaves. So, the second proof is better already!

In order to appeal to the inductive hypothesis, we then needed to transform our tree T' into one with n leaves. Although it may seem obvious that we can do this by finding two leaves with a common parent and removing those leaves, this actually warrants proof. We proved this here by identifying an internal node u that is as deep in the tree as possible. This kind of argument is called an *extremal* argument because we are choosing an "extreme" element (in this case "extreme" with respect to depth) to make our proof work. Finally, we reduced T' to a regular binary tree T with n leaves. Since the inductive hypothesis applies to *all* regular binary trees with n leaves, it applies to T in particular, and thus we were able to appeal to the inductive hypothesis to complete the proof.