Lighting the Fuse

The final proof technique I want to demonstrate is proof by induction. Induction is useful when you want to show some statement P holds for all natural numbers $n \geq n_0$ for some fixed n_0.

Examples of these statements are the Fundamental Theorem of Arithmetic, $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$, the binomial theorem, $\cos nx$ is an n degree polynomial in $\cos x$, the AM-GM inequality etc.

The idea is to establish the truth for *some* $n \in \mathbb{N}$ and use it to establish truth for other n using recurrences, breaking larger numbers into smaller numbers, etc.

Now, you're familiar with induction. For completeness, I'll cover a basic example, point out some common pitfalls, and then get to some more convoluted examples. Let's start with a straightforward example:

Straightforward / Textbook Examples

You're familiar with the following identity.

Problem 1. Show that

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

Let's prove it by induction. We will show that the statement holds for all $n \geq 1$.

Proof. For $n = 1$, the statement clearly holds. *(This is Step 1. of an Induction Proof – the base case.)*

Now, suppose the statement holds for some natural number $m \geq 1$. *(This is Step 2. of an Induction Proof – the induction hypothesis.) (We assume the statement holds for only one specific m – this is the "weak" induction hypothesis.)*

In Step 2, we assume the statement to be true for *some* subset of natural numbers, and try to prove it true for some other natural number. Here, I have assumed the statement is true for a natural number m, and I want to show it is true for $m + 1$.

Step 3 is just to complete the proof, which I will do now.

Consider the statement for $m+1$. We have

$$
\begin{align}
\sum_{k=1}^{m+1} k &= \sum_{k=1}^{m} k + (m+1) \tag{1}\\
&= \frac{m(m+1)}{2} + (m+1) \tag{2}\\
&= \frac{m(m+1) + 2(m+1)}{2} \tag{3}\\
&= \frac{(m+1)(m+2)}{2} \tag{4}\\
&= \frac{(m+1)((m+1)+1)}{2} \tag{5}
\end{align}
$$

Where in transitioning from (1) to (2), we've used the induction hypothesis, that the statement is true for m. □

This is an example of using a recurrence ($P(m) \implies P(m+1)$) alongside a base case ($P(1)$) to complete the proof.

Now let's have a look at strong induction.

Problem 2. Show that every positive integer $n > 1$ is either prime or a product of primes.

Proof. (Base Case) $n = 2$ is prime.

(Induction Hypothesis) Assume the statement holds for all integers k such that $2 \leq k \leq m$ for some $m \geq 2$. *(We assume the statement holds for all integers up to m – this is the "strong" induction hypothesis.)*

(Complete the Proof) Consider $m+1$. If $m+1$ is prime, we are done. Otherwise, it can be expressed as a product of two integers a and b, where $2 \leq a, b < m+1$. By the induction hypothesis, both a and b can be expressed as products of primes. Therefore, $m+1$ can also be expressed as a product of primes. □

This is an example of the "building block" approach. I did not exploit the truth of the statement at m. I just used some arbitrary numbers smaller than $m+1$ (not even knowing which ones) as "building blocks" for my proof.

Some Exotic Examples

Euclid's Lemma for $n = 2$

Recall we used Euclid's Lemma when proving an integer is even iff its square is even.

Here, we will prove the lemma for $n = 2$, i.e. given an integer m, there is a unique n such that precisely one of $m = 2n$ or $m = 2n + 1$ holds.

We will only concern ourselves with existence. Uniqueness and showing exactly one holds is not relevant to induction.

Problem 3. Show that for every integer m, there exists a unique integer n such that either $m = 2n$ or $m = 2n + 1$.

Proof. We will take three subcases – $m < 0$, $m = 0$, and $m > 0$.

Subcase $m = 0$ is trivial, as $n = 0$ satisfies the condition.

Consider the subcase $m > 0$. (Base Cases) For $m = 1$, we have $1 = 2 \times 0 + 1$ and for $m = 2$, we have $2 = 2 \times 1$. So the base cases hold.

(Induction Hypothesis) Suppose for some positive integer k, Euclid's Lemma holds.

(Completion) Consider $k + 1$. If k has the form $2n$, then $k + 1 = 2n + 1$.

If k has the form $2n + 1$, then $k + 1 = 2(n + 1)$.

Thus Euclid's Lemma holds for $k + 1$. □

Finally consider the subcase $m < 0$.

(Base Cases) For $m = -1$, we have $-1 = 2 \times (-1) + 1$ and for $m = -2$, we have $-2 = 2 \times (-1)$. So the base cases hold.

(Induction Hypothesis) Suppose for some negative integer k, Euclid's Lemma holds.

(Completion) Consider $k - 1$. If k has the form $2n$, then $k - 1 = 2(n - 1) + 1$.

If k has the form $2n + 1$, then $k - 1 = 2n$.

Thus Euclid's Lemma holds for $k - 1$. □

And that does it. It is impossible for us to have left out any integer (think about why this is true). I list this as an exotic example because of the subcase division and the two base cases.

AM-GM Inequality

Let me state the problem first.

Theorem 1. For any non-negative real numbers $x_1, x_2, \ldots, x_n$, we have

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \cdots x_n}$$

with equality if and only if $x_1 = x_2 = \cdots = x_n$.

Again, we will concern ourselves with just the inequality. The condition for equality is not relevant to induction.

We will perform an induction on *the number of variables*, and the way we will ensure every $n > 0$ is hit is unlike anything you will ever encounter.

Proof. For $n = 1$, the inequality is trivially true, as both sides are equal to x_1.

(Induction Hypothesis) Suppose that given *any* k positive reals with $1 \leq k \leq m$, the inequality holds.

(Completion) First we will show the inequality holds for any $2m$ positive reals. Let $x_1, x_2, \ldots, x_{2m}$ be any $2m$ positive reals.

Group them into $x_1, x_2 \ldots x_m$ and $x_{m+1}, x_{m+2} \ldots x_{2m}$.

By Induction Hypothesis, we can apply the inequality to both groups:

$$\frac{x_1 + x_2 + \cdots + x_m}{m} \geq \sqrt[m]{x_1 x_2 \cdots x_m} \tag{6}$$

$$\frac{x_{m+1} + x_{m+2} + \cdots + x_{2m}}{m} \geq \sqrt[m]{x_{m+1} x_{m+2} \cdots x_{2m}} \tag{7}$$

Note the LHS for both inequalities. They're both positive real numbers. First we observe that the AM of the LHSs is $\geq$ the AM of the RHSs.

$$\frac{\frac{x_1 + x_2 + \ldots + x_m}{m} + \frac{x_{m+1} + \ldots + x_{2m}}{m}}{2} \geq \frac{\sqrt[m]{x_1 x_2 \ldots x_m} + \sqrt[m]{x_{m+1} \ldots x_{2m}}}{2} \tag{8}$$

Applying AM-GM on RHS for 2 variables (by IH) we get

$$\frac{\sqrt[m]{x_1 x_2 \ldots x_m} + \sqrt[m]{x_{m+1} \ldots x_{2m}}}{2} \geq \sqrt{\sqrt[m]{x_1 x_2 \ldots x_m} \cdot \sqrt[m]{x_{m+1} \ldots x_{2m}}} \tag{9}$$

$$= \sqrt[2m]{x_1 x_2 \ldots x_{2m}} \tag{10}$$

Combining (8), (9), and (10) we get the AM-GM inequality for $2m$ variables.

$$\frac{x_1 + x_2 + \cdots + x_{2m}}{2m} \geq \sqrt[2m]{x_1 x_2 \cdots x_{2m}}$$

Finally, we'll show the inequality holds for $m - 1$ positive reals with the same IH.

Let $x_1, x_2, \ldots, x_{m-1}$ be any $m - 1$ positive reals. Define x_m as

$$x_m = \frac{x_1 + x_2 + \cdots + x_{m-1}}{m - 1}$$

And apply the inequality to $x_1, x_2, \ldots, x_{m-1}, x_m$.

$$\frac{x_1 + \cdots + x_{m-1} + \frac{x_1 + \cdots + x_{m-1}}{m-1}}{m} \geq \sqrt[m]{x_1 \ldots x_{m-1} \cdot \left(\frac{x_1 + \cdots + x_{m-1}}{m - 1} \right)}$$

The LHS simplifies to

$$\frac{x_1 + \cdots + x_{m-1}}{m - 1}$$

Thus

$$\left(\frac{x_1 + \cdots + x_{m-1}}{m-1}\right)^m \geq x_1 \ldots x_{m-1} \cdot \left(\frac{x_1 + \cdots + x_{m-1}}{m-1}\right)$$

Removing one copy of the fraction from either side,

$$\left(\frac{x_1 + \cdots + x_{m-1}}{m-1}\right)^{m-1} \geq x_1 \ldots x_{m-1}$$

And we're done. $\square$

Or are we? Those of you still awake would have noticed our proof for $k \to 2k$ relies on the inequality being true for 2 variables. But the inequality being true for 2 variables relies on the $k \to 2k$ implication holding in the first place.

This is a common pitfall in induction proofs. We need to find a different proof for 2 variables. That is easily done by considering $(\sqrt{a_1} - \sqrt{a_2})^2 \geq 0$.

What I wanted to emphasise is that you have to have a *very* clear image of how your chain of implications works, and fix it wherever necessary. More on this in the next.