STUDENT'S COMPANIONS IN BASIC MATH: THE EIGHTH

Induction and Recursion

The principle of mathematical induction is this: to establish an infinite sequence of propositions

$$P_1, P_2, P_3, \ \ldots \tag{1}$$

(or, simply put, P_n $(n \geq 1)$), it is enough to verify the following two things

(1) P_1, and

(2) $P_n \Rightarrow P_{n+1}$.

These two things will give a "domino effect" for the validity of all P_n $(n \geq 1)$. The principle is based on the following basic property about the set $\mathbf{N}$ of all natural numbers:

Well ordering Principle for $\mathbf{N}$. *If S is a nonempty subset of $\mathbf{N}$, then there exists an element m in S such that $m \leq n$ for all n in S.*

The number m in the above statement is called **the least element** in S. Now we explain why the well-ordering principle tells us that the method of induction works. Assuming that (1) and (2) above are verified. Let S be the set of all those n for which P_n fails. It is enough to show that S is an empty set. Suppose the contrary that S is nonempty. Then the well-ordering principle tells us that S has a least element, say m. By (1), m cannot be 1. Thus $m > 1$ and as a consequence $k \equiv m - 1$ is also a natural number. Since $k < m$ and m is the minimal element of S, k is not in S. So P_k is valid. By (2), $P_{k+1} \equiv P_m$ is also valid, contradicting the fact that m belongs to S. [If we replace $\mathbf{N}$ by general well-ordered sets called ordinals, we get the so-called transfinite induction and the argument here explains why it works. We do not consider this advanced topic here.]

EXERCISE 1. Use induction to show $1 + 2 + 3 + \cdots + n = \dfrac{n(n+1)}{2}$.

EXERCISE 2. Use induction to show $1 + 3 + 5 + \cdots + (2n - 1) = n^2$. (Remember: in mathematics, "show" means "prove".)

EXERCISE 3. Prove the Cauchy-Schwarz inequality

$$(a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2 \leq (a_1^2 + a_2^2 \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2)$$

for $n = 1$ and $n = 2$. Then use induction to prove this inequality for general n.

QUESTION 4. What is wrong with the following "proof" of "$n = n + 1$"?

> "Assume that $n = n + 1$ holds for $n = k$, that is, $k = k + 1$. Add both sides by 1, we get $k + 1 = k + 1 + 1$, which shows that the identity $n = n + 1$ also holds for $n = k + 1$. By the principle of induction, $n = n + 1$ is valid for all n."

QUESTION 5. What is wrong with the following "proof" of the statement "every man is bald" by induction?

> "All we need to do is to establish the statement 'a man with n hairs or less is bald' for all n. When $n = 0$, 1 or 2, the statement is clearly valid. Now assume that the statement is true for $n = k$, that is, assume that "a man with k hairs or less is bald" is valid. Then the statement $n = k + 1$ is also valid because an additional hair cannot change the fact of being bald. So by the principle of induction, the statement 'a man with n hairs is bald' is valid for all n. This proves all men are bald."

QUESTION 6. What is wrong with the following proof of the statement "all men in a group are pigs"?

> "Let us agree that some men are pigs. We may and we do assume that the group has at least one pig because we can ask a pig to join them. We prove by induction on the number n of persons in the group. The statement is clear for $n = 1$, because this group has at least one pig. Assume the validity of the statement for $n = k$. Now take a group of $k + 1$ persons. Then all men in any subgroup of size k which has at least one pig are pigs, by our induction hypothesis. From this we see that all $k + 1$ men in the group are pigs; (take a special case $k = 99$ to see how this argument works)."

In some rare occasions we use a technique called **reverse induction**: to establish the validity of a sequence of propositions P_n $(n \geq 1)$, it is enough to establish the following

(a) P_n is valid for infinitely many n.

(b) If P_{n+1} is valid, then so is P_n.

To see that P_n is valid for all n after (a) and (b) is checked, we assume the contrary that P_k is not valid for some k. By (a), we know that there is some N such that P_N is valid and $k < N$. Let S be the set of all those m with $m \le N$ and P_m fails. Then S is a finite set of numbers and S is nonempty because k is in S. Hence there is a number m_0 in S which is maximal among S. m_0 cannot be N since P_N is valid. So $m_0 < N$ and consequently $m_0 + 1 \le N$. By the maximality of m_0 we know that $m_0 + 1$ is not in S; in other words, P_{m_0+1} is valid. By (b), we know P_{m_0} is valid. This contradicts the fact that m_0 is in S.

EXAMPLE 7. We use this principle of reverse induction to prove the famous arithmetic-geometric means inequality:

$$\sqrt[n]{a_1 a_2 \cdots a_n} \le \frac{a_1 + a_2 + \cdots + a_n}{n} \tag{2}$$

where $a_1, a_2, \ldots, a_n$ are positive numbers. The left hand side is called the geometric mean of $a_1, a_2, \ldots, a_n$ and the right hand side is called the arithmetic mean. So the inequality says, the geometric mean of a finite set of positive numbers is less than or equal to their arithmetic mean:

$$G = G(a_1, \ldots, a_n) \ \le\ A = A(a_1, \ldots, a_n).$$

where

$$G = \sqrt[n]{a_1 a_2 \cdots a_n} \quad \text{and} \quad A = \frac{a_1 + a_2 + \cdots + a_n}{n}.$$

The following three steps will accomplish the proof of (2). Step one: Check that (2) holds for $n = 2$, or , using different notation,

$$\sqrt{AB} \le \frac{A+B}{2}, \quad \text{for } A,\, B > 0. \tag{3}$$

Step 2: Check that if (2) holds for $n = m$, then it holds for $n = 2m$. Step 3: Check that if (2) holds for $n = m + 1$, then it holds for $n = m$. Steps 1 and 2 tell us that (2) holds for $n = 2^k$ $(k = 1, 2, \ldots)$ and hence it holds for infinitely many n. Step 3 completes the reverse induction. To facilitate step 1, we use the symbol $\Leftrightarrow$ for "if and only if":

$$\begin{aligned} \sqrt{AB} \le \frac{A+B}{2} \quad &\Leftrightarrow \quad AB \le \frac{(A+B)^2}{4} \\ &\Leftrightarrow \quad (A+B)^2 \ge 4AB \\ &\Leftrightarrow \quad A^2 + 2AB + B^2 \ge 4AB \\ &\Leftrightarrow \quad A^2 - 2AB + B^2 \ge 0 \\ &\Leftrightarrow \quad (A-B)^2 \ge 0 \end{aligned}$$

The last inequality is obvious. So (3) is valid. Next: step 2. Assume that (2) holds for $n = m$. Consider $2m$ positive numbers: $a_1, \ldots, a_m, a_{m+1}, \ldots, a_{2m}$. Naturally we split them into two groups, each with m numbers. Apply the induction hypothesis to each group: $G(a_1, \ldots, a_m) \le A(a_1, \ldots, a_m)$ and $G(a_{m+1}, \ldots, a_{2m}) \le A(a_{m+1}, \ldots, a_{2m})$. Let $A = G(a_1, \ldots, a_m)$ and $B = G(a_{m+1}, \ldots, a_{2m})$ in (3) to obtain

$$\sqrt{G(a_1, \ldots, a_m)G(a_{m+1}, \ldots, a_{2m})} \le \frac{G(a_1, \ldots, a_m) + G(a_{m+1}, \ldots, a_{2m})}{2}$$
$$\le \frac{A(a_1, \ldots, a_m) + A(a_{m+1}, \ldots, a_{2m})}{2}.$$

A correct answer to the following question will complete step two:

QUESTION 8. Why is the last inequality gives (2) for $n = 2m$?

Final step: assume that (2) is valid for $n = m + 1$ and $a_1, \ldots, a_m$ are given. Write down (2) for $n = m + 1$:

$$G(a_1, \ldots, a_m, a_{m+1}) \le A(a_1, \ldots, a_m, a_{m+1}), \tag{4}$$

which is assumed to be valid. Here, $a_1, \ldots, a_m$ are given. We have freedom to choose a_{m+1}. A natural choice is

$$a_{m+1} = A \equiv \frac{a_1 + a_2 + \cdots + a_m}{m}. \tag{5}$$

This will work. But I leave the rest of the final step to you as an exercise, because I don't want to take the fun (of doing it) away from you.

EXERCISE 9. Check that, with the choice of a_{m+1} given by (5), inequality (4) is reduced to (2) for $n = m$.

It is possible to give a more streamlined proof of (2) by means of the convexity of the exponential function. But this needs some machinery in convexity theory. We will give an outline of this proof after describing some basic theory of convexity.

QUESTION 10. (1) What is wrong with the following proof "every man is obese" by using the reverse induction?

> "It is enough to check the validity of the statement 'a man weighted n lbs or more is obese'. The statement is obviously true for $n \ge 400$. So it is true for infinitely many

n. Now suppose that it is true for $n = m+1$. Then it is also true for $n = m$ because losing one pound cannot change the fact of being obese."

(2) What is wrong with the following argument of reverse induction to prove that "every student is doing well in the exam"?

"It is enough to prove that 'a student with a final mark of $n\%$ or more is doing well in the exam'. This statement is obvious when $n = 100$, in that case the student receives a perfect mark of 100%. Now assume that the statement is true for $n = m+1$. Then it is also true for $n = m$ because losing only one mark is not considered as a serious defect in writing the exam."

An important role played by induction is to give proper definitions for many mathematical expressions involving natural numbers. The usual pattern (or its variant) of this **definition by induction** or **by recursion** is such: to define a sequence $\{f(n)\}_{n\geq 1}$, we take the following two steps:

(I) Specify what $f(1)$ is,

(II) Specify how $f(n+1)$ can be decided from a subset of $f(1), f(2), \ldots, f(n)$.

EXAMPLE 11. You know the meanings of the expressions a^n, $n!$ and $\sum_{k=1}^{n} a_k$. Strictly speaking, they should be defined by induction. We write down the two steps (I), (II) described above for defining each of them:

$a^1 = a$; $a^{n+1} = a^n a$. (With $f(n) = a^n$, $f(1) = a$ and $f(n+1) = f(n)f(1)$.)

$1! = 1$; $(n+1)! = n!\,(n+1)$. (With $f(n) = n!$, $f(1) = 1$ and $f(n+1) = f(n)(n+1)$.)

$\sum_{k=1}^{1} a_k = a_1$; $\sum_{k=1}^{n+1} a_k = (\sum_{k=1}^{n} a_k) + a_{n+1}$.

(With $f(n) = \sum_{k=1}^{n} a_k$, $f(1) = a_1$ and $f(n+1) = f(n) + a_{n+1}$.)

Certainly, na is also defined by induction: $1a = a$ and $(n+1)a = na + a$.

EXAMPLE 12. The famous **Fibonacci numbers** F_n $(n = 0, 1, 2, 3, \ldots)$ are defined by induction (or recursion):

$$F_n = F_{n-1} + F_{n-2}, \qquad n \geq 2, \tag{6}$$

which says, starting from the third term, every term in the sequence of Fibonacci numbers is the sum of the previous two terms. In order to determine this sequence, we have to specify the first two terms. Here they are: $F_0 = 1$ and $F_1 = 1$. It is easy to produce a few terms at the beginning of the sequence

$$1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ \ldots$$

(This sequence was originally described by the inventor as how rabbits multiply: F_n is the number of pairs of rabbits on the nth day. There is an interesting story of this but it does not concern us here.) The mathematical question here is, how can we find a closed expression for F_n for general n? There are many ways to answer this question, such as: treat it as a difference equation, or rewrite (5) as a matrix identity

$$\begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{n-1} \\ F_{n-2} \end{bmatrix}$$

and use the method of diagonalization in linear algebra, or use generating functions. We briefly describe the last method here. Form the power series $f(x) = \sum_{n=0}^{\infty} F_n x^n$. It is called the **generating function** for the Fibonacci sequence. It converges when $|x|$ is small enough. Now

$$\begin{aligned} f(x) &= F_0 + F_1 x + \sum\nolimits_{n\geq 2} F_n x^n = F_0 + F_1 x + \sum\nolimits_{n\geq 2} (F_{n-1} + F_{n-2})\, x^n \\ &= 1 + x + \sum\nolimits_{n\geq 2} F_{n-1}\, x^n + \sum\nolimits_{n\geq 2} F_{n-2}\, x^n \\ &= 1 + x + \sum\nolimits_{k\geq 1} F_k\, x^{k+1} + \sum\nolimits_{\ell\geq 0} F_\ell\, x^{\ell+2} \qquad (k = n-1,\ \ell = n-2) \\ &= 1 + x + x\left(\sum\nolimits_{k\geq 1} F_k\, x^k\right) + x^2\left(\sum\nolimits_{\ell\geq 0} F_\ell\, x^\ell\right) \\ &= 1 + x + x(f(x) - 1) + x^2 f(x) = 1 + x f(x) + x^2 f(x). \end{aligned}$$

So we have $(1 - x - x^2) f(x) = 1$. We can factorize the polynomial $1 - x - x^2$ as $1 - x - x^2 = (1 - r_+ x)(1 - r_- x)$, where $r_\pm = (1 \pm \sqrt{5})/2$ (with $r_+ + r_- = 1$ and $r_+ r_- = -1$).

QUESTION 13. Why can we do so?

Now

$$\begin{aligned} f(x) &= \frac{1}{1 - x - x^2} = \frac{1}{(1 - r_+ x)(1 - r_- x)} = \left(\frac{r_+}{1 - r_+ x} - \frac{r_-}{1 - r_- x}\right)\frac{1}{r_+ - r_-} \\ &= \left(r_+ \sum\nolimits_{n\geq 0} r_+^n\, x^n - r_- \sum\nolimits_{n\geq 0} r_-^n\, x^n\right)\frac{1}{\sqrt{5}} = \frac{1}{\sqrt{5}} \sum\nolimits_{n\geq 0} (r_+^{n+1} - r_-^{n+1})\, x^n. \end{aligned}$$

Comparing this with $f(x) = \sum_{n\geq 0} F_n\, x^n$, we obtain

$$F_n = \frac{1}{\sqrt{5}}(r_+^{n+1} - r_-^{n+1}) \equiv \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}.$$

The amazing thing here is the appearance of the irrational number $\sqrt{5}$ in this answer for F_n, which is always an integer.

EXERCISE 14. Use the recursion relation (6) to prove the following identity

$$F_0 + F_1 + F_2 + \cdots + F_{n-1} = F_{n+1} - 1, \quad (n \geq 1).$$

Also prove this by induction.

EXERCISE 15. Prove the amazing identity $F_{m+n} = F_m F_n + F_{m-1}F_{n-1}$ by induction.

Recursion is an important notion for constructing so called **recursive functions**, which is a key notion in logic and theoretical computer science. [A good and inexpensive book in this area is Nigel Cutland's "Computability" published by Cambridge University Press. But it may be too demanding to read at this level.]

In the rest we study the notion of convexity, which is important in many areas like inequalities and optimization.

Take two points $\mathbf{a}$ and $\mathbf{b}$ in the d-dimensional Euclidean space $\mathbf{R}^d$. Denote by $[\mathbf{a}, \mathbf{b}]$ the line segment obtained by joining these two points, consisting of points of the form $(1-\lambda)\mathbf{a} + \lambda\mathbf{b}$ with $0 \leq \lambda \leq 1$:

$$[\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbf{R}^n : \ \mathbf{x} = (1-\lambda)\mathbf{a} + \lambda\mathbf{b} \text{ for some } \lambda, \text{ with } 0 \leq \lambda \leq 1\}.$$

A set C in $\mathbf{R}^d$ is convex if, for all points $\mathbf{a}, \mathbf{b}$ in $\mathbf{R}^d$, the line segment $[\mathbf{a}, \mathbf{b}]$ is contained in C:

$$\forall\ \mathbf{a},\ \mathbf{b} \in C, \quad 0 \leq \lambda \leq 1 \quad \Rightarrow \quad (1-\lambda)\mathbf{a} + \lambda\mathbf{b} \in C. \tag{7}$$

[You should make some free hand sketches of sets in the plane to understand this definition.] Here is an important description of convex sets:

Theorem. *A set* C *in* $\mathbf{R}^d$ *is convex if and only if it satisfies the following condition*: (C) *For each finite set of points* $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ *in* C, *a point of the form*

$$\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \cdots + \lambda_n\mathbf{a}_n \qquad (\lambda_1, \lambda_2, \ldots, \lambda_n \geq 0 \text{ with } \lambda_1 + \lambda_2 + \cdots + \lambda_n = 1) \tag{8}$$

is in C.

A points of the form (8) is called a **convex combination** of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. Physically, the point described by (8) is the barycentre of a set of mass points, where the kth point $\mathbf{a}_k$ has a fraction λ_k of mass for each k.

EXAMPLE 16. We are asked to prove this theorem. The "if" part is clear, if you observe that condition (7) for convexity is just condition (C) for $n = 2$ with $\mathbf{a} = \mathbf{a}_1$, $\mathbf{b} = \mathbf{a}_2$, and $\lambda = \lambda_2$. Now we assume (7) and prove (C) by induction on n. When $n = 1$, (C) is clear. Assume that (C) holds for $n = k$. Given $k+1$ points $\mathbf{a}_1, \ldots, \mathbf{a}_k, \mathbf{a}_{k+1}$ in C and $k+1$ nonnegative numbers $\lambda_1, \lambda_2, \ldots, \lambda_{k+1}$, with $\lambda_1 + \cdots + \lambda_{k+1} = 1$, we have to show that $\mathbf{b} \equiv \lambda_1\mathbf{a}_1 + \cdots + \lambda_k\mathbf{a}_k + \lambda_{k+1}\mathbf{a}_{k+1}$ is in C. Here we assume that $\lambda_{k+1} \neq 0$, otherwise this case is reduced to the case of $n = k$. Let $\mu_j = \lambda_j/(1 - \lambda_{k+1})$ for $j = 1, \ldots, k$. Then $\mu_j \geq 0$ and $\mu_1 + \cdots + \mu_k = 1$. So, by induction hypothesis, the point $\mathbf{a} = \mu_1\mathbf{a}_1 + \cdots + \mu_n\mathbf{a}_n$ is in C. Now $\mathbf{b} = (1 - \lambda_{k+1})\mathbf{a} + \lambda_{k+1}\mathbf{a}_{k+1}$. By condition (7), we know that $\mathbf{b}$ is in C.

EXERCISE 17. Prove that the set $D = \{(x, y) \in \mathbf{R}^2 |\ x^2 + y^2 \leq 1\}$, called the **unit disk**, is convex.

Let f be a function defined on an interval I (I can be open, or closed; finite, or infinite). We say that f is a **convex function** if, for all u, v in I and for all real number λ with $0 \leq \lambda \leq 1$, we have

$$f((1 - \lambda)u + \lambda v) \leq (1 - \lambda)f(u) + \lambda f(v).$$

(You should sketch a figure to understand this condition.)

EXERCISE 18. Use the above definition to prove that $f(x) = x^2$ is a convex function.

EXERCISE 19. Prove that if f is a convex function defined on an interval I, then, for all points $x_1, \ldots, x_n$ in I and real numbers $\lambda_1, \ldots, \lambda_n \geq 0$ with $\lambda_1 + \cdots + \lambda_n = 1$, we have

$$f(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n). \tag{9}$$

EXERCISE 20. Prove that if both f_1 and f_2 are convex functions defined on I, then so is the function $f = \max\{f_1, f_2\}$ given by $f(x) = \max\{f_1(x), f_2(x)\}$. (For example, letting $f_1(x) = x$ and $f_2(x) = -x$, we have $f(x) = |x|$. So the function $|x|$ is convex.)

QUESTION 21. Given two convex functions f_1 and f_2, is $g = \min\{f_1, f_2\}$ necessarily a convex function?

If a function f is smooth, we can use calculus to test its convexity: *A smooth function defined on an interval I is convex if and only if $f''(x) \geq 0$ for all x in I.* The proof of this assertion, which requires the mean value theorem, is omitted. For example, $f(x) = x^2$ is a convex function because $f''(x) = 2 > 0$. Also, e^x is a convex function because its second derivative is still e^x which is always positive. Applying (8) to e^x, we have

$$e^{\lambda_1 x_1 + \cdots + \lambda_n x_n} \leq \lambda_1 e^{x_1} + \cdots + \lambda_n e^{x_n}. \tag{10}$$

where $\lambda_1, \ldots, \lambda_n \geq 0$ and $\lambda_1 + \cdots + \lambda_n = 1$. Given positive numbers $a_1, \ldots, a_n$, we let $x_j = \ln a_j$ $(1 \leq j \leq n)$ so that $a_j = e^{x_j}$. Then (10) becomes

$$a_1^{\lambda_1} a_2^{\lambda_2} \cdots a_n^{\lambda_n} \leq \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_n a_n, \tag{11}$$

where $\lambda_1, \ldots, \lambda_n \geq 0$ and $\lambda_1 + \cdots + \lambda_n = 1$. When we choose $\lambda_1 = \cdots = \lambda_n = 1/n$, (11) becomes

$$(a_1 a_2 \cdots a_n)^{1/n} \leq \frac{a_1 + a_2 + \cdots + a_n}{n}.$$

which is the geometric-arithmetic means inequality described in *EXAMPLE 7* above.

Finally, we mention that the Taylor formula for a polynomial of degree n

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (x - a)^k$$

(see identity (11) in the *SEVENTH COMPANION*) can be proved by induction. I leave this proof to you as *PROBLEM 22.*