1. Sums and Products

In math, very often we have some interesting numbers which we would like to find their sum or product. Below we will look at a few methods for doing these operations. (Here we will also consider integrals which we can view as summing uncountably many numbers.)

Pairing Method. Recall to find $S = 1 + 2 + \cdots + 100$ we can say $S = 100 + 99 + \cdots + 1$ also and hence $2S = (1 + 100) + (2 + 99) + \cdots + (100 + 1) = 100 \times 101 = 10100$ yielding $S = 5050$. *This suggest that in handling numbers we can try to pair them first and hope this can simplify the problem.*

Examples. (1) (2000 APMO) Find $S = \displaystyle\sum_{i=0}^{101} \frac{x_i^3}{1 - 3x_i + 3x_i^2}$ for $x_i = \frac{i}{101}$.

Solution. Note $1 - 3x + 3x^2 = (1-x)^3 + x^3$. Let $f(x) = \dfrac{x^3}{(1-x)^3 + x^3}$. Then $f(x) + f(1-x) = 1$. Since $1 - x_i = x_{101-i}$, we have

$$2S = \sum_{i=0}^{101} \big(f(x_i) + f(x_{101-i}) \big) = 102 \qquad \Longrightarrow \quad S = 51.$$

---

(2) (2000 HKMO Math Camp) Express $\dfrac{\cos 1^\circ + \cos 2^\circ + \cdots + \cos 44^\circ}{\sin 1^\circ + \sin 2^\circ + \cdots + \sin 44^\circ}$ in the form $a + b\sqrt{c}$, where a, b and c are integers.

Solution 1. Recall $\cos a + \cos b = 2\cos\dfrac{a+b}{2}\cos\dfrac{a-b}{2}$ and $\sin a + \sin b = 2\sin\dfrac{a+b}{2}\cos\dfrac{a-b}{2}$. Taking $a = (45-n)^\circ$ and $b = n^\circ$ for $n = 1, 2, \ldots, 22$, we have

$$\frac{\cos 1^\circ + \cos 2^\circ + \cdots + \cos 44^\circ}{\sin 1^\circ + \sin 2^\circ + \cdots + \sin 44^\circ} = \frac{2\cos\frac{45^\circ}{2}\big(\cos\frac{43^\circ}{2} + \cos\frac{41^\circ}{2} + \cdots + \cos\frac{1^\circ}{2}\big)}{2\sin\frac{45^\circ}{2}\big(\cos\frac{43^\circ}{2} + \cos\frac{41^\circ}{2} + \cdots + \cos\frac{1^\circ}{2}\big)}$$
$$= \frac{\cos\frac{45^\circ}{2}}{\sin\frac{45^\circ}{2}} = \sqrt{\frac{\sqrt{2}+1}{\sqrt{2}-1}} = 1 + \sqrt{2}.$$

Solution 2. Since

$$\cos n^\circ + \sin n^\circ = \sqrt{2}(\cos 45^\circ \cos n^\circ + \sin 45^\circ \sin n^\circ) = \sqrt{2}\cos(45-n)^\circ,$$

$$\begin{aligned}&(\cos 1^\circ + \cos 2^\circ + \cdots + \cos 44^\circ) + (\sin 1^\circ + \sin 2^\circ + \cdots + \sin 44^\circ)\\ =&\sqrt{2}(\cos 44^\circ + \cos 43^\circ + \cdots + \cos 1^\circ).\end{aligned}$$

Therefore, $\dfrac{\cos 1^\circ + \cos 2^\circ + \cdots + \cos 44^\circ}{\sin 1^\circ + \sin 2^\circ + \cdots + \sin 44^\circ} = \dfrac{1}{\sqrt{2}-1} = 1 + \sqrt{2}$.

---

(3) (1980 Putnam Exam) Evaluate $\displaystyle\int_0^{\pi/2} \frac{dx}{1 + (\tan x)^{\sqrt{2}}}$.

Solution. Let $r = \sqrt{2}$ and $I = \displaystyle\int_0^{\pi/2} \frac{dx}{1 + \tan^r x} = \int_0^{\pi/2} \frac{\cos^r x\, dx}{\cos^r x + \sin^r x}$. Since $\cos(\frac{\pi}{2} - t) = \sin t$ and $\sin(\frac{\pi}{2} - t) = \cos t$, the change of variable $x = \dfrac{\pi}{2} - t$ will yield $I = \displaystyle\int_0^{\pi/2} \frac{\sin^r t\, dt}{\sin^r t + \cos^r t}$. So

$$2I = \int_0^{\pi/2} \frac{\cos^r x + \sin^r x}{\cos^r x + \sin^r x}\, dx = \int_0^{\pi/2} dx = \frac{\pi}{2} \qquad \Longrightarrow \quad I = \frac{\pi}{4}.$$

Telescoping Method. A particularly simple type of sum everybody can do is of the form

$$(a_1 - a_2) + (a_2 - a_3) + \cdots + (a_{n-1} - a_n) = a_1 - a_n.$$

This type of sum is called a *telescoping sum.* Similarly, there are *telescoping products*, where the factors are of the form a_i/a_{i+1} and their product is a_1/a_n. Some summation or product problems are of these forms. So *in summing, we should try to see if the terms can be put in the form* $a_i - a_{i+1}$. Here are some examples.

Examples. (4) Simplify $\sin 1 + \sin 2 + \cdots + \sin n$.

Solution 1. Recall $\sin a \sin b = \dfrac{\cos(a-b)-\cos(a+b)}{2}$. Setting $b = \frac{1}{2}$, we get $\sin a = \dfrac{\cos(a-\frac{1}{2})-\cos(a+\frac{1}{2})}{2\sin\frac{1}{2}}$. So

$$\begin{aligned}&\sin 1+\sin 2+\cdots+\sin n\\ =&\frac{(\cos\frac{1}{2}-\cos\frac{3}{2})+(\cos\frac{3}{2}-\cos\frac{5}{2})+\cdots+(\cos(n-\frac{1}{2})-\cos(n+\frac{1}{2}))}{2\sin\frac{1}{2}}\\ =&\frac{\cos\frac{1}{2}-\cos(n+\frac{1}{2})}{2\sin\frac{1}{2}}.\end{aligned}$$

Solution 2. (Use complex arithmetic) Let $z = \cos 1 + i\sin 1$. By de Moivre's formula, $z^k = (\cos 1 + i\sin 1)^k = \cos k + i\sin k$. Note $\sin 1+\sin 2+\cdots+\sin n = \text{Im}\,(z+z^2+\cdots+z^n)$. Now

$$\begin{aligned}z+z^2+\cdots+z^n &= \frac{z(z^n-1)}{z-1} = \frac{z^{1/2}(z^n-1)}{z^{1/2}-z^{-1/2}} = \frac{z^{n+1/2}-z^{1/2}}{2i\sin(1/2)}\\ &= \frac{\big(\cos(n+\frac{1}{2})-\cos\frac{1}{2}\big)+i\big(\sin(n+\frac{1}{2})-\sin\frac{1}{2}\big)}{2i\sin\frac{1}{2}}\\ &= \frac{\big(\sin(n+\frac{1}{2})-\sin\frac{1}{2}\big)+i\big(\cos\frac{1}{2}-\cos(n+\frac{1}{2})\big)}{2\sin\frac{1}{2}}.\end{aligned}$$

So $\displaystyle\sum_{k=1}^{n}\sin k = \frac{\cos\frac{1}{2}-\cos(n+\frac{1}{2})}{2\sin\frac{1}{2}}$. Also, $\displaystyle\sum_{k=1}^{n}\cos k = \frac{\sin(n+\frac{1}{2})-\sin\frac{1}{2}}{2\sin\frac{1}{2}}$.

---

(5) (1977 Putnam Exam) Evaluate $\displaystyle\prod_{n=2}^{\infty}\frac{n^3-1}{n^3+1}$. (Here $\displaystyle\prod_{n=2}^{\infty}a_n = \lim_{k\to\infty}a_2a_3\cdots a_k$.)

Solution. Note that

$$\frac{n^3-1}{n^3+1} = \frac{(n-1)(n^2+n+1)}{(n+1)(n^2-n+1)} = \frac{(n-1)\big((n+1)^2-(n+1)+1\big)}{(n+1)(n^2-n+1)}.$$

So for large k,

$$\prod_{n=2}^{k}\frac{n^3-1}{n^3+1} = \Big(\frac{1\cdot 7}{3\cdot 3}\Big)\Big(\frac{2\cdot 13}{4\cdot 7}\Big)\Big(\frac{3\cdot 21}{5\cdot 13}\Big)\cdots\Big(\frac{(k-1)(k^2+k+1)}{(k+1)(k^2-k+1)}\Big) = \frac{2(k^2+k+1)}{3k(k+1)}.$$

Taking limit as $k\to\infty$, we get the answer is $\dfrac{2}{3}$.

---

(6) Show that $2\sqrt{101}-2 < \displaystyle\sum_{n=1}^{100}\frac{1}{\sqrt{n}} < 20$.

Solution. (Note it may be difficult to find the exact sum. We have to bound the terms from above and below.) To get telescoping effects, we use

$$\sqrt{n+1}-\sqrt{n} = \frac{1}{\sqrt{n+1}+\sqrt{n}} < \frac{1}{2\sqrt{n}} < \frac{1}{\sqrt{n}+\sqrt{n-1}} = \sqrt{n}-\sqrt{n-1}.$$

Summing from $n=1$ to 100, then multiplying by 2, we get the inequalities.

Binomial Sums. For sums involving binomial coefficients, we will rely on the binomial theorem and sometimes a bit of calculus to find the answers. Trivially, from $(1+x)^n = \displaystyle\sum_{k=0}^{n}\binom{n}{k}x^k$, we get $2^n = \displaystyle\sum_{k=0}^{n}\binom{n}{k}$ by setting $x=1$ and $0 = \displaystyle\sum_{n=0}^{k}(-1)^k\binom{n}{k}$ by setting $x=-1$. These equations explain why the sum of the n-th row of the Pascal triangle is 2^n and the alternate sum is 0. Below we will look at more examples.

Examples. (7) Simplify $\displaystyle\sum_{k=0}^{n}\binom{n}{k}^2$.

Solution. Since $\binom{n}{k} = \binom{n}{n-k}$, the sum is the same as $\sum_{k=0}^{n} \binom{n}{k}\binom{n}{n-k}$. Note this sum is the coefficient of x^n in

$$\left(1+\binom{n}{1}x+\binom{n}{2}x^2+\cdots+x^n\right)\left(1+\binom{n}{1}x+\binom{n}{2}x^2+\cdots+x^n\right)$$
$$=(1+x)^n(1+x)^n=(1+x)^{2n}=1+\cdots+\binom{2n}{n}x^n+\cdots+x^{2n}.$$

Therefore, $\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}$. (*Remark*: By looking at the coefficients of x^j on both sides of the identity $(1+x)^m(1+x)^n = (1+x)^{m+n}$, we will get the more general identity

$$\sum_{k=0}^{j}\binom{m}{k}\binom{n}{j-k} = \binom{m+n}{j}.)$$

---

(8) (1962 Putnam Exam) Evaluate in closed form $\sum_{k=0}^{n} k^2\binom{n}{k}$.

Solution 1. For $n \geq k \geq 2$, $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n}{k}\binom{n-1}{k-1} = \frac{n(n-1)}{k(k-1)}\binom{n-2}{k-2}$. So

$$\begin{aligned}\sum_{k=0}^{n} k^2\binom{n}{k} &= n+\sum_{k=2}^{n}(k(k-1)+k)\binom{n}{k}\\ &= n+\sum_{k=2}^{n}\left(n(n-1)\binom{n-2}{k-2}+n\binom{n-1}{k-1}\right)\\ &= n+n(n-1)2^{n-2}+n(2^{n-1}-1) = n(n+1)2^{n-2}.\end{aligned}$$

The cases $n = 0$ and 1 are easily checked to be the same.

Solution 2. (Use calculus) Differentiating both sides of $(1+x)^n = \sum_{k=0}^{n}\binom{n}{k}x^k$, we get $n(1+x)^{n-1} = \sum_{k=0}^{n} k\binom{n}{k}x^{k-1}$. Multiplying both sides by x, then differentiating

both sides again, we get $n(1+x)^{n-1}+n(n-1)x(1+x)^{n-2} = \sum_{k=0}^{n} k^2\binom{n}{k}x^{k-1}$.

Setting $x = 1$, we get $\sum_{k=0}^{n} k^2\binom{n}{k} = n2^{n-1}+n(n-1)2^{n-2} = n(n+1)2^{n-2}$.

Fubini's Principle. When we have m rows of n numbers, to find the sum of these mn numbers, we can sum each row first then add up the row sums. This will be the same as summing each column first then add up the column sums. This simple fact is known as *Fubini's Principle.* There is a similar statement for the product of the mn numbers. In short, we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij} \quad \text{and} \quad \prod_{i=1}^{m}\prod_{j=1}^{n} a_{ij} = \prod_{j=1}^{n}\prod_{i=1}^{m} a_{ij}.$$

Historically the original Fubini's principle was about interchanging the order of double integrals, namely if $|f|$ is integrable on the domain, then

$$\int_a^b\int_c^d f(x,y)\,dxdy = \int_c^d\int_a^b f(x,y)\,dydx.$$

Examples. (9) For a $n \times n$ chessboard with n odd, each square is written a $+1$ or a -1. Let p_i be the product of the numbers in the i-th row and q_j be the product of the numbers in the j-th column. Show that $p_1+p_2+\cdots+p_n+q_1+q_2+\cdots+q_n \neq 0$.

Solution. By Fubini's principle, $p_1p_2\cdots p_n = q_1q_2\cdots q_n$. Note each p_i or q_j is ± 1. Suppose there are s (-1)'s among $p_1, p_2, \cdots, p_n$ and t (-1)'s among $q_1, q_2, \cdots, q_n$. Then either s and t are both odd or both even. Now

$$\begin{aligned}p_1+p_2+\cdots+p_n+q_1+q_2+\cdots+q_n &= -s+(n-s)-t+(n-t)\\ &= 2(n-s-t) \neq 0\end{aligned}$$

because $n-s-t$ is odd.

Note: If two integer variables are either both odd or both even, then we say they are of the *same parity*.

(10) (1982 Putnam Exam) Evaluate $\displaystyle\int_0^\infty \frac{\operatorname{Arctan}(\pi x) - \operatorname{Arctan} x}{x} dx$.

Solution.

$$\begin{aligned}\int_0^\infty \frac{\operatorname{Arctan}(\pi x) - \operatorname{Arctan} x}{x} dx &= \int_0^\infty \frac{1}{x} \operatorname{Arctan}(ux)\Big|_{u=1}^{u=\pi} dx \\ &= \int_0^\infty \int_1^\pi \frac{1}{1+(xu)^2} du dx \\ &= \int_1^\pi \int_0^\infty \frac{1}{1+(xu)^2} dx du \\ &= \int_1^\pi \frac{\pi}{2u} du = \frac{\pi}{2} \ln \pi.\end{aligned}$$

The interchange is valid since the integrand of the double integral is nonnegative and continuous on the domain and the integral is finite.

---

(11) Let n be a positive integer and p be a prime. Find the highest power of p dividing $n!$.

Solution. Write $1 \times 2 \times \cdots \times n$. For $k = 1, 2, \ldots, n$, if the highest power of p dividing k is j, then write j 1's in a column below the factor k in $1 \times 2 \times \cdots \times n$. (If p does not divide k, then $j = 0$, so do not write any 1.) The total number of 1's below $1 \times 2 \times \cdots \times n$ is the highest power of p dividing $n!$. Summing the column sums is difficult, but summing the row sums is easy. In the first row, there is one 1 in every p consecutive integers, so the first row sum is $[n/p]$. In the second row, there is one 1 in every p^2 consecutive integers, so the second row sum is $[n/p^2]$. Keep going. The i-th row sum is $[n/p^i]$. So the total number of 1's is

$$[n/p] + [n/p^2] + [n/p^3] + \cdots.$$

This is the highest power of p dividing $n!$.

---

(12) (1987 IMO) Let $P_n(k)$ be the number of permutations of $1, 2, 3, \ldots, n$ which have exactly k fixed points. Prove that $\displaystyle\sum_{k=0}^{n} k P_n(k) = n!$. (A *fixed point* is a number that is not moved by the permutation.)

Solution. There are $n!$ permutations of $1, 2, 3, \ldots, n$. Call them $f_1, f_2, \ldots, f_{n!}$. Write each in a separate row. For each permutation, replace each fixed point of f by 1 and replace all other numbers in the permutation by 0. Then the row sum gives the number of fixed points of f. Now $\displaystyle\sum_{k=0}^{n} k P_n(k)$ is the sum of the row sums, grouped according to the $P_n(k)$ rows that have the same row sum k. By Fubini's principle, this is also the sum of the column sums. For the j-th column, the number of 1's is the number of times j is a fixed point among the $n!$ permutations. If j is fixed, then the number of ways of permuting the other $n-1$ numbers is $P_{n-1}^{n-1} = (n-1)!$. So the column sum is $(n-1)!$ for each of the n columns. Therefore, the sum of column sums is $n(n-1)! = n!$. This is $\displaystyle\sum_{k=0}^{n} k P_n(k)$.

Exercises

1. Find $\dfrac{1}{1+\cot 1^\circ} + \dfrac{1}{1+\cot 5^\circ} + \dfrac{1}{1+\cot 9^\circ} + \cdots + \dfrac{1}{1+\cot 85^\circ} + \dfrac{1}{1+\cot 89^\circ}$.

2. (1988 Singapore MO) Compute

$$\frac{1}{2\sqrt{1}+1\sqrt{2}} + \frac{1}{3\sqrt{2}+2\sqrt{3}} + \cdots + \frac{1}{100\sqrt{99}+99\sqrt{100}}.$$

3. For n a positive integer and $0 < x \le \dfrac{\pi}{2}$, prove that $\cot \dfrac{x}{2^n} - \cot x \ge n$.

4. (1990 Hungarian MO) For positive integer n, show that

$$\sin^3 \frac{x}{3} + 3 \sin^3 \frac{x}{3^2} + \cdots + 3^{n-1} \sin^3 \frac{x}{3^n} = \frac{1}{4}\left(3^n \sin \frac{x}{3^n} - \sin x\right).$$

5. For a positive integer n, show that $\displaystyle\sum_{k=0}^{[n/4]} \binom{n}{4k} = 2^{n-2} + (\sqrt{2})^{n-2} \cos \frac{n\pi}{4}$.

*6. Prove that $\tan^2 1^\circ + \tan^2 3^\circ + \tan^2 5^\circ + \cdots + \tan^2 89^\circ = 4005$. (*Hint*: Find a polynomial of degree 45 having roots $\tan^2 1^\circ, \tan^2 3^\circ, \tan^2 5^\circ, \ldots, \tan^2 89^\circ$.)

*7. (1990 Austrian-Polish Math Competition) Let $n > 1$ be an integer and let $f_1, f_2, \ldots, f_{n!}$ be the $n!$ permutations of $1, 2, \ldots, n$. (Each f_i is a bijective function from $\{1, 2, \ldots, n\}$ to itself.) For each permutation f_i, let us define $S(f_i) = \sum_{k=1}^{n} |f_i(k) - k|$. Find $\frac{1}{n!}\sum_{i=1}^{n!} S(f_i)$.

*8. (1991 Canadian MO) Let n be a fixed positive integer. Find the sum of all positive integers with the following property: In base 2, it has exactly $2n$ digits consisting of n 1's and n 0's. (The leftmost digit cannot be 0.)

2. Inequalities (Part I)

We often compare numbers or math expressions, such as in finding maxima or minima or in applying the sandwich theorem. So we need to know some useful inequalities. Here we will look at some of these and see how they can be applied.

AM-GM-HM Inequality. For $a_1, a_2, \ldots, a_n > 0$,

$$AM = \frac{a_1 + a_2 + \cdots + a_n}{n} \geq GM = \sqrt[n]{a_1 a_2 \cdots a_n} \geq HM = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}}.$$

Either equality holds if and only if $a_1 = a_2 = \cdots = a_n$.

Examples. (1) By the AM-GM inequality, for $x > 0$, $x + \frac{1}{x} \geq 2\sqrt{x \cdot \frac{1}{x}} = 2$ with equalilty if and only if $x = 1$.

(2) By the AM-HM inequality, if $a_1, a_2, \ldots, a_n > 0$, then

$$(a_1 + a_2 + \cdots + a_n)\left(\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}\right) \geq n^2.$$

(3) If $a, b, c > 0$ and $abc = 1$, find the minimum of $(a + b + c)(ab + bc + ca)$.

Solution. By the AM-GM inequality,

$$\frac{a+b+c}{3} \geq \sqrt[3]{abc} = 1 \quad \text{and} \quad \frac{ab+bc+ca}{3} \geq \sqrt[3]{(ab)(bc)(ca)} = 1.$$

So $(a + b + c)(ab + bc + ca) \geq 9$ with equality if and only if $a = b = c = 1$. Therefore, the minimum is 9.

(4) For positive integer n, show that $\left(1 + \frac{1}{n}\right)^n \leq \left(1 + \frac{1}{n+1}\right)^{n+1}$.

Solution. Let $a_1 = a_2 = \cdots = a_n = 1 + \frac{1}{n}$, $a_{n+1} = 1$. By the AM-GM inequality,

$$AM = \frac{n(1 + \frac{1}{n}) + 1}{n+1} = 1 + \frac{1}{n+1} \geq GM = \sqrt[n+1]{\left(1 + \frac{1}{n}\right)^n \cdot 1}.$$

Taking the $(n+1)$-st power of both sides, we get $\left(1+\frac{1}{n+1}\right)^{n+1} \geq \left(1+\frac{1}{n}\right)^n$.

(5) (1964 IMO) Let a, b, c be the sides of a triangle. Prove that

$$a^2(b+c-a)+b^2(c+a-b)+c^2(a+b-c) \leq 3abc.$$

Solution. Let $x=\frac{a+b-c}{2}$, $y=\frac{b+c-a}{2}$, $z=\frac{c+a-b}{2}$, then $x, y, z>0$ and $a=z+x$, $b=x+y$, $c=y+z$. The inequality to be proved becomes

$$(z+x)^2 2y+(x+y)^2 2z+(y+z)^2 2x \leq 3(z+x)(x+y)(y+z).$$

This is equivalent to $x^2y+y^2z+z^2x+xy^2+yz^2+zx^2 \geq 6xyz$, which is true because $x^2+y^2z+z^2x+xy^2+yz^2+zx^2 \geq 6\sqrt[6]{x^6y^6z^6}=6xyz$ by the AM-GM inequality.

Cauchy-Schwarz Inequality. For real numbers $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$,

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

Equality holds if and only if $a_ib_j=a_jb_i$ for all $i, j=1, \ldots, n$.

Examples. (6) Find the maximum and minimum of $a\cos\theta+b\sin\theta$, where $0 \leq \theta < 2\pi$.

Solution. By the Cauchy-Schwarz inequality,

$$(a\cos\theta+b\sin\theta)^2 \leq (a^2+b^2)(\cos^2\theta+\sin^2\theta)=a^2+b^2.$$

So $-\sqrt{a^2+b^2} \leq a\cos\theta+b\sin\theta \leq \sqrt{a^2+b^2}$. Equality holds if and only if $a\sin\theta=b\cos\theta$, i.e. $\tan\theta=b/a$. There are two such θ's in $[0, 2\pi)$ corresponding to the left and right equalities. So the maximum is $\sqrt{a^2+b^2}$ and the minimum is $-\sqrt{a^2+b^2}$.

(7) (1978 USAMO) For real numbers a, b, c, d, e such that $a+b+c+d+e=8$ and $a^2+b^2+c^2+d^2+e^2=16$, find the maximum of e.

Solution. By the Cauchy-Schwarz inequality,

$$(a+b+c+d)^2 \leq (1^2+1^2+1^2+1^2)(a^2+b^2+c^2+d^2).$$

So $(8-e)^2 \leq 4(16-e^2)$. Expanding and simplifying, we get $e(5e-16) \leq 0$. This means $0 \leq e \leq 16/5$. Examining the equality case of the Cauchy-Schwarz inequality, we see that when $a=b=c=d=6/5$, e will attain the maximum value of $16/5$.

(8) (1995 IMO) If $a, b, c>0$ and $abc=1$, then prove that

$$\frac{1}{a^3(b+c)}+\frac{1}{b^3(c+a)}+\frac{1}{c^3(a+b)} \geq \frac{3}{2}.$$

Solution. Substituting $x=\frac{1}{a}=bc$, $y=\frac{1}{b}=ca$, $z=\frac{1}{c}=ab$, the inequality becomes

$$\frac{x^2}{z+y}+\frac{y^2}{x+z}+\frac{z^2}{y+x} \geq \frac{3}{2}.$$

Now $x+y+z=\frac{x}{\sqrt{z+y}}\sqrt{z+y}+\frac{y}{\sqrt{x+z}}\sqrt{x+z}+\frac{z}{\sqrt{y+x}}\sqrt{y+x}$. By the Cauchy-Schwarz inequality, we get

$$(x+y+z)^2 \leq \left(\frac{x^2}{z+y}+\frac{y^2}{x+z}+\frac{z^2}{y+x}\right)\underbrace{\big((z+y)+(x+z)+(y+x)\big)}_{=2(x+y+z)}.$$

Using the last inequality and the AM-GM inequality, we get

$$\frac{x^2}{z+y}+\frac{y^2}{x+z}+\frac{z^2}{y+x} \geq \frac{x+y+z}{2} \geq \frac{3\sqrt[3]{xyz}}{2}=\frac{3}{2}.$$

Rearrangement (or Permutation) Inequality. If $a_1 \geq a_2 \geq \cdots \geq a_n$ and $b_1 \geq b_2 \geq \cdots \geq b_n$, then

$$a_1b_1+a_2b_2+\cdots+a_nb_n \geq a_1b_{r_1}+a_2b_{r_2}+\cdots+a_nb_{r_n} \geq a_1b_n+a_2b_{n-1}+\cdots+a_nb_1,$$

where $b_{r_1}, b_{r_2}, \ldots, b_{r_n}$ is a permutation of $b_1, b_2, \ldots, b_n$.

Example. (9) (1978 IMO) Let $c_1, c_2, \ldots, c_n$ be distinct positive integers. Prove that

$$c_1 + \frac{c_2}{2^2} + \cdots + \frac{c_n}{n^2} \geq 1 + \frac{1}{2} + \cdots + \frac{1}{n}.$$

Solution. Let $a_1, a_2, \ldots, a_n$ be the c_i's arranged in increasing order. Since the a_i's are distinct positive integers, we have $a_1 \geq 1, a_2 \geq 2, \ldots, a_n \geq n$. Now, since $a_1 < a_2 < \cdots < a_n$ and $1 > \frac{1}{2^2} > \cdots > \frac{1}{n^2}$, by the rearrangement inequality, we get

$$c_1 + \frac{c_2}{2^2} + \cdots + \frac{c_n}{n^2} \geq a_1 + \frac{a_2}{2^2} + \cdots + \frac{a_n}{n^2} \geq 1 + \frac{2}{2^2} + \cdots + \frac{n}{n^2}.$$

(10) Redo example (8) using the rearrangement inequality.

Solution. (Due to Ho Wing Yip) We define x, y, z as in example (10). Without loss of generality, we may assume $x \geq y \geq z$ because the inequality is symmetric. Then $xyz = 1$, $x^2 \geq y^2 \geq z^2$ and $\frac{1}{z+y} \geq \frac{1}{x+z} \geq \frac{1}{y+x}$. By the rearrangement inequality,

$$\frac{x^2}{z+y} + \frac{y^2}{x+z} + \frac{z^2}{y+x} \geq \frac{x^2}{y+x} + \frac{y^2}{z+y} + \frac{z^2}{x+z},$$
$$\frac{x^2}{z+y} + \frac{y^2}{x+z} + \frac{z^2}{y+x} \geq \frac{x^2}{x+z} + \frac{y^2}{y+x} + \frac{z^2}{z+y}.$$

Adding these inequalities and dividing by 2, we get

$$\frac{x^2}{z+y} + \frac{y^2}{x+z} + \frac{z^2}{y+x} \geq \frac{1}{2}\Big(\frac{y^2+x^2}{y+x} + \frac{z^2+y^2}{z+y} + \frac{x^2+z^2}{x+z}\Big).$$

Applying the simple inequality $a^2+b^2 \geq (a+b)^2/2$ to the numerators of the right sides, then the AM-GM inequality, we get

$$\frac{x^2}{z+y} + \frac{y^2}{x+z} + \frac{z^2}{y+x} \geq \frac{1}{2}\big(\frac{y+x}{2} + \frac{z+y}{2} + \frac{x+z}{2}\big)$$
$$= \frac{x+y+z}{2} \geq \frac{3\sqrt[3]{xyz}}{2} = \frac{3}{2}.$$

Chebysev's Inequality. If $a_1 \geq a_2 \geq \cdots \geq a_n$ and $b_1 \geq b_2 \geq \cdots \geq b_n$, then

$$a_1b_1 + a_2b_2 + \cdots + a_nb_n \geq \frac{(a_1+a_2+\cdots+a_n)(b_1+b_2+\cdots+b_n)}{n}$$
$$\geq a_1b_n + a_2b_{n-1} + \cdots + a_nb_1.$$

Either equality holds if and only if $a_1 = a_2 = \cdots = a_n$ or $b_1 = b_2 = \cdots = b_n$.

Examples. (11) (1974 USAMO) For $a, b, c > 0$, prove that $a^a b^b c^c \geq (abc)^{(a+b+c)/3}$.

Solution. By symmetry, we may assume $a \geq b \geq c$. Then $\log a \geq \log b \geq \log c$. By Chebysev's inequality,

$$\log(a^a b^b c^c) = a\log a + b\log b + c\log c$$
$$\geq \frac{(a+b+c)(\log a + \log b + \log c)}{3} = \log(abc)^{\frac{a+b+c}{3}}.$$

The desired inequality follows by exponentiation.

(12) If $0 \leq a_k < 1$ for $k = 1, 2, \ldots, n$ and $S = a_1 + a_2 + \cdots + a_n$, then prove that

$$\sum_{k=1}^{n} \frac{a_k}{1-a_k} \geq \frac{nS}{n-S}.$$

Solution. Without loss of generality, we may assume $a_1 \geq a_2 \geq \cdots \geq a_n \geq 0$. Then $0 < 1-a_1 \leq 1-a_2 \leq \cdots \leq 1-a_n$ and $\frac{a_1}{1-a_1} \geq \frac{a_2}{1-a_2} \geq \cdots \geq \frac{a_n}{1-a_n}$. By Chebysev's inequality,

$$S = \frac{a_1}{1-a_1}(1-a_1) + \frac{a_2}{1-a_2}(1-a_2) + \cdots + \frac{a_n}{1-a_n}(1-a_n)$$
$$\leq \frac{1}{n}\sum_{k=1}^{n}\frac{a_k}{1-a_k}\sum_{k=1}^{n}(1-a_k) = \frac{n-S}{n}\sum_{k=1}^{n}\frac{a_k}{1-a_k}.$$

The result follows.

In math as well as in statistics, we often need to take averages (or means) of numbers. Other than AM, GM, HM, there are so-called power means and symmetric means, which include AM and GM as special cases.

Power Mean Inequality. For $a_1, a_2, \ldots, a_n > 0$ and $s < t$,

$$M_s = \left(\frac{a_1^s + a_2^s + \cdots + a_n^s}{n}\right)^{1/s} \le M_t = \left(\frac{a_1^t + a_2^t + \cdots + a_n^t}{n}\right)^{1/t}.$$

Equality holds if and only if $a_1 = a_2 = \cdots = a_n$.

Remarks. Clearly, $M_1 = AM$ and $M_{-1} = HM$. Now $M_2 = \sqrt{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n}}$ is called the *root-mean-square* (RMS) of the numbers. It appears in statistics and physics. Also, taking limits, it can be shown that $M_{+\infty}$ is $MAX = \max\{a_1, a_2, \ldots, a_n\}$, M_0 is GM and $M_{-\infty}$ is $MIN = \min\{a_1, a_2, \ldots, a_n\}$. So we have

$$MAX \ge RMS \ge AM \ge GM \ge HM \ge MIN.$$

Maclaurin's Symmetric Mean Inequality. For $a_1, a_2, \ldots, a_n > 0$,

$$AM = S_1 \ge S_2^{1/2} \ge \cdots \ge S_n^{1/n} = GM,$$

where S_j is the average of all possible products of $a_1, a_2, \ldots, a_n$ taken j at a time. Any one of the equalities holds if and only if $a_1 = a_2 = \cdots = a_n$.

Remarks. To be clear on the meaning of S_j, take $n = 4$. In that case, we have

$$S_1 = \frac{a_1 + a_2 + a_3 + a_4}{4}, \quad S_2 = \frac{a_1a_2 + a_1a_3 + a_1a_4 + a_2a_3 + a_2a_4 + a_3a_4}{6},$$

$$S_3 = \frac{a_1a_2a_3 + a_1a_2a_4 + a_1a_3a_4 + a_2a_3a_4}{4} \quad \text{and} \quad S_4 = a_1a_2a_3a_4.$$

Examples. (13) Show that $x^5 + y^5 + z^5 \le x^5\sqrt{\frac{x^2}{yz}} + y^5\sqrt{\frac{y^2}{zx}} + z^5\sqrt{\frac{z^2}{xy}}$ for positive x, y, z.

Solution. Let $a = \sqrt{x}, b = \sqrt{y}, c = \sqrt{z}$, then the inequality becomes

$$a^{10} + b^{10} + c^{10} \le \frac{a^{13} + b^{13} + c^{13}}{abc}.$$

Now $a^{13} + b^{13} + c^{13} = 3M_{13}^{13} = 3M_{13}^{10}M_{13}^3 \ge 3M_{10}^{10}M_0^3 = (a^{10} + b^{10} + c^{10})abc$.

(14) If $a, b, c > 0$, then prove that $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} \le \frac{a^8 + b^8 + c^8}{a^3b^3c^3}$.

Solution. The inequality is equivalent to

$$a^8 + b^8 + c^8 \ge a^3b^3c^3(\frac{1}{a} + \frac{1}{b} + \frac{1}{c}) = (abc)^2(bc + ca + ab).$$

By the power mean inequality and the symmetric mean inequality,

$$\begin{aligned} a^8 + b^8 + c^8 = 3M_8^8 &\ge 3M_1^8 = 3S_1^8 = 3S_1^6S_1^2 \\ &\ge (S_3^{1/3})^6 3(S_2^{1/2})^2 = (abc)^2(bc + ca + ab). \end{aligned}$$

Multiplying by 3 on both sides, we are done.

(15) If $a_1, a_2, \ldots, a_n \ge 0$ and $(1 + a_1)(1 + a_2)\cdots(1 + a_n) = 2^n$, then show that $a_1a_2\cdots a_n \le 1$.

Solution. By the symmetric mean inequality,

$$\begin{aligned} 2^n &= (1 + a_1)(1 + a_2)\cdots(1 + a_n) \\ &= 1 + nS_1 + \binom{n}{2}S_2 + \cdots + \binom{n}{n-1}S_{n-1} + S_n \\ &\ge 1 + nS_n^{1/n} + \binom{n}{2}S_n^{2/n} + \cdots + \binom{n}{n-1}S_n^{(n-1)/n} + S_n = (1 + S_n^{1/n})^n. \end{aligned}$$

So $2 \ge 1 + S_n^{1/n}$. Then $a_1a_2\cdots a_n = S_n \le 1$.

Exercises

1. Redo example (11).

2. Redo example (13).

3. Redo example (15).

4. For $x_1, x_2, \ldots, x_n > 0$, show that $\frac{x_1^2}{x_2} + \frac{x_2^2}{x_3} + \cdots + \frac{x_n^2}{x_1} \geq x_1 + x_2 + \cdots + x_n$.

5. For $0 < a, b, c < 1$ and $a+b+c = 2$, show that $8(1-a)(1-b)(1-c) \leq abc$.

6. If $a, b, c, d > 0$ and $c^2 + d^2 = (a^2 + b^2)^3$, then show that $\frac{a^3}{c} + \frac{b^3}{d} \geq 1$.

7. For $a_1, a_2, \ldots, a_n > 0$ and $a_1 + a_2 + \cdots + a_n = 1$, find the minimum of

$$(a_1 + \frac{1}{a_1})^2 + (a_2 + \frac{1}{a_2})^2 + \cdots + (a_n + \frac{1}{a_n})^2.$$

8. If $a, b, c, d > 0$ and $S = a^2 + b^2 + c^2 + d^2$, then show that

$$\frac{a^3+b^3+c^3}{a+b+c} + \frac{a^3+b^3+d^3}{a+b+d} + \frac{a^3+c^3+d^3}{a+c+d} + \frac{b^3+c^3+d^3}{b+c+d} \geq S.$$

9. If $x_1, x_2, \ldots, x_n > 0$ and $x_1 + x_2 + \cdots + x_n = 1$, then show that

$$\sum_{k=1}^{n} \frac{x_k}{\sqrt{1-x_k}} \geq \frac{1}{\sqrt{n-1}} \sum_{k=1}^{n} \sqrt{x_k}.$$

10. Let a, b, c be the sides of a triangle. Show that

$$a^2b(a-b) + b^2c(b-c) + c^2a(c-a) \geq 0.$$

3. Number Theory

§1 Divisibility.

Definitions. (i) If a, b, c are integers such that $a = bc$ and $b \neq 0$, then we say b *divides* a and denote this by $b|a$. (For example, 2 divides 6, so we write $2|6$.)

(ii) A positive integer $p > 1$ is a *prime number* if 1 and p are the only positive integers dividing p. If a positive integer $n > 1$ is not prime, it is a *composite number.*

There is a famous proof of the fact that there are infinitely many prime numbers. It goes as follow. Suppose there are only finitely many prime numbers, say they are $p_1, p_2, \ldots, p_n$. Then the number $M = p_1 p_2 \cdots p_n + 1$ is greater than $p_1, p_2, \ldots, p_n$. So M cannot be prime, hence there is a prime number p_i dividing M. However, p_i also divides $M - 1$. Hence p_i will divide $M - (M-1) = 1$, a contradiction.

Fundamental Theorem of Arithmetic (or Prime Factorization Theorem). Every positive integer n can be written as the product of prime powers $n = 2^{e_1}3^{e_2}5^{e_3}7^{e_4} \cdots p_k^{e_k}$, where the e_i's are nonnegative integers, in one and only one way (except for reordering of the primes).

Examples. $90 = 2^1 3^2 5^1$ and $924 = 2^2 3^1 7^1 11^1$.

Questions. Do positive rational numbers have prime factorizations? (Yes, if exponents are allowed to be any integers.) Do positive real numbers have prime factorizations (allowing rational exponents)? (No, π does not.)

Corollaries. (1) $m = p_1^{d_1} p_2^{d_2} \cdots p_k^{d_k}$ divides $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ if and only if $0 \leq d_i \leq e_i$ for $i = 1, 2, \cdots, k$.
(2) The number $n = 2^{e_1}3^{e_2} \cdots p_k^{e_k}$ has exactly $(e_1+1)(e_2+1)\cdots(e_k+1)$ positive divisors.
(3) A positive integer n is the m-th power of a positive integer b (i.e. $n = b^m$) if and only if in the prime factorization of $n = 2^{e_1}3^{e_2}5^{e_3} \ldots p_k^{e_k}$, every e_i is a multiple of m.

Examples. (1) Since $90 = 2^1 3^2 5^1$, it has $(1+1)(2+1)(1+1) = 12$ positive divisors. They are $2^{d_1}3^{d_2}5^{d_3}$, where $d_1 = 0, 1$, $d_2 = 0, 1, 2$ and $d_3 = 0, 1$.

(2) Suppose n is a positive integer such that $2n$ has 28 positive divisors and $3n$ has 30 positive divisors. How many positive divisors does $6n$ have?

Solution. Write $n = 2^{e_1}3^{e_2}\cdots p_k^{e_k}$. Then $(e_1+2)(e_2+1)\cdots(e_k+1) = 28$ and $(e_1+1)(e_2+2)\cdots(e_k+1) = 30$. Now $a = (e_3+1)\cdots(e_k+1)$ divides 28 and 30, so it must be 1 or 2. If $a = 1$, then $(e_1+2)(e_2+1) = 28$ and $(e_1+1)(e_2+2) = 30$, which have the unique solution $e_1 = 5$, $e_2 = 3$. It follows $6n$ has $(e_1+2)(e_2+2)a = 35$ positive divisors. If $a = 2$, then $(e_1+2)(e_2+1) = 14$ and $(e_1+1)(e_2+2) = 15$, which have no integer solutions by simple checking.

---

(3) (1985 IMO) Given a set M of 1985 distinct positive integers, none of which has a prime divisor greater than 26. Prove that M contains at least one subset of four distinct elements whose product is the fourth power of an integer.

Solution. Let $M = \{n_1, n_2, n_3, \ldots, n_{1985}\}$. Taking prime factorizations, suppose $n_i = 2^{e_{1,i}}3^{e_{2,i}}5^{e_{3,i}}\cdots 23^{e_{9,i}}$. Since 23 is the ninth prime number, there are $2^9 = 512$ possible parity (i.e. odd-even) patterns for the numbers $e_{i,1}, e_{2,i}, e_{3,i}, \ldots, e_{9,i}$. So among any 513 of them, there will be two (say n_i, n_j) with the same pattern. Then $n_in_j = b_{ij}^2$. Note b_{ij} cannot have any prime divisor greater than 26.

Remove these pairs one at a time. Since $1985 - 2\times 512 = 961 > 513$, there are at least 513 pairs. Consider the b_{ij}'s for these pairs. There will be two (say b_{ij}, b_{kl}) such that $b_{ij}b_{kl} = c^2$. Then $n_in_jn_kn_l = b_{ij}^2b_{kl}^2 = c^4$.

Definitions. Let $a_1, a_2, \ldots, a_n$ be integers, not all zeros.

(i) The *greatest common divisor* (or *highest common factor*) of $a_1, a_2, \ldots, a_n$ is the largest positive integer dividing all of them. We denote this number by $(a_1, a_2, \ldots, a_n)$ or $\gcd(a_1, a_2, \ldots, a_n)$. If $(a_1, a_2, \ldots, a_n) = 1$, then we say $a_1, a_2, \ldots, a_n$ are *coprime* or *relatively prime.* In particular, two coprime integers have no common prime divisors!

(ii) The *least common multiple* of $a_1, a_2, \ldots, a_n$ is the least positive integer which is a multiple of each of them. We denote this number by $[a_1, a_2, \ldots, a_n]$ or $\mathrm{lcm}(a_1, a_2, \ldots, a_n)$.

Example. (4) $(6, 8) = 2$, $[6, 8] = 24$; $(6, 8, 9) = 1$, $[6, 8, 9] = 72$.

Theorem. If $a_i = 2^{e_{1,i}}3^{e_{2,i}}\cdots p_k^{e_{k,i}}$, then

$$(a_1, a_2, \ldots, a_n) = 2^{\min\{e_{1,i}\}}3^{\min\{e_{2,i}\}}\cdots p_k^{\min\{e_{k,i}\}}$$

and

$$[a_1, a_2, \ldots, a_n] = 2^{\max\{e_{1,i}\}}3^{\max\{e_{2,i}\}}\cdots p_k^{\max\{e_{k,i}\}}.$$

For $n = 2$, $(a_1, a_2)[a_1, a_2] = a_1a_2$. (The last equation need not be true for more than 2 numbers.)

Example. (5) $6 = 2^13^1$, $8 = 2^33^0$, so $(6, 8) = 2^13^0 = 2$, $[6, 8] = 2^33^1 = 24$.

Prime factorization is difficult for large numbers. So to find gcd's, we can also use the following fact.

Euclidean Algorithm. If a, b are integers not both zeros, then $(a, b) = (a - bm, b) = (a, b - an)$ for any positive integers m, n. In particular, if $a > b > 0$ and $a = bm + r$, then $(a, b) = (r, b)$.

Examples. (6) $(2445, 652) = (489, 652) = (489, 163) = 163$.

---

(7) (IMO 1959) Prove that the fraction $\frac{21n+4}{14n+3}$ is irreducible for every natural number n.

Solution. $(21n + 4, 14n + 3) = (7n + 1, 14n + 3) = (7n + 1, 1) = 1$.

The following are some useful facts about relatively prime integers.

(1) For nonnegative integers a, b not both zeros, (a, b) is the least positive integer of the form $am + bn$, where m, n are integers. In particular, if $(a, b) = 1$, then there are integers m, n such that $am + bn = 1$.

Reasons. Clearly (a, b) divides positive numbers of the form $am+bn$, hence $(a, b) \le am + bn$. By symmetry, we may assume $a \ge b$. We will induct on a. If $a = 1$, then $b = 0$ or 1 and $(a, b) = 1 = a \cdot 1 + b \cdot 0$. Suppose this is true for all cases $a < a_0$. By Euclidean algoritnm, $(a_0, b) = (r, b)$, where $a_0 = bq + r$, $0 \le r < b$. Since $b < a_0$, by the inductive hypothesis, there are integers m, n such that $(a_0, b) = (r, b) = rm + bn = (a_0 - bq)m + bn = a_0m + b(n - qm)$. So the case $a = a_0$ is true.

(2) If $n|ab$ and $(a, n) = 1$, then $n|b$. (This is because $ar + ns = 1$ for some integers r, s so that $n|(ab)r + n(sb) = b$.) In particular, if p is prime and $p|ab$, then $p|a$ or $p|b$. From this, we get that if $(a, n) = 1$ and $(b, n) = 1$, then $(ab, n) = 1$.

(3) If $n^k = ab$ and $(a, b) = 1$, then each of a and b is the k-th power of an integer. (This follows from taking prime factorization of n and using the second part of the last fact.)

§2 Modulo Arithmetic.

Division Algorithm. Let b be a positive integer. For any integer a, there are integers q, r such that $a = bq + r$ and $0 \leq r < b$. (r is called the *remainder* of a upon division by b. Remainders are always nonnegative.)

Note that when 19 is divided by 5, the remainder is 4, but when -19 is divided by 5, the remainder is 1 because $-19 = 5(-4) + 1$. When the integers $\ldots, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, \ldots$ is divided by 5, the respective remainders form the periodic sequence

$$\ldots, 4, 0, 1, 2, 3, 4, 0, 1, 2, 3, 4, 0, 1, 2, \ldots.$$

Definitions. (i) We say a is *congruent* to a' modulo b and denote this by $a \equiv a' (\mathrm{mod}\ b)$ if and only if a and a' have the same remainder upon division by b.

(ii) For a positive integer n, a *complete set of residues modulo n* is a set of n integers $r_1, r_2, \ldots, r_n$ such that every integer is congruent to exactly one of $r_1, r_2, \ldots, r_n$ modulo n. (For example, $0, 1, 2, \ldots, n-1$ form a complete set of residues modulo n for every positive integer n.)

Basic Properties. (i) $a \equiv a' (\mathrm{mod}\ b)$ if and only if $b | a - a'$. (Often this is used as the definition of the congruent relation.)

(ii) If $a \equiv a' (\mathrm{mod}\ b)$ and $c \equiv c' (\mathrm{mod}\ b)$, then $a + c \equiv a' + c' (\mathrm{mod}\ b)$, $a - c \equiv a' - c' (\mathrm{mod}\ b)$, $ac \equiv a'c' (\mathrm{mod}\ b)$, $a^n \equiv a'^n (\mathrm{mod}\ b)$ and $P(a) \equiv P(a') (\mathrm{mod}\ b)$ for any polynomial $P(x)$ with integer coefficients.

Example. (8) Find the remainder of 1978^{20} upon division by $5^3 = 125$.

Solution. $1978^{20} = (2000 - 22)^{20} \equiv (-22)^{20} = 484^{10} \equiv (-16)^{10} = 256^5 \equiv 6^5 = 2^5 3^5 \equiv 32(-7) \equiv 26 (\mathrm{mod}\ 125)$.

Other than finding remainders, modulo arithmetic is also useful in many situations. For example, (mod 2) is good for parity check. The fact that a number

is divisible by 3 if and only if the sum of digits is divisible by 3 can be easily explained by

$$a_n 10^n + a_{n-1} 10^{n-1} + \cdots + a_1 10 + a_0 \equiv a_n + a_{n-1} + \cdots + a_1 + a_0 (\mathrm{mod}\ 3).$$

In working with squares, (mod 4) is useful in doing parity check since $(2n)^2 = 4n^2 \equiv 0 (\mathrm{mod}\ 4)$ and $(2n+1)^2 = 4n^2 + 4n + 1 \equiv 1 (\mathrm{mod}\ 4)$. Similarly for cubes, we have $k^3 \equiv -1, 0$ or $1 (\mathrm{mod}\ 9)$ according to $k \equiv -1, 0, 1 (\mathrm{mod}\ 3)$ respectively. For fourth powers, $k^4 \equiv 0$ or $1\ (\mathrm{mod}\ 16)$ according to k is even or odd respectively. Also, as in the reasoning above for (mod 3), one can show that every nonnegative integer in base 10 is congruent to the sum of its digits (mod 9). To determine the units digits, we use (mod 10).

The following facts are very useful in dealing with some problems.

Further Properties. (iii) (Cancellation Property) If $am \equiv am' (\mathrm{mod}\ b)$ and $(a, b) = 1$, then $b | a(m - m')$, so $b | m - m'$, i.e. $m \equiv m' (\mathrm{mod}\ b)$.

(iv) (Existence of Multiplicative Inverse) If $(a, b) = 1$, then there exists a unique $m (\mathrm{mod}\ b)$ such that $am \equiv 1 (\mathrm{mod}\ b)$. We may denote this m by a^{-1}. (*Reasons.* Since $(a, b) = 1$, there exist integers m, n such that $1 = (a, b) = am + bn \equiv am (\mathrm{mod}\ b)$. If $am' \equiv 1 (\mathrm{mod}\ b)$, then $m \equiv m' (\mathrm{mod}\ b)$ by the cancellation property.)

(v) For a positive integer c, if $(a, b) = 1$, then we define $a^{-c} \equiv (a^{-1})^c (\mathrm{mod}\ b)$. Since $a^c (a^{-1})^c \equiv 1 (\mathrm{mod}\ b)$, so we also have $a^{-m} \equiv (a^c)^{-1} (\mathrm{mod}\ b)$. From this, we can check that $a^{r+s} \equiv a^r a^s (\mathrm{mod}\ b)$ and $(a^r)^s \equiv a^{rs} (\mathrm{mod}\ b)$ for all integers r, s.

(vi) For nonnegative integers a, b not both zeros, if $r^a \equiv 1 (\mathrm{mod}\ s)$ and $r^b \equiv 1 (\mathrm{mod}\ s)$, then there are integers m, n isuch that $r^{(a,b)} = r^{am+bn} = (r^a)^m (r^b)^n \equiv 1 (\mathrm{mod}\ s)$.

Fermat's Little Theorem. If p is prime and $(a, p) = 1$, then $a^{p-1} \equiv 1 (\mathrm{mod}\ p)$. (This means that $a^{p-2} \equiv a^{-1}\ (\mathrm{mod}\ p)$ because $a(a^{p-2}) \equiv 1 (\mathrm{mod}\ p)$.)

Euler's Theorem. If $(a, n) = 1$, then $a^{\phi(n)} \equiv 1 (\mathrm{mod}\ n)$, where the Euler ϕ-function $\phi(n)$ is the number of positive integers less than or equal to n, which are relatively prime to n. (For a prime p, $\phi(p) = p - 1$ and hence Euler's theorem generalizes Fermat's little theorem.)

To understand the reasons behind these theorems, we will define a *reduced set of residues modulo n* to be a set of $\phi(n)$ integers $r_1, r_2, \ldots, r_{\phi(n)}$ such that every integer relatively prime to n is congruent to exactly one of $r_1, r_2, \ldots, r_{\phi(n)}$ modulo n. To prove Euler's theorem, we note that if $r_1, r_2, \ldots, r_{\phi(n)}$ is a reduced set of residues modulo n and $(a, n) = 1$, then $ar_1, ar_2, \ldots, ar_{\phi(n)}$ are relatively prime to n and they also form a reduced set of residues modulo n. (This is because $r_i = r_j$ if and only if $ar_i \equiv ar_j \pmod{n}$ by the properties above.) So each ar_i is congruent to a unique r_j modulo n and hence

$$a^{\phi(n)} r_1 r_2 \cdots r_{\phi(n)} = (ar_1)(ar_2)\cdots(ar_{\phi(n)}) \equiv r_1 r_2 \cdots r_{\phi(n)} \pmod{n}.$$

Since $(r_1 r_2 \cdots r_{\phi(n)}, n) = 1$, applying the cancellation property, we get $a^{\phi(n)} \equiv 1 \pmod{n}$.

Wilson's Theorem. If p is a prime number, then $(p-1)! \equiv -1 \pmod{p}$. (The converse is also true, if $n > 4$ is composite, then $(n-1)! \equiv 0 \pmod{n}$.)

Reasons. The case $p = 2$ is clear. Let $p > 2$ be prime and $1 \le a < p$, then $(a, p) = 1$ implies $a, 2a, \ldots, a(p-1)$ form a reduced set of residues modulo p and hence there is a unique b such that $1 \le b < p$ and $ab \equiv 1 \pmod{p}$. We have $a = b$ if and only if p divides $a^2 - 1 = (a-1)(a+1)$, that is $a = 1$ or $p-1$. Hence, for the $p-3$ integers $2, 3, \ldots, p-2$, we can form $(p-3)/2$ pairs a, b with $ab \equiv 1 \pmod{p}$. Then $(p-1)! \equiv 1 \cdot 1^{(p-3)/2}(p-1) \equiv -1 \pmod{p}$.

For the converse of Wilson's theorem, if $n > 4$ is composite, then let p be a prime dividing n and supoose $n \neq p^2$. Then $p < n$ and $n/p < n$ so that $p \neq n/p$ and $n = p(n/p)$ divides $(n-1)!$. If $n = p^2$, then $p \neq 2$ and $p < 2p < p^2 = n$ so again n divides $2p^2 = p(2p)$, which divides $(n-1)!$.

Examples. (9) $\phi(1) = 1$. If p is prime, then $1, 2, 3, \ldots, p-1$ are re;atively prime to p and so $\phi(p) = p - 1$. (Again, this means Fermat's little theorem is a special case of Euler's theorem). For $k \ge 1$, p prime, since the numbers $p, 2p, 3p, \ldots, p^k$ are the only numbers less than p^k not relatively prime to p^k, so $\phi(p^k) = p^k - p^{k-1} = p^k(1 - \frac{1}{p})$.

---

(10) Let us find the units digit of 7^{7^7}. Note $\phi(10) = 4$, since only 1,3,7,9 are less than or equal to 10 and relatively prime to 10. Since $(7, 10) = 1$, so $7^{\phi(10)} =$

$7^4 \equiv 1 \pmod{10}$ by Euler's theorem. Now $7^7 \equiv (-1)^7 \equiv 3 \pmod{4}$ and so $7^{7^7} = 7^{4n+3} = (7^4)^n 7^3 \equiv 1^n 7^3 = 343 \equiv 3 \pmod{10}$. So the units digit of 7^{7^7} is 3.

Chinese Remainder Theorem. Let $m_1, m_2, \ldots, m_k$ be positive integers such that $(m_i, m_j) = 1$ for every pair $i \neq j$. Then the equations $x \equiv b_1 \pmod{m_1}$, $x \equiv b_2 \pmod{m_2}, \ldots, x \equiv b_k \pmod{m_k}$ have a common solution. In fact, every two solutions are congruent $\pmod{m_1 m_2 \cdots m_k}$ and we say the solution is unique $\pmod{m_1 m_2 \cdots m_k}$.

Reasons. One solution x can be found as follow: let $M_j = \frac{m_1 m_2 \ldots m_k}{m_j}$, then

$$x = M_1^{\phi(m_1)} b_1 + M_2^{\phi(m_2)} b_2 + \cdots + M_k^{\phi(m_k)} b_k$$

is a solution since $m_i | M_j$ for $i \neq j$, $(M_i, m_i) = 1$ imply $x \equiv M_i^{\phi(m_i)} b_i \equiv b_i \pmod{m_i}$ for $i = 1, 2, \ldots, k$ by Euler's theorem. Next, to show the solution is unique $\pmod{m_1 m_2 \cdots m_k}$, let $x' \equiv b_i \pmod{m_i}$ also. Then $x - x' \equiv 0 \pmod{m_i}$ for $i = 1, 2, \ldots, k$, i.e. $x - x'$ is a common multiple of the m_i's. Since $(m_i, m_j) = 1$ for $i \neq j$, so their lcm $m_1 m_2 \cdots m_k | x - x'$. Then $x \equiv x' \pmod{m_1 m_2 \cdots m_k}$.)

Computation Formulas. If $(a, b) = 1$, then $\phi(ab) = \phi(a)\phi(b)$. If $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ and $e_i \ge 1$ for $i = 1, 2, \ldots, k$, then

$$\phi(n) = n(1 - \frac{1}{p_1})(1 - \frac{1}{p_2}) \cdots (1 - \frac{1}{p_k}).$$

Reasons. For the first statement, observe that if $1 \le x \le ab$ and $(x, ab) = 1$, then $(x, a) = 1$ and $(x, b) = 1$. So the remainders r, s of x upon divisions by a, b are relatively prime to a, b, respectively by the Euclidean algorithm. Conversely, if $1 \le r \le a, (a, r) = 1$ and $1 \le s \le b, (b, s) = 1$, then $x \equiv r \pmod{a}$ and $x \equiv s \pmod{b}$ have a unique solution less than or equal to ab by the Chinese remainder theorem. Thus, the pairing $x \leftrightarrow (r, s)$ is a one-to-one correspondence. The second statement follows from the first statement and the fact $\phi(p_i^k) = p_i^k(1 - \frac{1}{p_i})$.

Examples. (11) $\phi(100) = \phi(2^2 5^2) = 100(1 - \frac{1}{2})(1 - \frac{1}{5}) = 40$.

---

(12) To solve the system $x \equiv 3 \pmod{7}$, $x \equiv 2 \pmod{5}$, we may use the formula in the paragraph below the Chinese remainder theorem to get $x \equiv 5^{\phi(7)} 3 + 7^{\phi(5)} 2 =$

$5^6 3 + 7^4 2 \equiv 17 (\mathrm{mod}\ 35)$. (However, in general the formula may involve large numbers.) Alternatively, we can solve as follow: $x \equiv 3 (\mathrm{mod}\ 7) \Leftrightarrow x = 7n + 3$,

$$x \equiv 2 (\mathrm{mod}\ 5) \Leftrightarrow 7n + 3 \equiv 2 (\mathrm{mod}\ 5) \Leftrightarrow 7n \equiv -1 \equiv 4 (\mathrm{mod}\ 5) \Leftrightarrow$$

$$n \equiv 3(7n) \equiv 2 (\mathrm{mod}\ 5) \Leftrightarrow n = 5k + 2.$$

Then $x = 7(5k + 2) + 3 = 35k + 17$ (or $x \equiv 17 (\mathrm{mod}\ 35)$.)

---

(13) (IMO 1989) Prove that for each positive integer n there exist n consecutive positive integers, none of which is an integral power of a prime number.

Solution. Let $p_1, p_2, \ldots, p_{2n-1}, p_{2n}$ be $2n$ distinct prime numbers. Now by the Chinese remainder theorem,

$$x \equiv -1 (\mathrm{mod}\ p_1 p_2), \quad x \equiv -2 (\mathrm{mod}\ p_3 p_4), \quad \ldots, \quad x \equiv -n (\mathrm{mod}\ p_{2n-1} p_{2n})$$

have a common solution. Then each of the numbers $x + 1, x + 2, \ldots, x + n$ is divisible by two different prime numbers. Hence each cannot be a prime power.

---

(14) If q is a prime factor of $a^2 + b^2$ and $q \equiv 3 (\mathrm{mod}\ 4)$, then $q|a$ and $q|b$. (This fact is sometimes useful, for example in exercises 11 and 22.)

Solution. Suppose q does not divide a, say. Then $(q, a) = 1$. Let $c = a^{q-2}$, then $ac = a^{q-1} \equiv 1 (\mathrm{mod}\ q)$ by Fermat's little theorem. Now $q|a^2 + b^2$ implies $b^2 \equiv -a^2 (\mathrm{mod}\ q)$, so $(bc)^2 \equiv -1 (\mathrm{mod}\ q)$. Then q does not divide bc and $(bc)^{q-1} \equiv (-1)^{(q-1)/2} = -1 (\mathrm{mod}\ q)$, contradicting Fermat's little theorem. So $q|a$ and similarly $q|b$.

---

(15) (1978 IMO) Let m and n be natural numbers with $1 \leq m < n$. In their decimal representations, the last three digits of 1978^m are equal, respectively, to the last three digits of 1978^n. Find m and n such that $m + n$ has its least value.

Solution. Since the last three digits are equal, so $1978^n \equiv 1978^m (\mathrm{mod}\ 1000)$, i.e.

$$1000 = 2^3 5^3 | 1978^n - 1978^m = 1978^m (1978^{n-m} - 1).$$

So, $2^3 | 1978^m$ (because $1978^{n-m} - 1$ is odd) and the least m is 3. Let $d = n - m$. The problem now is to look for the least positive integer d such that $5^3 | 1978^d - 1$ (i.e. $1978^d \equiv 1 (\mathrm{mod}\ 5^3)$.) Since $\phi(5^3) = 100$, so $1978^{100} \equiv 1 (\mathrm{mod}\ 5^3)$ by Euler's theorem. Thus the least such d is at most 100.

Suppose the least $d < 100$. Let $100 = dq + r$ with $0 \leq r < d$, then $1978^r \equiv (1978^d)^q 1978^r = 1978^{100} \equiv 1 (\mathrm{mod}\ 5^3)$. Since d is to be the least such exponent, r must be 0. Then $d|100$. Also, $5|1978^d - 1$, so $1 \equiv 1978^d \equiv 3^d (\mathrm{mod}\ 5)$. The only such d's are multiples of 4. So $d = 4$ or 20. However, example (8) shows that $1978^{20} \not\equiv 1 (\mathrm{mod}\ 5^3)$. (This also shows that $1978^4 \not\equiv 1 (\mathrm{mod}\ 5^3)$ because $1978^4 \equiv 1 (\mathrm{mod}\ 5^3)$ implies $1978^{20} = (1978^4)^5 \equiv 1 (\mathrm{mod}\ 5^3)$.) So $d \geq 100$. Therefore the least $d = 100$ and the least $m + n = d + 2m = 106$ when $m = 3$ and $n = d + m = 103$.

Exercises

1. For each of the following statements, determine if each is true or false. If true, give an explanation. If false, provide a counterexample.
   (a) If p is a prime number and $p|n^k$, then $p^k|n^k$.
   (b) If $(ab, c) = 1$, then $(a, c) = 1$ and $(b, c) = 1$.
   (c) If $a^2 \equiv b^2 (\mathrm{mod}\ c^2)$, then $a \equiv b (\mathrm{mod}\ c)$.

2. As in the last exercise, determine if each of the following statement is true or false. Provide reason or counterexample.
   (a) If p is an odd prime, then $\phi(p^2 - 1) = \phi(p - 1)\phi(p + 1)$. How about $\phi(p^2 - 4) = \phi(p - 2)\phi(p + 2)$?
   (b) If $n > 1$, then show that the sequence $n, \phi(n), \phi(\phi(n)), \phi(\phi(\phi(n)))$, $\phi(\phi(\phi(\phi(n))))$, $\ldots$ must be all 1's after the n-th term.

3. If p is a prime number, show that p divides the binomial coefficients $C_n^p = \frac{p!}{n!(p-n)!}$ for $n = 1, 2, \ldots, p - 1$.

4. Find all intergers x such that $x \equiv 1 (\mathrm{mod}\ 2)$, $x \equiv 2 (\mathrm{mod}\ 3)$ and $x \equiv 3 (\mathrm{mod}\ 5)$.

5. Compute the last 2 digits of 7^{7^7}.(*Hint*: Consider (mod 4) and (mod 25).)

6. (1972 USAMO) Prove that for any positive integers a, b, c,

$$\frac{[a,b,c]^2}{[a,b][b,c][c,a]} = \frac{(a,b,c)^2}{(a,b)(b,c)(c,a)}.$$

7. Show that the greatest power of a prime number p dividing $n!$ is

$$\sum_{k=1}^{\infty}[\frac{n}{p^k}] = [\frac{n}{p}] + [\frac{n}{p^2}] + [\frac{n}{p^3}] + \cdots,$$

where $[x]$ is the greatest integer less than or equal to x.

8 (1972 IMO 1972) Let m and n be arbitrary non-negative integers. Prove that

$$\frac{(2m)!(2n)!}{m!n!(m+n)!}$$

is an integer. (*Hint*: One solution uses the last exercise. Another solution is to get a recurrence relation.)

9. Do there exist 21 consecutive positive integers each of which is divisible by one or more primes p from the interval $2 \le p \le 13$?

10. Show that there are infinitely many prime numbers of the form $4n-1$. (*Hint*: Modify the proof that there are infinitely many prime numbers.)

11. Show that there are infinitely many prime numbers of the form $4n+1$. (*Hint*: Use example 14.)

Remarks. There is a famous theorem called *Dirichlet's Theorem on Prime Progression,* which states that for every pair of relatively prime positive integers a, b, the arithmetic progression $a, a+b, a+2b, a+3b, \ldots$ must contain infinitely many prime numbers.

Another famous theorem known as Chebysev's theorem asserts that for every $x > 1$, there is always a prime number p between x and $2x$. This is

also called *Bertrand's Postulate* because it was experimentally verified by Bertrand for x from 1 to 1, 000, 000 before Chebysev proved it.

---

§3 Divisibility Problems

Examples. (16) (1998 IMO) Determine all pairs (a, b) of positive integers such that ab^2+b+7 divides a^2b+a+b.

Solution. Considering the expressions as polynomials of a and treating b as constant, it is natural to begin as follow. If ab^2+b+7 divides a^2b+a+b, then ab^2+b+7 divides

$$a(ab^2+b+7) - b(a^2b+a+b) = 7a - b^2.$$

If $7a-b^2 = 0$, then 7 divides b and so $b = 7k$, $a = 7k^2$ for some positive integer k. It is easy to check these pairs $(a, b) = (7k^2, 7k)$ satisfy the condition.

If $7a-b^2 < 0$, then $ab^2+b+7 \le |7a-b^2| = -7a+b^2$, but this contradicts $-7a+b^2 < b^2 < ab^2+b+7$.

If $7a-b^2 > 0$, then $ab^2+b+7 \le 7a-b^2$. If $b \ge 3$, then $ab^2+b+7 > 9a > 7a > 7a-b^2$, a contradiction. So b can only be 1 or 2. If $b = 1$, then $a+8$ divides $7a-1 = 7(a+8)-57$. So $a+8$ divides 57, which implies $a = 11$ or 49. If $b = 2$, then $4a+9$ divides $7a-4$. Since $7a-4 < 2(4a+9) = 8a+18$, we get $7a-4 = 4a+9$, which has no integer solution. Finally, $(a, b) = (11, 1)$ and $(49, 1)$ are easily checked to be solutions.

---

(17) (1988 IMO) Let a and b be positive integers such that $ab+1$ divides a^2+b^2. Show that $\dfrac{a^2+b^2}{ab+1}$ is the square of an integer.

Solution. Let $k = (a^2+b^2)/(ab+1)$. Assume there exists a case k is an integer, but not a perfect square. Among all such cases, consider the case when $\max\{a, b\}$ is least possible. Note $a = b$ implies $0 < k = 2a^2/(a^2+1) < 2$ so that $k = 1$. Hence, by symmetry, we may assume $a > b$. Now $x^2+b^2-k(xb+1) = 0$ has a as a root. The other root is the integer $c = kb-a = (b^2-k)/a$. Now $cb+1 = (c^2+b^2)/k > 0$ and $c = (b^2-k)/a \ne 0$ imply c is a positive integer.

Also, $c = (b^2 - k)/a < (a^2 - k)/a < a$. Now $k = (c^2 + b^2)/(cb + 1)$ is a nonsquare integer and $\max\{b, c\} < a = \max\{a, b\}$ contradict $\max\{a, b\}$ is least possible. Therefore, all such k's are perfect squares.

Remarks. Considering to the roots of a quadratic expression is a useful trick in some number theory problems!

---

(18) (2003 IMO) Determine all pairs of positive integers (a, b) such that $\dfrac{a^2}{2ab^2 - b^3 + 1}$ is a positive integer.

Solution. Let $k = a^2/(2ab^2 - b^3 + 1)$ be a positive integer. Then $a^2 - 2kb^2a + kb^3 - k = 0$. (Note it is possible to consider roots. However, the following is a variation that is also useful.) Multiplying by 4 and completing squares, we get $(2a - 2kb^2)^2 = (2kb^2 - b)^2 + (4k - b^2)$. Let $M = 2a - 2kb^2$ and $N = 2kb^2 - b$, then $M^2 = N^2 + (4k - b^2)$.

If $4k - b^2 = 0$, then b is even and $M = \pm N$. If $M = -N$, then we get $b = 2a$. If $M = N$, then $2a = 4kb^2 - b = b^4 - b$. Thus, we get $(a, b) = (b/2, b)$ or $((b^4 - b)/2, b)$ with b an even integer. We can easily check these are solutions.

If $4k - b^2 > 0$, then since $N = 2kb^2 - b = b(2kb - 1) \geq 1(2 - 1) = 1$, so $M^2 \geq (N + 1)^2$. We have

$$4k - b^2 = M^2 - N^2 \geq (N + 1)^2 - N^2 = 2N + 1 = 4kb^2 - 2b + 1,$$

which implies $4k(b^2 - 1) + (b - 1)^2 \leq 0$. Since the left side is also nonnegative, this forces $b = 1$ and $k = a^2/(2a - 1 + 1) = a/2$. Then $(a, b) = (2k, 1)$, which can be checked to be a solution for every positive integer k.

If $4k - b^2 < 0$, then $M^2 \leq (N - 1)^2$. So

$$4k - b^2 = M^2 - N^2 \leq (N - 1)^2 - N^2 = -2N + 1 = -4kb^2 + 2b + 1,$$

which implies $0 \leq (1 - 4k)b^2 + 2b + (1 - 4k)$. However, the right side equals $(1 - 4k)(b + \dfrac{1}{1 - 4k})^2 + \dfrac{8k(2k - 1)}{1 - 4k} < 0$, a contradiction.

---

(19) (1972 Putnam Exam) Show that if n is an integer greater than 1, then n does not divide $2^n - 1$.

Solution. Assume n divides $2^n - 1$ for some integer $n > 1$. Since $2^n - 1$ is odd, so n is odd. Let p be the smallest prime divisor of n. Then p divides $2^n - 1$, hence $2^n \equiv 1 (\text{mod } p)$. By Fermat's little theorem, $2^{p-1} \equiv 1 (\text{mod } p)$. Now the greatest common divisor d of n and $p - 1$ must be 1 because d divides n, $d \leq p - 1$ and p is the smallest prime divisor of n. Then there are integers r, s such that $rn + s(p - 1) = 1$, which implies

$$2 = 2^d = (2^n)^r (2^{p-1})^s \equiv 1 (\text{mod } p),$$

a contradiction.

Exercises

12. (1992 IMO) Find all integers a, b, c with $1 < a < b < c$ such that $(a - 1)(b - 1)(c - 1)$ is a divisor of $abc - 1$.

13. (1994 IMO) Determine all ordered pairs (m, n) of positive integers such that $(n^3 + 1)/(mn - 1)$ is an integer. (*Comments*: There are 9 solutions.)

14. Redo Example 18 by considering roots of quadratic expression as in the solution of Example 17.

15. (1999 IMO) Determine all pairs (n, p) of positive integers such that p is a prime, $n \leq 2p$, and $(p - 1)^n + 1$ is divisible by n^{p-1}. (*Hint*: For $p \geq 3$, consider the smallest prime divisor q of n.)

16. (2000 CHKMO) Find all prime numbers p and q such that $\dfrac{(7^p - 2^p)(7^q - 2^q)}{pq}$ is an integer.

17. (2003 IMO) Let p be a prime number. Prove that there exists a prime number q such that for every integer n, the number $n^p - p$ is not divisible by q. (*Hint*: Note not all of the prime divisors of $M = (p^p - 1)/(p - 1)$ are congruent to 1 (mod p^2). Let q be such a prime divisor of M.)

18. (1990 IMO) Determine all integers $n > 1$ such that $(2^n + 1)/n^2$ is an integer. (*Hint*: Write $n = 3^k r$ with $(3, r) = 1$. Show $k = 0$ or 1 by factoring $2^n + 1$

and considering (mod 9). Show $r = 1$ by considering the smallest prime divisor of r if $r > 1$.)

---

§4 Diophantine Equations–Equations which integral solutions are sought.

Examples. (20) (1979 USAMO) Determine all integral solutions of $n_1^4 + n_2^4 + \cdots + n_{14}^4 = 1599$.

Solution. We have $(2n)^4 = 16n^4 \equiv 0 (\text{mod } 16)$ and

$$(2n+1)^4 = 16n^4 + 32n^3 + 8n(3n+1) + 1 \equiv 1 (\text{mod } 16).$$

So, $n_1^4 + n_2^4 + \cdots + n_{14}^4 \equiv 0, 1, 2, \ldots, 14 (\text{mod } 16)$, but $1599 \equiv 15 (\text{mod } 16)$. So, there can be no solution.

---

(21) (1976 USAMO) Determine all integral solutions of $a^2 + b^2 + c^2 = a^2b^2$.

Solution. Suppose (a, b, c) is a solution. If a, b, c are odd, then $a^2 + b^2 + c^2 \equiv 3 (\text{mod } 4)$, but $a^2b^2 \equiv 1 (\text{mod } 4)$. If two are odd and one even, then $a^2 + b^2 + c^2 \equiv 2 (\text{mod } 4)$, but $a^2b^2 \equiv 0$ or $1 (\text{mod } 4)$. If one is odd and two even, then $a^2 + b^2 + c^2 \equiv 1 (\text{mod } 4)$, but $a^2b^2 \equiv 0 (\text{mod } 4)$.

Therefore, a, b, c must all be even, say $a = 2a_0$, $b = 2b_0$, $c = 2c_0$. Then we get $a_0^2 + b_0^2 + c_0^2 = 4a_0^2b_0^2$. If at least one of a_0, b_0, c_0 is odd, then $a_0^2 + b_0^2 + c_0^2 \equiv 1, 2$ or $3 (\text{mod } 4)$, but $4a_0^2b_0^2 \equiv 0 (\text{mod } 4)$. So a_0, b_0, c_0 must all be even again, say $a_0 = 2a_1$, $b_0 = 2b_1$, $c_0 = 2c_1$. Then $a_1^2 + b_1^2 + c_1^2 = 16a_1^2b_1^2$. So a_1, b_1, c_1 must all be even. From this we see inductively that a, b, c can be divisible by any power of 2. Therefore, $a = b = c = 0$.

---

(22) (1993 APMO) Determine all positive integer n for which the equation $x^n + (2+x)^n + (2-x)^n = 0$ has an integer as a solution.

Solution. If n is even, the terms on the left side are nonnegative and cannot all be 0. So there will not be any integer solution. If $n = 1$, then $x = -4$ is the solution. If n is odd and at least 3, then any solution x must be even, otherwise the left side is odd. Suppose $x = 2y$. Then the equation becomes $y^n + (1+y)^n + (1-y)^n = 0$. Obviously $y \neq 0$. We have $0 = y^n + (1+y)^n + (1-y)^n \equiv 2 (\text{mod } |y|)$. So $y|2$,

forcing $y = \pm 1$ or ± 2. However, simple checkings show these are not solutions. So $n = 1$ is the only solution.

---

(23) Determine all integral solutions of $y^2 = 1 + x + x^2 + x^3 + x^4$.

Solution. Completing squares, we have $(x^2 + \frac{x}{2} + 1)^2 = y^2 + \frac{5}{4}x^2$ and $(x^2 + \frac{x}{2})^2 + \frac{3}{4}(x + \frac{2}{3})^2 + \frac{2}{3} = y^2$. So $x^2 + \frac{x}{2} + 1 \geq |y| > x^2 + \frac{x}{2}$. If x is odd, then $|y| = x^2 + \frac{x}{2} + \frac{1}{2}$. Substituting back in to the equation and simplifying we get $x^2 - 2x - 3 = 0$, yielding $x = -1$ or 3. If x is even, then $|y| = x^2 + \frac{x}{2} + 1$ and so $\frac{5}{4}x^2 = 0$ forcing $x = 0$. Therefore, the solutions are $(x, y) = (-1, \pm 1)$, $(3, \pm 11)$ and $(0, \pm 1)$.

---

(24) Determine all nonzero integral solutions of $(a^2 + b)(a + b^2) = (a - b)^3$.

Solution. Expanding, then simplifying the equation, we get $2b^2 + (a^2 - 3a)b + (3a^2 + a) = 0$. Applying the quadratic formula, we get

$$b = \frac{3a - a^2 \pm \sqrt{a^4 - 6a^3 - 15a^2 - 8a}}{4} = \frac{3a - a^2 \pm \sqrt{a(a-8)(a+1)^2}}{4}.$$

So $a(a-8) = (a-4)^2 - 4^2$ must be a perfect square, say $x^2 = (a-4)^2 - 4^2$. Then $(a-4)^2 - x^2 = 16$. So $(a-4+x)(a-4-x) = (\pm 1)(\pm 16)$ or $(\pm 2)(\pm 8)$ or $(\pm 4)(\pm 4)$. From these, we get the nonzero $a = -1, 8, 9$. These lead to the solutions $(a, b) = (-1, -1)$, $(8, -10)$, $(9, -6)$, $(9, -21)$. (Check: The 4 solutions yield the 4 equations $0 \times 0 = 0$, $54 \times 108 = 18^3$, $75 \times 45 = 15^3$, $60 \times 450 = 30^3$.)

---

(25) Find all positive integral solutions of $3^x + 4^y = 5^z$.

Solution. We will show there is exactly one set of solution, namely $x = y = z = 2$. To simplify the equation, we consider modulo 3. We have $1 = 0 + 1^y \equiv 3^x + 4^y = 5^z \equiv (-1)^z \pmod 3$. It follows that z must be even, say $z = 2w$. Then $3^x = 5^z - 4^y = (5^w + 2^y)(5^w - 2^y)$. Now $5^w + 2^y$ and $5^w - 2^y$ are not both divisible by 3, since their sum is not divisible by 3. So, $5^w + 2^y = 3^x$ and $5^w - 2^y = 1$. Then, $(-1)^w + (-1)^y \equiv 0 (\text{mod } 3)$ and $(-1)^w - (-1)^y \equiv 1 (\text{mod } 3)$. From these, we get w is odd and y is even. If $y > 2$, then $5 \equiv 5^w + 2^y = 3^x \equiv 1$ or $3 \pmod 8$,

a contradiction. So $y = 2$. Then $5^w - 2^y = 1$ implies $w = 1$ and $z = 2$. Finally, we get $x = 2$.

Finally, we come to the most famous Diophantine equation. Let us define *Pythagorean triples* to be triples (a, b, c) of positive integers satisfying $a^2 + b^2 = c^2$. For example, $(3, 4, 5)$ and $(5, 12, 13)$ are Pythagorean triples. Clearly, if $a^2 + b^2 = c^2$, then $(ad)^2 + (bd)^2 = (cd)^2$ for any positive integer d. So, solutions of $a^2 + b^2 = c^2$ with a, b, c relatively prime (i.e. having no common prime divisors) are more important. These are called *primitive solutions.* Below we will establish a famous theorem giving all primitive solutions.

Theorem. If u, v are relatively prime positive integers of opposite parity and $u > v$, then $a = u^2 - v^2, b = 2uv, c = u^2 + v^2$ give a primitive solution of $a^2 + b^2 = c^2$. Conversely, every primitive solution is of this form, with a possible permutation of a and b.

(For example, $u = 2, v = 1$ yields $a = 3, b = 4, c = 5$.)

Reasons. For the first statement, $a^2 + b^2 = u^4 + 2u^2v^2 + v^4 = c^2$. Suppose two of a, b, c have a common prime divisor p, then the equation will imply all three have p as a common divisor. Note $p \neq 2$ since a, c are odd. Then $p|(c + a)/2 = u^2$ and $p|(c - a)/2 = v^2$. This contradicts u, v being relatively prime. So a, b, c must be relatively prime.

For the second statement, if $a^2 + b^2 = c^2$, then $a^2 + b^2 \equiv 0$ or $1 \pmod 4$. So, if a, b, c are also relatively prime, then one of a or b is odd and the other is even. Let us say a is odd and b is even. Then c is odd and it follows $m = (c + a)/2$ and $n = (c - a)/2$ are positive integers. Note $a(= m - n)$ and $c(= m + n)$ relatively prime implies m, n cannot have a common prime divisor (for if $p|m$ and $p|n$, then $p|m - n = a$ and $p|m + n = c$). Now $(b/2)^2 = (c^2 - a^2)/4 = mn$. It follows that both m and n are perfect squares with no common prime divisors. Let us say $m = u^2$ and $n = v^2$. Then $a = u^2 - v^2$, $b = 2uv$ and $c = u^2 + v^2$. Since a is odd, u and v are of opposite parity.

Remark. The general solutions of $a^2 + b^2 = c^2$ are either trivial with a or b equals 0 or nontrivial with a, b of the form $\pm(u^2 - v^2)d, \pm 2uvd$ and c of the form $\pm(u^2 + v^2)d$, where u, v are as above and d is positive.

Example. (26) Find all positive integral solutions of $3^x + 4^y = 5^z$ using the theorem on Pythagorean triples.

Solution. Let x, y, z be a solution, then $1 \equiv (-1)^z \pmod 3$ and $(-1)^x \equiv 1 \pmod 4$. So x and z are even, say $x = 2a$ and $z = 2b$. Then $(3^a)^2 + (2^y)^2 = (5^b)^2$. Since $3^a, 2^y$ and 5^b are relatively prime, by the theorem on Pythagorean triples, $3^a = u^2 - v^2$ and $2^y = 2uv$, where $u > v$ and one is odd, the other even. Now $2^y = 2uv$ implies $u = 2^{y-1} > v = 1$. Then $3^a = 2^{2(y-1)} - 1 = (2^{y-1} - 1)(2^{y-1} + 1)$. Since the two factors on the right differ by 2 and must be powers of 3, we have $2^{y-1} - 1 = 1$, which gives $y = 2, u = 2, a = 1, x = 2$ and $z = 2$.

Exercises

19. Show that $15x^2 - 7y^2 = 9$ has no integral solutions.

20. Find all integral solution(s) of $x^3 + 2y^3 + 4z^3 = 9w^3$. (*Hint:* Consider (mod 9) first.)

21. Find all integral solution(s) of $3 \cdot 2^x + 1 = y^2$.

22. Show that $y^2 = x^3 + 7$ has no integral solutions. (*Hint:* Note $y^2 + 1 = (x + 2)(x^2 - 2x + 4)$ and use example 14.)

*23. Show that $x^4 + y^4 = z^2$ has no positive integral solutions. (*Hint:* Suppose (x, y, z) is a positive solution with z least possible, then show there is another solution with a smaller z value using the theorem on Pythagorean triples.) Then also conclude that $x^4 + y^4 = z^4$ has no positive integral solution.

Remarks. The famous *Fermat's Last Theorem* is the statement that for every integer $n > 2$, the equation $x^n + y^n = z^n$ has no positive integral solution. Fermat claimed to have a proof 350 years ago, but nobody found his proof. Only a few year ago, Andrew Wiles finally proved it. His proof was 200 pages long.

4. Combinatorics

Combinatorics is the study of counting objects. There are many basic, yet useful principles that allow us to count efficiently.

§1. Addition and Multiplication Principles.

Addition Principle. *Suppose events $A_1, A_2, \ldots, A_n$ have $a_1, a_2, \ldots, a_n$ outcomes respectively. If all of these outcomes are distinct, then the number of outcomes due to event A_1 or event A_2 or ... or event A_n is $a_1 + a_2 + \cdots + a_n$.*

Remarks. When you need to count something, the difficulty is how to break up the things into groups that are easy to count.

Examples. (1) Flipping a coin (event A_1) results in two outcomes: head or tail. Tossing a dice (event A_2) results in six outcomes: 1,2,3,4,5,6. So flipping a coin *or* tossing a dice results in $6+2=8$ outcomes.

---

(2) Find the number of squares having all their vertices belonging to an 10×10 array of equally spaced dots.

Solution. Each such square has a unique circumscribed square with sides parallel to the sides of the array! Use these circumscribed squares as *events*. If a circumscribed square is $k \times k$ (that is, each side has $k+1$ dots), then there are k distinct squares (*outcomes*) inscribed in this circumscribed square. For $k = 1, 2, \ldots, 9$, there are $(10-k)^2$ circumscribed squares of dimension $k \times k$. So the answer is

$$\sum_{k=1}^{9}(10-k)^2 \cdot k = \sum_{k=1}^{9}(100k - 20k^2 + k^3) = 100\sum_{k=1}^{9}k - 20\sum_{k=1}^{9}k^2 + \sum_{k=1}^{9}k^3 = 825,$$

using the formulas

$$\sum_{k=1}^{n}k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^{n}k^2 = \frac{n(n+1)(2n+1)}{6} \quad \text{and} \quad \sum_{k=1}^{n}k^3 = \left(\frac{n(n+1)}{2}\right)^2.$$

Multiplication Principle. *Suppose events $A_1, A_2, \ldots, A_n$ have $a_1, a_2, \ldots, a_n$ outcomes respectively. Then the number of outcomes of event A_1, followed by event A_2, ..., followed by event A_n is $a_1 a_2 \cdots a_n$.*

Examples. (3) Let n, k be integers such that $0 \le k \le n$.

(a) Find the number P_k^n (or ${}_nP_k$) of <u>*permutations*</u> of n distinct objects taken k at a time. This is the number of ways of taking k of the n objects one after the other without replacement so ordering is important. (If k is not mentioned, then the default value is $k = n$.)

(b) Find the number C_k^n (or ${}_nC_k$ or $\binom{n}{k}$) of <u>*combinations*</u> of n distinct objects taken k at a time. This is the number of ways of taking k of the n objects at the same time so ordering is not important.

Solutions (a) There are n outcomes of taking the first object, followed by $n-1$ outcomes of taking the second object, ..., there are $n-k+1$ outcomes of taking the k-th objects. By the multiplication principle, $P_k^n = n(n-1)\cdots(n-k+1) = \dfrac{n!}{(n-k)!}$.

(b) Divide all permutations of n objects taken k at a time into groups that have the same k objects. In each group, the number of ways the k objects were taken was $P_k^k = k!$. Since order is not important in combinations, the $k!$ ways in each group is only counted once. Hence C_k^n is the number of groups, hence $C_k^n = \dfrac{P_k^n}{k!} = \dfrac{n!}{k!(n-k)!}$.

---

(4) Find the number of positive divisors of a positive integer N having prime factorization $2^{e_1}3^{e_2}\cdots p_n^{e_n}$. Find the number of ordered pairs (a, b) of positive integers a, b such that $\text{lcm}(a, b) = N$.

Solution. Every divisor of N is of the form $2^{A_1}3^{A_2}\cdots p_n^{A_n}$, where $A_i = 0, 1, 2, \ldots, e_i$ for $i = 1, 2, \ldots, n$. So the number of outcomes in the event of filling in A_1 followed by the event of filling in A_2 ... followed by the event of filling in A_n is $(e_1+1)(e_2+1)\cdots(e_n+1)$. This is the number of positive divisors of N.

Let $a = 2^{m_1}3^{m_2}\cdots p_n^{m_n}$ and $b = 2^{k_1}3^{k_2}\cdots p_n^{k_n}$. Then $\text{lcm}(a, b) = N$ if and only if $\max(m_1, k_1) = e_1$ *and* $\max(m_2, k_2) = e_2$ *and ... and* $\max(m_n, k_n) = e_n$. Consider the event $\max(m_1, k_1) = e_1$, there are $2e_1+1$ possible outcomes, namely $(m_1, k_1) = (0, e_1), (1, e_1), \ldots, (e_1, e_1), (e_1, e_1-1), \ldots, (e_1, 1), (e_1, 0)$. Similarly, there are $2e_i + 1$ outcomes for the event $\max(m_i, k_i) = e_i$. So the number of pairs (a, b) such that $\text{lcm}(a, b) = n$ is $(2e_1+1)(2e_2+1)\cdots(2e_n+1)$.

---

§2. Bijection Principle.

Bijection Principle. *If there is a one-to-one correspondence between the outcomes of event A and outcomes of event B, then A and B have the same number of outcomes.*

Reminder. When we are asked to count the number of outcomes of an event, sometimes it may be possible to set up a one-to-one correspondence with another event whose outcomes are easier to count.

Examples. (5) Let n be a positive integer. In how many ways can one write a sum of at least two positive integers that add up to n? Consider the same set of integers written in a different order as being different. (For example, there are 3 ways to express 3 as $3 = 1+1+1 = 2+1 = 1+2$.)

Solution. (First the answer can be discovered by trying the cases $n = 4, 5$ and observe pattern!) We study the case $n = 3$. We have $3 = (1)+(1)+(1) = (1+1)+(1) = (1)+(1+1)$. Each of the sum is in a *one-to-one correspondence* with $(1\square 1\square 1)$, where each blank square is filled with a $+$ or a $)+($. Note since each sum has at least two terms, there is at least on $)+($.

For a general n, we have $n = (1\square 1\square \cdots \square 1)$ with $n-1$ blank squares. By the multiplication principle, in filling in the blank squares one followed by the other, there are 2^{n-1} ways, but filling all blank squares with $+$ is not allowed. So there are $2^{n-1}-1$ ways.

---

(6) How many paths are there going from $(0, 0)$ to $(10, 20)$ on the coordinate plane such that either the x or the y coordinate of every point on the path is an integer and the x and y coordinates on the path are always nondecreasing at every moment?

Solution. Note such a path is a staircase from $(0, 0)$ to $(10, 20)$ with corners at lattice points (i.e. points with x, y integers). Break the path into unit length pieces. Then each piece is either moving left or up. By projecting the path to the x and y-axes, we see that the length is 30 units and there are 10 left and 20 up pieces. There is a *one-to-one correspondence* between a path and a sequence of 10 lefts and 20 ups in some order. Hence, among the 30 pieces, it depends where we take the 10 lefts. Therefore, there are C_{10}^{30} paths.

---

(7) Each of the vertices of a regular nonagon (i.e. 9-sided polygon) has been colored either red or blue. Prove that there exist two congruent monochromatic (i.e. vertices having the same color) triangles.

Solution. By the pigeonhole principle, there are at least five vertices of the same color, say red. So there are at least $C_3^5 = 10$ triangles having red vertices. For each triangle ABC (vertices in clockwise order), take vertex A, go around the nonagon in the clockwise direction and count the number of sides of the nonagon travelled from A to B, from B to C and from C to A.

Arrange these three numbers in increasing order $x \le y \le z$. There is a *one-to-one correspondence* between the sides of triangle ABC and (x, y, z). Since x, y, z are positive integers, $x \le y \le z$ and $x+y+z = 9$, there are 7 outcomes, namely $(x, y, z) = (1, 1, 7), (1, 2, 6), (1, 3, 5), (1, 4, 4), (2, 2, 5), (2, 3, 4), (3, 3, 3)$. Since $10 > 7$, by the pigeonhole principle, there must be two congruent red triangles.

---

(8) Let m and n be integers greater than 1. Let S be a set with n elements, and let $A_1, A_2, \ldots, A_m$ be subsets of S. Assume that for any two elements x and y in S, there is a set A_i containing either x or y, but not both. Prove that $n \le 2^m$.

Solution. For each element z in S, define $f(z) = (z_1, z_2, \ldots, z_m)$, where $z_i = 1$ if z is in the set A_i, otherwise $z_i = 0$. If $x \ne y$, then $f(x)$ and $f(y)$ differ in some coordinates. So the n elements in S correspond to n different $f(z)$. By the multiplication principle, there are exactly $2 \times 2 \times \cdots \times 2 = 2^m$ possible $(z_1, z_2, \ldots, z_m)$'s. Therefore, $n \le 2^m$.

---

§3 Pigeonhole Principle.

Pigeonhole Principle. *If $n+1$ or more objects are put into n boxes, then at least two of the objects will be in the same box. More generally, if m objects are put into n boxes, then at least $\left\lceil \frac{m}{n} \right\rceil$ objects will be in the same box.*

Examples. (9) (1954 Putnam Exam) Five point are chosen from inside of the unit square. Show that there are two points with distance at most $\frac{1}{2}\sqrt{2}$.

Solution. Divide the inside of the unit square into 4 squares with side $\frac{1}{2}$. By the pigeonhole principle, there are two points in the same square. Then their distance is at most the length of the diagonal, which is $\frac{1}{2}\sqrt{2}$.

---

(10) Eleven numbers are chosen from $1, 2, \ldots, 20$. Show that the sum of two of them is 21.

Solution. Consider the 10 sets $\{1, 20\}, \{2, 19\}, \cdots, \{10, 11\}$. By the pigeonhole principle, two of them will be from the same set. Hence the sum is 21.

---

(11) From any set of m integers, where $m > 1$, show that there must be a subset the sum of whose elements is divisible by m.

Solution. Let $a_1, a_2, \ldots, a_m$ be the integers. Consider the $m+1$ numbers

$$S_0 = 0, \quad S_1 = a_1, \quad S_2 = a_1 + a_2, \quad \ldots, \quad S_m = a_1 + a_2 + \cdots + a_m.$$

By the pigeonhole principle, there are S_i, S_j with $i > j$ such that $S_i \equiv S_j \pmod{m}$. Then $a_{j+1} + \cdots + a_i = S_i - S_j \equiv 0 \pmod{m}$. So $\{a_{j+1}, \ldots, a_i\}$ is a subset the sum of whose elements is divisible by m.

---

(12) Suppose $n+1$ numbers are chosen from $1, 2, \ldots, 2n$. Show that there are two of them such that one divides the other.

Solution. Factor each of the $n+1$ numbers into the form $2^m k$ with k odd. There are n possibilities for k, namely $k = 1, 3, \ldots, 2n-1$. By the pigeonhole principle two of them have the same k factor. Then the one with the smaller exponent m divides the one with the larger exponent m.

---

(13) Each pair of 6 distinct points are joined by a red or blue line segment. Show that there is a red or blue triangle. (Note if 6 is replaced by 5, the problem will not be true as one can color the edges of a pentagon red and the diagonals blue to get a counterexample.)

Solution. Take one of the 6 points, say A. By the pigeonhole principle, among the 5 segments from A, there are 3 of the same color, say AB, AC, AD are red. Either triangle BCD is blue or one of the side, say BC, is red, then triangle ABC is red.

Alternative Formulation. Among any 6 people, either there are 3 who knows each other or there are 3 with no pair knows each other.

Solution. Associate each person a point. Draw a red segment joining the points if the corresponding people knows each other, a blue segment if they don't know each other. Then use the result above.

Remarks. There is a famous theorem known as *Ramsey's Theorem*, which asserts that for any positive integers p and q, there is a smallest positive integer $n = R(p, q)$ such that the following statement is true:

> *If n points are given with no three collinear and all line segments connecting pairs of them are colored red or blue, then either there are p points, all segments connecting them are red or there are q points, all segments connecting them are blue.*

Example (13) is the statement $R(3, 3) = 6$. Very few Ramsey numbers $R(p, q)$ are known. Also, Ramsey's theorem can be extended to more than 2 colors. Briefly, given integers $p_1, \ldots, p_k$, there is a least n such that if all segments connecting n points are colored by k colors $C_1, \ldots, C_k$, then either there are p_1 points with segments all C_1 colored or $\ldots$ or there are p_k points with segments all C_k colored.

---

§4 Principle of Inclusion and Exclusion.

Principle of Inclusion and Exclusion (PIE). *Let $|S|$ denote the number of elements in a set S.*

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|,$$

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

In general,

$$|A_1\cup A_2\cup\cdots\cup A_n| = \sum_{1\le i\le n}|A_i| - \sum_{1\le i<j\le n}|A_i\cap A_j| + \sum_{1\le i<j<k\le n}|A_i\cap A_j\cap A_k| - \cdots.$$

Examples. (14) Find the number of positive integers at most 1000 which are divisible by 10 or 12 or 14.

Solution. Let A_k be the set of positive integers at most 1000 which are divisible by k. Since $A_i\cap A_j = A_{\mathrm{lcm}(i,j)}$, lcm(10, 12) = 60, lcm(10, 14) = 70, lcm(12, 14) = 84 and lcm(10, 12, 14) = 420, so by PIE,

$$\begin{aligned}
&|A_{10}\cup A_{12}\cup A_{14}|\\
=&|A_{10}| + |A_{12}| + |A_{14}| - |A_{10}\cap A_{12}| - |A_{10}\cap A_{14}|\\
&\qquad - |A_{12}\cap A_{14}| + |A_{10}\cap A_{12}\cap A_{14}|\\
=&\left[\frac{1000}{10}\right] + \left[\frac{1000}{12}\right] + \left[\frac{1000}{14}\right] - \left[\frac{1000}{60}\right] - \left[\frac{1000}{70}\right] - \left[\frac{1000}{84}\right] + \left[\frac{1000}{420}\right]\\
=&100 + 83 + 71 - 16 - 14 - 11 + 2 = 215.
\end{aligned}$$

---

(15) (Derangement Problem) How many ways can n letters be put into n envelopes so that no letter goes into the right envelope?

Solution. Totally there are $n!$ ways of putting n letters into n envelopes. We will count the opposite situation, where at least one letter goes into the right envelopes. Let A_i be all possible ways of putting letters into envelopes such that the i-th letter goes to the right envelope. Since the other letters may go randomly into the other $n-1$ envelopes, $|A_i| = (n-1)!$. Similarly, $|A_i\cap A_j| = (n-2)!$ for $i\neq j$ as the other $n-2$ letters may be randomly placed, etc. By PIE,

$$\begin{aligned}
|\bigcup_{i=1}^{n} A_i| &= \sum_{1\le i\le n}|A_i| - \sum_{1\le i<j\le n}|A_i\cap A_j| + \sum_{1\le i<j<k\le n}|A_i\cap A_j\cap A_k| - \cdots\\
&= \binom{n}{1}(n-1)! - \binom{n}{2}(n-2)! + \binom{n}{3}(n-3)! - \cdots\\
&= n!(\frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \cdots \pm \frac{1}{n!}).
\end{aligned}$$

So the number of ways no letters go into the right envelopes is

$$n! - |A_1\cup A_2\cup\cdots\cup A_n| = n!(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots \pm \frac{1}{n!}).$$

(Note the probability that no letters will go to the right envelopes is the above expression divided by $n!$, which is close to $\frac{1}{e} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots$ when n is large.)

---

(16) (Euler ϕ-function) For a positive integer n, show that the number of integers in $\{1, 2, \ldots, n\}$ that are relatively prime to n is $\phi(n) = n\prod_{i=1}^{m}(1 - \frac{1}{p_i})$, where $n = p_1^{e_1}p_2^{e_2}\cdots p_m^{e_m}$ is the prime factorization of n.

Solution. Instead we count the number of integers in $\{1, 2, \ldots, n\}$ that are not relatively prime to n. Then these integers are divisible by at least one of the primes $p_1, p_2, \ldots, p_m$. Let A_i be the number of integers in $\{1, 2\ldots, n\}$ that are divisible by p_i. By PIE,

$$\begin{aligned}
|\bigcup_{i=1}^{m} A_i| &= \sum_{1\le i\le m}|A_i| - \sum_{1\le i<j\le n}|A_i\cap A_j| + \sum_{1\le i<j<k\le m}|A_i\cap A_j\cap A_k| - \cdots\\
&= \sum_{1\le i\le m}\frac{n}{p_i} - \sum_{1\le i<j\le m}\frac{n}{p_ip_j} + \sum_{1\le i<j<k\le n}\frac{n}{p_ip_jp_k} - \cdots.
\end{aligned}$$

Hence

$$\begin{aligned}
\phi(n) &= n - |A_1\cup A_2\cup\cdots\cup A_m|\\
&= n - \sum_{1\le i\le m}\frac{n}{p_i} + \sum_{1\le i<j\le m}\frac{n}{p_ip_j} - \cdots\\
&= n\Big(1 - (\frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_m}) + (\frac{1}{p_1p_2} + \cdots + \frac{1}{p_{m-1}p_m}) + \cdots\Big)\\
&= n(1 - \frac{1}{p_1})(1 - \frac{1}{p_2})\cdots(1 - \frac{1}{p_m})\\
&= n\prod_{i=1}^{m}(1 - \frac{1}{p_i}).
\end{aligned}$$

§5 Recurrence Relations and Generating Functions.

Given a sequence $a_0, a_1, a_2, a_3, \ldots$, a *recurrence relation* is typically a formula for a_n in terms of $a_0, a_1, a_2, \ldots, a_{n-1}$ and n. For example, the famous *Fibonacci sequence* is defined by the initial conditions $F_0 = 1, F_1 = 1$ and the recurrence relation $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$. Using the recurrence relation, we see that the Fibonacci sequence F_n is $1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \ldots$. For a sequence $a_0, a_1, a_2, a_3, \ldots$, the *generating function* of the sequence is $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots$. (Note this series may not converge for all real x.) We can often use recurrence relations and generating functions to help in counting things.

Examples. (17) Let a_n be the number of regions formed on a plane by n lines, no two of which are parallel and no three concurrent. Find a_n.

Solution. Clearly, $a_1 = 2, a_2 = 4, a_3 = 7$ by drawing pictures. To solve the problem by recursion, observe that any two of the lines must intersect. Suppose $n-1$ lines formed a_{n-1} regions. The n-th line will intersect them at $n-1$ points. These $n-1$ points divide the n-th line into n parts and each part cuts one of the a_{n-1} regions into two. So $a_n = a_{n-1} + n$. Then

$$\begin{aligned} a_n - a_1 &= (a_n - a_{n-1}) + (a_{n-1} - a_{n-2}) + \cdots + (a_2 - a_1) \\ &= n + (n-1) + \cdots + 2 = \frac{(n+2)(n-1)}{2}. \end{aligned}$$

Therefore, $a_n = a_1 + \dfrac{(n+2)(n-1)}{2} = \dfrac{n^2+n+2}{2}$.

---

(18) Show that the n-th term of the Fibonacci sequence is given by

$$F_n = \frac{1}{\sqrt{5}}\Big(\big(\frac{1+\sqrt{5}}{2}\big)^{n+1} - \big(\frac{1-\sqrt{5}}{2}\big)^{n+1}\Big). \qquad \text{(Binet's Formula)}$$

Solution. Consider the generating function $f(x) = F_0 + F_1x + F_2x^2 + F_3x^3 + \cdots$ of the Fibonacci sequence. Using $F_0 = F_1$ and $F_n = F_{n-1} + F_{n-2}$, we have

$$\begin{aligned} xf(x) &= F_0x + F_1x^2 + F_2x^3 + F_3x^4 + \cdots \\ x^2f(x) &= \qquad F_0x^2 + F_1x^3 + F_2x^4 + \cdots \\ (x^2+x)f(x) &= F_1x + F_2x^2 + F_3x^3 + F_4x^4 + \cdots = f(x) - 1. \end{aligned}$$

So $f(x) = \dfrac{-1}{x^2+x-1}$. By partial fraction, we get $\dfrac{-1}{x^2+x-1} = \dfrac{A}{x-r} + \dfrac{B}{x-s}$, where $A = -\dfrac{1}{\sqrt{5}}$, $B = \dfrac{1}{\sqrt{5}}$, $r = \dfrac{-1+\sqrt{5}}{2}$, $s = \dfrac{-1-\sqrt{5}}{2}$. Note $|s| > |r| > 1$. So for $|x| < |r|$, we have $|x/r| < 1$, $|x/s| < 1$ and

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{5}}\Big(\frac{1}{r-x} - \frac{1}{s-x}\Big) = \frac{1}{\sqrt{5}}\Big(\frac{1/r}{1-(x/r)} - \frac{1/s}{1-(x/s)}\Big) \\ &= \frac{1}{\sqrt{5}}\Big(\big(\frac{1}{r} + \frac{x}{r^2} + \frac{x^2}{r^3} + \cdots\big) - \big(\frac{1}{s} + \frac{x}{s^2} + \frac{x^2}{s^3} + \cdots\big)\Big) \\ &= \sum_{n=0}^{\infty} \frac{1}{\sqrt{5}}\Big(\frac{1}{r^{n+1}} - \frac{1}{s^{n+1}}\Big)x^n. \end{aligned}$$

Therefore, $F_n = \dfrac{1}{\sqrt{5}}\Big(\dfrac{1}{r^{n+1}} - \dfrac{1}{s^{n+1}}\Big)$. Since $\dfrac{1}{r} = \dfrac{1+\sqrt{5}}{2}$, $\dfrac{1}{s} = \dfrac{1-\sqrt{5}}{2}$, Binet's formula follows.

Using generating functions, we can find formulas for terms of k-th order *linear recurrence relations* (which are of the form $a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_ka_{n-k}$ with constants $c_1, c_2, \ldots, c_k$.)

Theorem. *Let the sequence $a_0, a_1, a_2, \ldots$ satisfy $a_n = c_1a_{n-1} + c_2a_{n-2}$ (c_1, c_2 constants) for $n > 1$ and the characteristic equation $x^2 = c_1x + c_2$ has roots r_1, r_2. Then there are constants b_1, b_2 such that for $n = 0, 1, 2, \ldots$,*

$$a_n = \begin{cases} b_1r_1^n + b_2r_2^n & \text{if } r_1 \neq r_2 \\ b_1r_1^n + b_2nr_2^n & \text{if } r_1 = r_2. \end{cases}$$

Remarks. Similar theorem is true for higher order linear recurrence relation. If $a_0, a_1, a_2, \ldots$ satisfies $a_n = c_1a_{n-1} + c_2a_{n-2} + c_3a_{n-3}$ and $x^3 = c_1x^2 + c^2x + c_3$ has roots r_1, r_2, r_3, then there are constants b_1, b_2, b_3 such that for $n = 0, 1, 2, \ldots$,

$$a_n = \begin{cases} b_1r_1^n + b_2r_2^n + b_3r_3^n & \text{if } r_1, r_2, r_3 \text{ are distinct} \\ b_1r_1^n + b_2nr_1^n + b_3r_3^n & \text{if } r_1 = r_2 \neq r_3 \\ b_1r_1^n + b_2nr_1^n + b_3n^2r_1^n & \text{if } r_1 = r_2 = r_3. \end{cases}$$

(19) (1989 Putnam Exam) Prove that there exists a unique function f defined on $(0, +\infty)$ such that $f(x) > 0$ and $f(f(x)) = 6x - f(x)$.

Solution. For $a > 0$, define $a_0 = a$, $a_n = f(a_{n-1})$ for $n = 1, 2, 3, \ldots$. Then $a_n = f(a_{n-1}) = f(f(a_{n-2})) = 6a_{n-2} - f(a_{n-2}) = -a_{n-1} + 6a_{n-2}$. The characteristic equation $x^2 = -x + 6$ has roots -3 and 2. So $a_n = \alpha(-3)^n + \beta 2^n$. If $\alpha \neq 0$, then $a_n < 0$ when n is large, contradicting $f(x) > 0$. So $\alpha = 0$. Since $a_0 = a$, so $\beta = a$. We get $f(a) = a_1 = 2a$. Simple checking shows that $f(x) = 2x$ satisfies $f(f(x)) = 4x = 6x - f(x)$. Therefore, the unique function is $f(x) = 2x$.

---

(20) (1991 Chinese National Senior High Math Competition) Let a_n be the number of positive integers having digits 1, 3 and 4 only and sum of digits equal n. (Find a recurrence relation for a_n and) show that a_{2n} is a perfect square for $n = 1, 2, 3, \ldots$.

Solution. Let A_n be the set of all such integers. Then $A_1 = \{1\}$, $A_2 = \{11\}$, $A_3 = \{111, 3\}$, $A_4 = \{1111, 13, 31, 4\}, \ldots$. So $a_1 = 1, a_2 = 1, a_3 = 2, a_4 = 4$. Note the number $d_1 d_2 \cdots d_k$ (in digit form) is in A_n if and only if the number $d_1 d_2 \cdots d_{k-1}$ is in $\begin{cases} A_{n-1} & \text{if } d_k = 1 \\ A_{n-3} & \text{if } d_k = 3 \\ A_{n-4} & \text{if } d_k = 4. \end{cases}$ So $a_n = a_{n-1} + a_{n-3} + a_{n-4}$ for $n > 4$. Although we can get a formula for a_n using the characteristic equation method, it turns out that formula will not be to helpful to show a_{2n}'s are perfect squares.

Before proceeding further, we will use the recurrence relation to write out the terms of a_n. We get 1, 1, 2, 4, 6, 9, 15, 25, 40, 64, 104, 169, 273, 441, 714, ... The a_{2n} sequence is 1, 4, 9, 25, 64, 169, 441, ... and seems to be all perfect squares. What can we observe from the a_n sequence? Well, $a_{2n-1} + a_{2n} = a_{2n+1}$ and $a_{2n}a_{2n+2} = a_{2n+1}^2$ seem to be true. If true, these can finish the problem for us because $a_2 = 1^2$ and if $a_{2k} = m^2$, then $a_{2k+2} = a_{2k+1}^2/a_{2k} = (a_{2k+1}/m)^2$ will imply all a_{2n}'s are perfect squares by induction.

Now we can check the relations by mathematical induction. For the first relation, $a_1 + a_2 = 2 = a_3$. Suppose $a_{2n-1} + a_{2n} = a_{2n+1}$. Then $a_{2n+1} + a_{2n+2} = a_{2n-1} + a_{2n} + a_{2n+2} = a_{2n+3}$ by the recurrence relation, which completes the induction for the first relation. For the second relation, $a_2 a_4 = 4 = a_3^2$. Suppose

$a_{2n}a_{2n+2} = a_{2n+1}^2$. Then using the recurrence relation and the first relation, we get

$$\begin{aligned} a_{2n+2}a_{2n+4} &= a_{2n+2}(a_{2n+3} + a_{2n+1} + a_{2n}) = a_{2n+2}a_{2n+3} + a_{2n+2}a_{2n+1} + a_{2n+1}^2 \\ &= a_{2n+2}a_{2n+3} + a_{2n+1}(a_{2n+2} + a_{2n+1}) = a_{2n+2}a_{2n+3} + a_{2n+1}a_{2n+3} \\ &= a_{2n+3}(a_{2n+2} + a_{2n+1}) = a_{2n+3}^2, \end{aligned}$$

which completes the induction for the second relation.

---

(21) Suppose $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ are two different groups of n positive integers such that the numbers $a_i + a_j$ $(1 \leq i < j \leq n)$ are the same as the numbers $b_i + b_j$ $(1 \leq i < j \leq n)$. Show that n is a power of 2.

Solution. Let $f(x) = x^{a_1} + x^{a_2} + \cdots + x^{a_n}$ and $g(x) = x^{b_1} + x^{b_2} + \cdots + x^{b_n}$. Then

$$f^2(x) = \sum_{i=1}^{n} x^{2a_i} + 2\sum_{1\leq i<j\leq n} x^{a_i+a_j} = f(x^2) + 2\sum_{1\leq i<j\leq n} x^{a_i+a_j},$$

$$f^2(x) - f(x^2) = 2\sum_{1\leq i<j\leq n} x^{a_i+a_j} = 2\sum_{1\leq i<j\leq n} x^{b_i+b_j} = g^2(x) - g(x^2).$$

So $g(x^2) - f(x^2) = g^2(x) - f^2(x) = (g(x) + f(x))(g(x) - f(x))$. Since $g(1) = n = f(1)$, so $g(x) - f(x) = (x-1)^k p(x)$ for some $k \geq 1$ and polynomial $p(x)$. Then

$$g(x) + f(x) = \frac{g(x^2) - f(x^2)}{g(x) - f(x)} = \frac{(x^2-1)^k p(x^2)}{(x-1)^k p(x)} = (x+1)^k \frac{p(x^2)}{p(x)}.$$

Setting $x = 1$, we get $2n = f(1) + g(1) = 2^k$. Therefore $n = 2^{k-1}$.

For the next few examples, we will need to multiply power series. Observe that

$$\begin{aligned} &(a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots)(b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots) \\ =&a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \cdots. \end{aligned}$$

For sequences $a_0, a_1, a_2, a_3, \ldots$ and $b_0, b_1, b_2, b_3, \ldots$, we define their *convolution* to be the sequence $c_0, c_1, c_2, c_3, \ldots$, where

$$c_n = a_0b_n + a_1b_{n-1} + \cdots + a_nb_0 = \sum_{k=0}^{n} a_kb_{n-k}$$

is the coefficient sequence of the product of the generating functions for the two sequences. We will denote the convolution by $a_n \star b_n = c_n$.

On the open interval $(-1, 1)$, if we multiply $\dfrac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$ by itself, we get $\dfrac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \cdots$. (This also follows from differentiating the geometric series.) In short, $1 \star 1 = n + 1$. Now

$$\frac{x}{(1-x)^2} = x + 2x^2 + 3x^3 + \cdots.$$

Differentiating both sides, we get $\dfrac{1+x}{(1-x)^3} = 1 + 4x + 9x^2 + 16x^3 + \cdots$.

Examples. (22) For $n \geq 3$, show that

$$\binom{n}{1} - 2^2\binom{n}{2} + 3^2\binom{n}{3} - \cdots + (-1)^{n+1}n^2\binom{n}{n} = 0.$$

Solution. Note the left side is a convoluted expression. Now

$$1 - 4x + 9x^2 - 16x^3 + \cdots = \frac{1-x}{(1+x)^3},$$

$$\binom{n}{n}x + \binom{n}{n-1}x^2 + \cdots + \binom{n}{0}x^{n+1} = \binom{n}{0}x + \binom{n}{1}x^2 + \cdots + \binom{n}{n}x^{n+1}$$
$$= x(1+x)^n.$$

Their product is $x(1-x)(1+x)^{n-3}$, which is of degree $n-1$. So the coefficient of x^n is

$$\binom{n}{1} - 2^2\binom{n}{2} + 3^2\binom{n}{3} - \cdots + (-1)^{n+1}n^2\binom{n}{n} = 0.$$

The binomial theorem asserts that for any real number α,

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2}x^2 + \cdots = \sum_{k=0}^{\infty}\binom{\alpha}{k}x^k,$$

where $\dbinom{\alpha}{k} = \dfrac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}$. If α is a nonnegative integer, this is true for all x. However, if α is not a nonnegative integer (eg. $\alpha = -1$), then it is only true for $|x| < 1$. Considering the cases $\alpha = -n$ and $\alpha = 1/2$, we have

$$\binom{-n}{k} = \frac{-n(-n-1)\cdots(-n-k+1)}{k!}$$
$$= \frac{(-1)^k n(n+1)\cdots(n+k-1)}{k!} = (-1)^k\binom{n+k-1}{k},$$

$$\binom{1/2}{k} = \frac{\frac{1}{2}(\frac{1}{2}-1)\cdots(\frac{1}{2}-k+1)}{k!} = \frac{(-1)^{k-1}1\cdot 3\cdot 5\cdots(2k-3)}{2^k k!}.$$

Example. (23) (Catalan Numbers) Find the number T_n of ways a convex n-sided polygon can be divided into triangles by $n-3$ nonintersecting diagonals. (Here nonintersecting means no intersection inside the polygon.)

Solution. The first few cases are $T_3 = 1$, $T_4 = 2$ and $T_5 = 5$. Consider a $(n+1)$-sided convex polygon $v_1v_2\ldots v_{n+1}$. Fix edge v_nv_{n+1}. For any of these divisions, one of the triangles will be $v_kv_nv_{n+1}$ with $1 \leq k \leq n-1$. For each such k, the polygon $v_1v_2\ldots v_kv_{n+1}$ has $k+1$ sides and the polygon $v_kv_{k+1}\ldots v_n$ has $n-k+1$ sides. These yield $T_{k+1}T_{n-k+1}$ triangulations of $v_1v_2\ldots v_{n+1}$ having triangle $v_kv_nv_{n+1}$. (For the case $k = 1$ or $n-1$, T_2 should be set to 1.) Summing from $k = 1$ to $n-1$, we get $T_{n+1} = T_2T_n + T_3T_{n-1} + \cdots + T_nT_2$, which is a convoluted expression.

Consider the generating function $f(x) = T_2 + T_3x + T_4x^2 + T_5x^3 + \cdots$. We have

$$f^2(x) = T_2^2 + (T_2T_3 + T_3T_2)x + (T_2T_4 + T_3T_3 + T_4T_2)x^2 + \cdots$$
$$= T_3 + T_4x + T_5x^2 + T_6x^3 + \cdots.$$

So $xf^2(x) = T_3x + T_4x^2 + T_5x^3 + T_6x^4 + \cdots = f(x) - 1$. Solving for $f(x)$ with $x \neq 0$, we get

$$f(x) = \frac{1 \pm \sqrt{1-4x}}{2x} = \frac{1}{2x}\Big(1 \pm \sum_{k=0}^{\infty} \binom{1/2}{k}(-4x)^k\Big).$$

Since $f(0) = T_2 = 1$, the plus sign is rejected. Comparing coefficients, we get for $n = 3, 4, 5, \ldots,$

$$T_n = -\frac{1}{2}\binom{1/2}{n-1}(-4)^{n-1} = \frac{(2n-4)!}{(n-1)!(n-2)!} = \frac{C_{n-2}^{2n-4}}{n-1}.$$

Exercises.

1. (1980 USSR Math Olympiad) Let n be an odd integer greater than 1. Show that one of the numbers $2^1 - 1, 2^2 - 1, 2^3 - 1, \ldots, 2^n - 1$ is divisible by n. (*Hint*: None of them is congruent to $-1 (\text{mod } n)$.)

2. Show that there are integers a, b, c, not all zero, with absolute values less than 10^6 such that

$$|a + b\sqrt{2} + c\sqrt{3}| \leq \frac{1 + \sqrt{2} + \sqrt{3}}{1 + 10^6 + 10^{12}}.$$

   (*Hint*: Consider all numbers of the form $r + s\sqrt{2} + t\sqrt{3}$, where $r, s, t \in \{0, 1, 2, \ldots, 10^6 - 1\}$ and see how they are distributed.)

3. (1976 USA Math Olympiad) Each square of a 4×7 board is colored white or black. Prove that with any such coloring, there is always a rectangle whose four corner squares are of the same color. Is this true if the board is 4×6?

4. For $n \geq 3$, how many n digit numbers are there such that each digit is 1, 2 or 3 and the digits contain 1,2,3 each at least once? (*Hint*: Let A_1 be the set of n digit numbers, each of its digits is 2 or 3.)

5. In a group of 100 people, suppose everyone knows at least 51 other people in the group. Show that there are three people in the group who know each other.

6. (1978 Austrian-Polish Math Competition) There are 1978 clubs. Each club has 40 members. It is known that every pair of these clubs has exactly one common member. Show that there is one member who belongs to every club.

7. (1964 IMO) Seventeen people correspond by mail with each other. In their letters only 3 different topics are discussed. Each letter deals with only one of these topics. Show that there are at least three people who write to each other on the same topic. (*Hint*: Use 17 points and 3 colors. One color for each topic.)

8. Show that for integers a, b, c with $c \neq 0, 1, 4, 9, 16, \ldots$, if $(a + b\sqrt{c})^n = p + q\sqrt{c}$, then $(a - b\sqrt{c})^n = p - q\sqrt{c}$ for every positive integer n.

9. Let a_n be the number of ways in which a $2 \times n$ rectangle can be formed out of n 1×2 rectangles. Find a recurrence relation of a_n in terms of a_{n-1} and a_{n-2}, then find a_n in terms of n.

10. Let a_n be the number of strings of n symbols each of which is either 0, 1 or 2 such that no two consecutive 0's occur. Show that $a_n = 2a_{n-1} + 2a_{n-2}$ and find a_n in terms of n. (Note $a_1 = 3$, $a_2 = 8$, $a_3 = 22$.)

11. Show that there are $a_n = 2^{n-1}$ ways of arranging the integers $1, 2, \ldots, n$ in a row such that except for the leftmost, every number differ from some number to its left by $+1$ or -1. (*Hint*: The rightmost integer must be 1 or n.)

12. In example (12), show that $a_k = \begin{cases} F_n^2 & \text{if } k = 2n \text{ is even} \\ F_nF_{n+1} & \text{if } k = 2n+1 \text{ is odd,} \end{cases}$ where F_n is the n-th Fibonacci numbers.

5. Functional Equations

A *functional equation* is an equation whose variables are ranging over functions. Hence, we are seeking all possible functions satisfying the equation. We will let $\mathbb{Z}$ denote the set of all integers, $\mathbb{Z}^+$ or $\mathbb{N}$ denote the positive integers, $\mathbb{N}_0$ denote the nonnegative integers, $\mathbb{Q}$ denotes the rational numbers, $\mathbb{R}$ denotes the real numbers, $\mathbb{R}^+$ denote the positive real numbers and $\mathbb{C}$ denote the complex numbers.

In simple cases, a functional equation can be solved by introducing some substitutions to yield more informations or additional equations.

Example. (1) Find all functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$x^2 f(x) + f(1-x) = 2x - x^4$$

for all $x \in \mathbb{R}$.

Solution. Replacing x by $1-x$, we have $(1-x)^2 f(1-x) + f(x) = 2(1-x) - (1-x)^4$. Since $f(1-x) = 2x - x^4 - x^2 f(x)$ by the given equation, we have $(1-x)^2(2x - x^4 - x^2 f(x)) + f(x) = 2(1-x) - (1-x)^4$. Solving for $f(x)$, we have

$$f(x) = \frac{2(1-x) - (1-x)^4 - (1-x)^2(2x-x^4)}{1 - x^2(1-x)^2} = 1 - x^2.$$

Check: For $f(x) = 1 - x^2$, $x^2 f(x) + f(1-x) = x^2(1-x^2) + (1-(1-x)^2) = 2x - x^4$.

For certain type of functional equations, a standard approach to solving the problem is to determine some special values (such as $f(0)$ or $f(1)$), then inductively determine $f(n)$ for $n \in \mathbb{N}_0$, follow by reciprocal values $f(\frac{1}{n})$ and use density to find all $f(x)$, $x \in \mathbb{R}$. The following are examples of such approach.

Example. (2) Find all functions $f : \mathbb{Q} \to \mathbb{Q}$ such that

$$f(x+y) = f(x) + f(y) \qquad \text{(Cauchy Equation)}$$

for all $x, y \in \mathbb{Q}$.

Solution. Step 1 Taking $x = 0 = y$, we get $f(0) = f(0) + f(0) \Rightarrow f(0) = 0$.

Step 2 We will prove $f(kx) = kf(x)$ for $k \in \mathbb{N}$, $x \in \mathbb{Q}$ by induction. This is true for $k = 1$. Assume this is true for k. Taking $y = kx$, we get

$$f(x + kx) = f(x) + f(kx) = f(x) + kf(x) = (k+1) f(x).$$

Step 3 Taking $y = -x$, we get $0 = f(0) = f(x + (-x)) = f(x) + f(-x) \Rightarrow f(-x) = -f(x)$. So $f(-kx) = -f(kx) = -kf(x)$ for $k \in \mathbb{N}$. Therefore, $f(kx) = kf(x)$ for $k \in \mathbb{Z}$, $x \in \mathbb{Q}$.

Step 4 Taking $x = \frac{1}{k}$, we get $f(1) = f(k\frac{1}{k}) = kf(\frac{1}{k}) \Rightarrow f(\frac{1}{k}) = \frac{1}{k} f(1)$.

Step 5 For $m \in \mathbb{Z}, n \in \mathbb{N}$, $f(\frac{m}{n}) = mf(\frac{1}{n}) = \frac{m}{n} f(1)$. Therefore, $f(x) = cx$ with $c(= f(1)) \in \mathbb{Q}$.

Check: For $f(x) = cx$ with $c \in \mathbb{Q}$, $f(x+y) = c(x+y) = cx + cy = f(x) + f(y)$.

In dealing with functions on $\mathbb{R}$, after finding the function on $\mathbb{Q}$, we can often finish the problem by using the following fact.

Density of Rational Numbers. *For every real number x, there are rational numbers $p_1, p_2, \ldots$ increase to x and there are rational numbers $q_1, q_2, \ldots$ decrease to x. We denote this by $p_n \nearrow x$ and $q_n \searrow x$ as $n \to +\infty$.*

It follows from decimal representation of real numbers. For example, $\pi = 3.14159\ldots$ is the limits of $3, \frac{31}{10}, \frac{314}{100}, \frac{3141}{1000}, \ldots$ and also $4, \frac{32}{10}, \frac{315}{100}, \frac{3142}{1000}, \ldots$.

Example. (3) Find all functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(x+y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}$ and $f(x) \geq 0$ for $x \geq 0$.

Solution. Step 1 By example 2, $f(x) = xf(1)$ for $x \in \mathbb{Q}$.

Step 2 If $x \geq y$, then $x - y \geq 0$. So

$$f(x) = f((x-y)+y) = f(x-y) + f(y) \geq f(y).$$

So, f is increasing.

Step 3 If $x \in \mathbb{R}$, then by the density of rational numbers, there are $p_n, q_n \in \mathbb{Q}$ such that $p_n \leq x \leq q_n$, $p_n \nearrow x$ and $q_n \searrow x$ as $n \to +\infty$. So by steps 1 and 2, $p_n f(1) = f(p_n) \leq f(x) \leq f(q_n) = q_n f(1)$. As $n \to +\infty$, $p_n f(1) \nearrow xf(1)$

and $q_n f(1) \searrow x f(1)$. So $p_n f(1)$ and $q_n f(1)$ will squeeze $f(x)$ to $x f(1)$. We get $f(x) = x f(1)$ for all $x \in \mathbb{R}$. Therefore, $f(x) = cx$ with $c(= f(1)) \geq 0$.

Check: For $f(x) = cx$ with $c \geq 0$, $f(x+y) = c(x+y) = cx+cy = f(x)+f(y)$ and $f(x) = cx \geq 0$ for $x \geq 0$.

Remarks. (1) In example 3, if we replace the condition "$f(x) \geq 0$ for $x \geq 0$" by "f is monotone", then the answer is essentially the same, namely $f(x) = cx$ with $c = f(1)$. Also, if the condition "$f(x) \geq 0$ for $x \geq 0$" is replaced by "f is continuous at 0", then steps 2 and 3 in example 3 are not necessary. We can take rational $p_n \nearrow x$ and take limit of $p_n f(1) = f(p_n) = f(p_n - x) + f(x)$ to get $x f(1) = f(x)$ since $p_n - x \nearrow 0$.

(2) The Cauchy equation $f(x+y) = f(x)+f(y)$ for all $x, y \in \mathbb{R}$ has noncontinuous solutions (in particular, solutions not of the form $f(x) = cx$). This requires the concept of a *Hamel basis* of the vector space $\mathbb{R}$ over $\mathbb{Q}$ from linear algebra.

The following are some useful facts related to the Cauchy equation.

Fact 1. *Let $A = \mathbb{R}, [0,\infty)$ or $(0,\infty)$. If $f : A \to \mathbb{R}$ satisfies $f(x+y) = f(x)+f(y)$ and $f(xy) = f(x)f(y)$ for all $x, y \in A$, then either $f(x) = 0$ for all $x \in A$ or $f(x) = x$ for all $x \in A$.*

Proof. By example 2, we have $f(x) = f(1) = x$ for all $x \in \mathbb{Q}$. If $f(1) = 0$, then $f(x) = f(x \cdot 1) = f(x)f(1) = 0$ for all $x \in A$. Otherwise, we have $f(1) \neq 0$. since $f(1) = f(1)f(1)$, we get $f(1) = 1$. Then $f(x) = x$ for all $x \in A \cap \mathbb{Q}$.

If $y \geq 0$, then $f(y) = f(\sqrt{y})f(\sqrt{y}) = f(\sqrt{y})^2 \geq 0$ and $f(x+y) = f(x)+f(y) \geq f(x)$, which implies f is increasing. Now for any $x \in A \setminus \mathbb{Q}$, by the density of rational numbers, there are $p_n, q_n \in \mathbb{Q}$ such that $p_n < x < q_n$, $p_n \nearrow x$ and $q_n \searrow x$ as $n \to +\infty$. As f is increasing, we have $p_n = f(p_n) \leq f(x) \leq f(q_n) = q_n$. Taking limits, the sandwich theorem gives $f(x) = x$ for all $x \in A$.

Fact 2. *If a function $f : (0,\infty) \to \mathbb{R}$ satisfies $f(xy) = f(x)f(y)$ for all $x, y > 0$ and f is monotone, then either $f(x) = 0$ for all $x > 0$ or there exists c such that $f(x) = x^c$ for all $x > 0$.*

Proof. For $x > 0$, $f(x) = f(\sqrt{x})^2 \geq 0$. Also, $f(1) = f(1)f(1)$ implies $f(1) = 0$ or 1. If $f(1) = 0$, then $f(x) = f(x)f(1) = 0$ for all $x > 0$. If $f(1) = 1$, then

$f(x) > 0$ for all $x > 0$ (since $f(x) = 0$ implies $f(1) = f(x\frac{1}{x}) = f(x)f\big(\frac{1}{x}\big) = 0$, which would lead to a contradiction).

Define $g : \mathbb{R} \to \mathbb{R}$ by $g(w) = \ln f(e^w)$. Then

$$g(x+y) = \ln f(e^{x+y}) = \ln f(e^x)f(e^y) = \ln f(e^x) + \ln f(e^y) = g(x) + g(y).$$

Since f is monotone, it follows that g is also monotone. Then $g(w) = cw$ for all w. Therefore, $f(x) = x^c$ for all $x > 0$.

As an application of these facts, we look at the following example.

Example. (4) (2002 IMO) Find all functions f from the set $\mathbb{R}$ of real numbers to itself such that

$$\big(f(x) + f(z)\big)\big(f(y) + f(t)\big) = f(xy - zt) + f(xt + yz)$$

for all x, y, z, t in $\mathbb{R}$.

Solution. (*Due to Yu Hok Pun, 2002 Hong Kong IMO team member, gold medalist*) Suppose $f(x) = c$ for all x. Then the equation implies $4c^2 = 2c$. So c can only be 0 or $\frac{1}{2}$. Reversing steps, we can also check $f(x) = 0$ for all x or $f(x) = \frac{1}{2}$ for all x are solutions.

Suppose the equation is satisfied by a nonconstant function f. Setting $x = 0$ and $z = 0$, we get $2f(0)\big(f(y) + f(t)\big) = 2f(0)$, which implies $f(0) = 0$ or $f(y) + f(t) = 1$ for all y, t. In the latter case, setting $y = t$, we get the constant function $f(y) = \frac{1}{2}$ for all y. Hence we may assume $f(0) = 0$.

Setting $y = 1, z = 0, t = 0$, we get $f(x)f(1) = f(x)$. Since $f(x)$ is not the zero function, $f(1) = 1$. Setting $z = 0, t = 0$, we get $f(x)f(y) = f(xy)$ for all xy. In particular, $f(w) = f(\sqrt{w})^2 \geq 0$ for $w > 0$.

Setting $x = 0$, $y = 1$ and $t = 1$, we have $2f(1)f(z) = f(-z) + f(z)$, which implies $f(-z) = f(z)$ for all z. So f is even.

Define the function $g : (0,\infty) \to \mathbb{R}$ by $g(w) = f(\sqrt{w}) \geq 0$. Then for all $x, y > 0$,

$$g(xy) = f(\sqrt{xy}) = f(\sqrt{x}\sqrt{y}) = f(\sqrt{x})f(\sqrt{y}) = g(x)g(y).$$

Next f is even implies $g(x^2) = f(x)$ for all x. Setting $z = y$, $t = x$ in the given equation, we get

$$\big(g(x^2) + g(y^2)\big)^2 = g\big((x^2 + y^2)^2\big) = g(x^2 + y^2)^2$$

for all x, y. Taking square roots and letting $a = x^2$, $b = y^2$, we get $g(a + b) = g(a) + g(b)$ for all $a, b > 0$.

By fact 1, we have $g(w) = w$ for all $w > 0$. Since $f(0) = 0$ and f is even, it follows $f(x) = g(x^2) = x^2$ for all x.

Check: If $f(x) = x^2$, then the equation reduces to

$$(x^2 + z^2)(y^2 + t^2) = (xy - zt)^2 + (xt + yz)^2,$$

which is a well known identity and can easily be checked by expansion or seen from $|p|^2|q|^2 = |pq|^2$, where $p = x + iz, q = y + it \in \mathbb{C}$.

The concept of a fixed point is another useful idea in attacking a functional equations. Knowing all the fixed points are important in certain types of functional equations.

Definitions. w is a *fixed point* of a function f if w is in the domain of f and $f(w) = w$. Let $f^{(1)} = f$ and $f^{(n)} = f \circ f^{(n-1)}$ for $n = 2, 3, 4, \ldots$, the function $f^{(n)}$ is called the *n-th iterate of f*.

(5) (1983 IMO) Find all functions $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that $f(xf(y)) = yf(x)$ for all $x, y \in \mathbb{R}^+$ and as $x \to +\infty$, $f(x) \to 0$.

Solution. Step 1 Taking $x = 1 = y$, we get $f(f(1)) = f(1)$. Taking $x = 1$, $y = f(1)$, we get $f(f(f(1))) = f(1)^2$. Then

$$f(1)^2 = f(f(f(1))) = f(f(1)) = f(1) \Rightarrow f(1) = 1$$

since $f(1) \in \mathbb{R}^+$. So 1 is a fixed point of f.

Step 2 Taking $y = x$, we get $f(xf(x)) = xf(x)$. So $w = xf(x)$ is a fixed point of f for every $x \in \mathbb{R}^+$.

Step 3 Suppose f has a fixed point $x > 1$. By step 2, $xf(x) = x^2$ is also a fixed point, $x^2 f(x^2) = x^4$ is also a fixed point, ... So x^{2^n}'s are fixed points. Since

$x > 1$, $x^{2^n} \to +\infty$, but $f(x^{2^n}) = x^{2^n} \to +\infty$, not 0. This contradicts $f(x) \to 0$ as $x \to +\infty$. So f does not have any fixed point $x > 1$.

Step 4 Suppose f has a fixed point $x \in (0, 1)$. Then

$$1 = f(\frac{1}{x}x) = f(\frac{1}{x}f(x)) = xf(\frac{1}{x}) \Rightarrow f(\frac{1}{x}) = \frac{1}{x},$$

i.e. f has a fixed point $\frac{1}{x} > 1$, contradicting step 3. So f does not have any fixed point $x \in (0, 1)$.

Step 5 Steps 1, 3, 4 showed the only fixed point of f is 1. By step 2, we get $xf(x) = 1 \Rightarrow f(x) = \frac{1}{x}$ for all $x \in \mathbb{R}^+$.

Check: For $f(x) = \frac{1}{x}$, $f(xf(y)) = f(\frac{x}{y}) = \frac{y}{x} = yf(x)$. As $x \to +\infty$, $f(x) = \frac{1}{x} \to 0$.

---

(6) (1996 IMO) Find all functions $f : \mathbb{N}_0 \to \mathbb{N}_0$ such that $f(m + f(n)) = f(f(m)) + f(n)$ for all $m, n \in \mathbb{N}_0$.

Solution. Step 1 Taking $m = 0 = n$, we get $f(f(0)) = f(f(0)) + f(0) \Rightarrow f(0) = 0$. Taking $m = 0$, we get $f(f(n)) = f(n)$, i.e. $f(n)$ is a fixed point of f for every $n \in \mathbb{N}_0$. Also the equation becomes $f(m + f(n)) = f(m) + f(n)$.

Step 2 If w is a fixed point of f, then we will show kw is a fixed point of f for all $k \in \mathbb{N}_0$. The cases $k = 0, 1$ are known. Suppose kw is a fixed point, then $f(kw + w) = f(kw + f(w)) = f(kw) + f(w) = kw + w$ and so $(k + 1)w$ is also a fixed point.

Step 3 If 0 is the only fixed point of f, then $f(n) = 0$ for all $n \in \mathbb{N}_0$ by step 1. Obviously, the zero function is a solution.

Otherwise, f has a least fixed point $w > 0$. We will show the only fixed points are kw, $k \in \mathbb{N}_0$. Suppose x is a fixed point. By the division algorithm, $x = kw + r$, where $0 \le r < w$. We have

$$x = f(x) = f(r + kw) = f(r + f(kw)) = f(r) + f(kw) = f(r) + kw.$$

So $f(r) = x - kw = r$. Since w is the least positive fixed point, $r = 0$ and $x = kw$.

Since $f(n)$ is a fixed point for all $n \in \mathbb{N}_0$ by step 1, $f(n) = c_n w$ for some $c_n \in \mathbb{N}_0$. We have $c_0 = 0$.

Step 4 For $n \in \mathbb{N}_0$, by the division algorithm, $n = kw + r$, $0 \le r < w$. We have

$$f(n) = f(r + kw) = f(r + f(kw)) = f(r) + f(kw)$$
$$= c_r w + kw = (c_r + k)w = (c_r + \left[\frac{n}{w}\right])w.$$

Check: For each $w > 0$, let $c_0 = 0$ and let $c_1, \ldots, c_{w-1} \in \mathbb{N}_0$ be arbitrary. The functions $f(n) = (c_r + [\frac{n}{w}])w$, where r is the remainder of n divided by w, (and the zero function) are all the solutions. Write $m = kw + r, n = lw + s$ with $0 \le r, s < w$. Then

$$f(m+f(n)) = f(r+kw+(c_s+l)w) = c_r w+kw+c_s w+lw = f(f(m))+f(n).$$

Let S_n be the set of fixed points of $f^{(n)}$. Observe that if x is a fixed point of $f^{(n)}$, then $f(x)$ is also a fixed point of $f^{(n)}$ because $f^{(n)}(f(x)) = f^{(n+1)}(x) = f(f^{(n)}(x)) = f(x)$. So f takes S_n to itself. Also f is injective on S_n because if $f(a) = f(b)$ for $a, b \in S_n$, then $a = f^{(n)}(a) = f^{(n-1)}(f(a)) = f^{(n-1)}(f(b)) = f^{(n)}(b) = b$. This means that if S_n is a finite set, then f is a permutation of S_n.

Since $g(x) = x$ implies $g^{(2)}(x) = g(g((x)) = g(x) = x$, so the fixed points of g are also fixed points of $g^{(2)}$. Letting $g = f, f^{(2)}, f^{(4)}, f^{(8)}, \ldots$, respectively, we get $S_1 \subseteq S_2 \subseteq S_4 \subseteq S_8 \subseteq \cdots$.

Example. (7) Find all functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(f(x)) = x^2 - 2$ for all $x \in \mathbb{R}$.

Solution. Assume such f exists. It turns out S_2 and S_4 are useful for this problem. The fixed points of $f^{(2)}$ are the roots of $x = x^2 - 2$, i.e. $S_2 = \{-1, 2\}$. The fixed points of $f^{(4)}$ are the roots of $x = x^4 - 4x^2 + 2$. i.e. $S_4 = \{-1, 2, \frac{-1 \pm \sqrt{5}}{2}\}$. Let $c = \frac{-1+\sqrt{5}}{2}, d = \frac{-1-\sqrt{5}}{2}$. Since f permutes S_2 and $c, d \in S_4 \setminus S_2$, $f(c) = c$ or d. If $f(c) = c$, then $f^{(2)}(c) = c$ implies c is a fixed point of $f^{(2)}$, which is not

true. So $f(c) = d$ and hence $f(d) = c$. Then $c = f(d) = f(f(c)) = f^{(2)}(c)$, again a contradiction. So no such f can exist.

The above examples showed traditional or systematical ways of solving functional equations. The following examples show some other approaches to deal with these equations.

Example. (8) Find all functions $f : \mathbb{N} \to \mathbb{N}$ such that $f(f(m) + f(n)) = m + n$ for all $m, n \in \mathbb{N}$.

Solution. Clearly, the identity function $f(x) = x$ is a solution. We will show that is the only solution.

To show $f(1) = 1$, suppose $f(1) = t > 1$. Let $s = f(t-1) > 0$. Observe that if $f(m) = n$, then $f(2n) = f(f(m) + f(m)) = 2m$. So $f(2t) = 2$ and $f(2s) = 2t - 2$. Then $2s + 2t = f(f(2s) + f(2t)) = f(2t) = 2 \Rightarrow t < 1$, a contradiction. Therefore, $f(1) = 1$.

Inductively, suppose $f(n) = n$. Then $f(n+1) = f(f(n) + f(1)) = n + 1$. Therefore, $f(n) = n$ for all $n \in \mathbb{N}$ by mathematical induction.

---

(9) (1987 IMO) Prove that there is no function $f : \mathbb{N}_0 \to \mathbb{N}_0$ such that $f(f(n)) = n + 1987$.

Solution. Suppose there is such a function f. Then f is injective, i.e. $f(a) = f(b) \Rightarrow a + 1987 = f(f(a)) = f(f(b)) = b + 1987 \Rightarrow a = b$.

Suppose $f(n)$ misses exactly k distinct values $c_1, \ldots, c_k$ in $\mathbb{N}_0$, i.e. $f(n) \neq c_1, \ldots, c_k$ for all $n \in \mathbb{N}_0$. Then $f(f(n))$ misses the $2k$ distinct values $c_1, \ldots, c_k$ and $f(c_1), \ldots, f(c_k)$ in $\mathbb{N}_0$. (The $f(c_j)$'s are distinct because f is injective.) Now if $w \neq c_1, \ldots, c_k, f(c_1), \ldots, f(c_k)$, then there is $m \in \mathbb{N}_0$ such that $f(m) = w$. Since $w \neq f(c_j)$, $m \neq c_j$, so there is $n \in \mathbb{N}_0$ such that $f(n) = m$, then $f(f(n)) = w$. This shows $f(f(n))$ misses only the $2k$ values $c_1, \ldots, c_k, f(c_1), \ldots, f(c_k)$ and no others. Since $n + 1987$ misses the 1987 values $0, 1, \ldots, 1986$ and $2k \neq 1987$, this is a contradiction.

---

(10) (1999 IMO) Determine all functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$f(x - f(y)) = f(f(y)) + xf(y) + f(x) - 1$$

for all $x, y \in \mathbb{R}$.

Solution. Let $c = f(0)$. Setting $x = y = 0$, we get $f(-c) = f(c) + c - 1$. So $c \neq 0$. Let A be the range of f, then for $x = f(y) \in A$, we have $c = f(0) = f(x) + x^2 + f(x) - 1$. Solving for $f(x)$, this gives $f(x) = \dfrac{c+1}{2} - \dfrac{x^2}{2}$.

Next, if we set $y = 0$, we get

$$\{f(x-c) - f(x) : x \in \mathbb{R}\} = \{cx + f(c) - 1 : x \in \mathbb{R}\} = \mathbb{R}$$

because $c \neq 0$. This means $A - A = \{y_1 - y_2 : y_1, y_2 \in A\} = \mathbb{R}$.

Now for an arbitrary $x \in \mathbb{R}$, let $y_1, y_2 \in A$ be such that $y_1 - y_2 = x$. Then

$$\begin{aligned} f(x) &= f(y_1 - y_2) = f(y_2) + y_1 y_2 + f(y_1) - 1 \\ &= \frac{c+1}{2} - \frac{y_2^2}{2} + y_1 y_2 + \frac{c+1}{2} - \frac{y_1^2}{2} - 1 \\ &= c - \frac{(y_1 - y_2)^2}{2} = c - \frac{x^2}{2}. \end{aligned}$$

Since for $x \in A$, $f(x) = \dfrac{c+1}{2} - \dfrac{x^2}{2}$, so $c = 1$. Hence, $f(x) = 1 - \dfrac{x^2}{2}$ for all x.

Check: For $f(x) = 1 - \dfrac{x^2}{2}$, both sides equal $\dfrac{1}{2} + \dfrac{y^2}{2} - \dfrac{y^4}{8} + x - \dfrac{xy^2}{2} - \dfrac{x^2}{2}$.

Exercises

1. Find all functions $f : \mathbb{N}_0 \to \mathbb{Q}$ such that $f(1) \neq 0$ and

$$f(x + y^2) = f(x) + 2(f(y))^2 \quad \text{for all } x, y \in \mathbb{N}_0.$$

2. Find all functions $f : \mathbb{Q} \to \mathbb{R}$ such that $f(1) = 2$ and

$$f(xy) = f(x)\,f(y) - f(x + y) + 1 \quad \text{for all } x, y \in \mathbb{Q}.$$

3. Find all functions $f : \mathbb{Q} \to \mathbb{R}$ such that

$$f(x)\,f(y) = f(x + y) \quad \text{for all } x, y \in \mathbb{Q}.$$

4. Find all functions $f : \mathbb{R} \to \mathbb{R}$ such that
   (a) $f(x + y) = f(x) + f(y) + 2xy$ for all $x, y \in \mathbb{R}$ and
   (b) $\lim\limits_{x \to 0} \dfrac{f(x)}{x} = 1$.
   (*Hint*: For $n \in \mathbb{N}$, consider $y = x,\ y = 2x,\ \ldots,\ y = (n-1)x$.)

5. (1986 IMO) Find all functions $f : [0, \infty) \to [0, \infty)$ such that
   (a) $f(x f(y))\, f(y) = f(x + y)$ for $x, y \geq 0$ and
   (b) $f(2) = 0$ and $f(x) \neq 0$ for $0 \leq x < 2$.

6. Suppose $f : \mathbb{R} \to \mathbb{R}^+$ is such that

$$f(\sqrt{x^2 + y^2}) = f(x)\,f(y) \quad \text{for every } x, y \in \mathbb{R}.$$

   Find $f(x)$ for $x \in \mathbb{Q}$ in terms of $f(1)$.

*7. (1990 IMO) Let $\mathbb{Q}^+$ be the set of positive rational numbers. Construct a function $f : \mathbb{Q}^+ \to \mathbb{Q}^+$ such that

$$f(x f(y)) = \frac{f(x)}{y} \quad \text{for all } x, y \in \mathbb{Q}^+.$$

*8. (1994 IMO) Let S be the set of real numbers greater than -1. Find all functions $f : S \to S$ such that
   (a) $f(x + f(y) + x f(y)) = y + f(x) + y f(x)$ for all $x, y \in S$ and
   (b) $\frac{f(x)}{x}$ is strictly increasing for $-1 < x < 0$ and for $0 < x$.
   (*Hint*: Show f can only have a fixed point at 0.)

*9. (1992 IMO) Find all functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$f(x^2 + f(y)) = y + (f(x))^2 \quad \text{for all } x, y \in \mathbb{R}.$$

   (*Hint*: Assume $f(0) = 0$, then show $x > 0 \Rightarrow f(x) > 0$, and f is increasing.)

*10. Find all functions $f : \mathbb{Q} \to \mathbb{Q}$ such that $f(2) = 2$ and

$$f\left(\frac{x+y}{x-y}\right) = \frac{f(x) + f(y)}{f(x) - f(y)} \quad \text{for } x \neq y.$$

   (*Hint*: Try $y = cx$ for different $c \in \mathbb{Q}$ and $y = x - 2$.)

6. Inequalities (Part II)

For inequalitiies involving functions, the shapes of the graphs of the functions on intervals are very important.

Definitions. A continuous function f is *convex on an interval I* if

$$f\big(\frac{x_1+x_2}{2}\big) \le \frac{f(x_1)+f(x_2)}{2} \quad \text{for every } x_1, x_2 \in I.$$

Also, f is *strictly convex on I* if f is convex on I and equality holds only when $x_1 = x_2$ above.

A function f is *concave on an interval I* if the function $-f$ is convex on I. (In that case, $f\big(\frac{x_1+x_2}{2}\big) \ge \frac{f(x_1)+f(x_2)}{2}$ for every $x_1, x_2 \in I$.) Similiarly, f is *strictly convex on I* if $-f$ is strictly convex on I.

Second Derivative Test. *If $f''(x) \ge 0$ on the open interval $I = (a, b)$, then f is convex on I. If $f''(x) > 0$ on I, then f is strictly convex on I. The statements for concave and strictly concave functions are similar by reversing the inequality signs.*

For functions defined on intervals containing endpoints, continuity at the endpoints and nonnegative second derivatives inside the interval are sufficient for the functions to be convex on the intervals. Similar statements for strictly convex functions on such intervals are true.

Using the second derivative test, we can check that the following are examples of strictly convex functions on intervals:

$$x^p \text{ on } [0,\infty) \text{ for } p > 1, \qquad x^p \text{ on } (0,\infty) \text{ for } p < 0,$$

$$a^x \text{ on } (-\infty,\infty) \text{ for } a > 1 \qquad \tan x \text{ on } [0, \frac{\pi}{2}).$$

The following are examples of strictly concave functions on intervals:

$$x^p \text{ on } [0,\infty) \text{ for } 0 < p < 1, \qquad \log_a x \text{ on } (0,\infty) \text{ for } a > 1,$$

$$\cos x \text{ on } [-\pi/2, \pi/2], \qquad \sin x \text{ on } [0, \pi].$$

The most important inequality concerning these functions is the following.

Jensen's Inequality. *If f is convex on I and $x_1, x_2, \ldots, x_n \in I$, then*

$$f\big(\frac{x_1+x_2+\cdots+x_n}{n}\big) \le \frac{f(x_1)+f(x_2)+\cdots+f(x_n)}{n}.$$

For strictly convex functions, equality holds if and only if $x_1 = x_2 = \cdots = x_n$.

Generalized Jensen's Inequality. *If f is convex and continuous on I, $x_1, \ldots, x_n \in I$ and $0 < t_1, t_2, \ldots, t_n < 1$ with $t_1 + t_2 + \cdots + t_n = 1$, then*

$$f(t_1x_1 + t_2x_2 + \cdots + t_nx_n) \le t_1 f(x_1) + t_2 f(x_2) + \cdots + t_n f(x_n)$$

(with the same equality condition for strictly convex functions). For concave functions, all inequality signs reverse.

Examples. (1) For a triangle ABC, show that $\sin A + \sin B + \sin C \le \frac{3\sqrt{3}}{2}$ and determine when equality holds.

Solution. Since $f(x) = \sin x$ is strictly concave on $[0, \pi]$, so

$$\begin{aligned}\sin A + \sin B + \sin C &= f(A) + f(B) + f(C)\\ &\le 3f\big(\frac{A+B+C}{3}\big) = 3\sin\big(\frac{A+B+C}{3}\big) = \frac{3\sqrt{3}}{2}.\end{aligned}$$

Equality holds if and only if $A = B = C = \pi/3$, i.e. $\triangle ABC$ is equilateral.

---

(2) If $a, b, c > 0$ and $a + b + c = 1$, then find the minimum of

$$(a+\frac{1}{a})^{10} + (b+\frac{1}{b})^{10} + (c+\frac{1}{c})^{10}.$$

Solution. Note $0 < a, b, c < 1$. Let $f(x) = (x+\frac{1}{x})^{10}$ on $I = (0, 1)$, then f is strictly convex on I because

$$f''(x) = 90(x+\frac{1}{x})^8(1-\frac{1}{x^2})^2 + 10(x+\frac{1}{x})^9(\frac{2}{x^3}) > 0 \quad \text{for } x \in I.$$

By Jensen's inequality,

$$\frac{10^{10}}{3^9} = 3f(\frac{1}{3}) = 3f\big(\frac{a+b+c}{3}\big)$$
$$\leq f(a)+f(b)+f(c) = (a+\frac{1}{a})^{10}+(b+\frac{1}{b})^{10}+(c+\frac{1}{c})^{10}.$$

Therefore, the minimum is $\frac{10^{10}}{3^9}$, attained when $a=b=c=\frac{1}{3}$.

---

(3) Prove the AM-GM inequality.

Solution. If $a_1, a_2, \ldots, a_n > 0$, then since $f(x) = \log x$ is strictly concave on $(0,\infty)$, by Jensen's inequality,

$$\log\big(\frac{a_1+a_2+\cdots+a_n}{n}\big) \geq \frac{\log a_1+\log a_2+\cdots+\log a_n}{n} = \log(\sqrt[n]{a_1a_2\cdots a_n}).$$

Exponentiating both sides, we get the AM-GM inequality.

Remarks. If we use the generalized Jensen's inequality instead, we can get the weighted AM-GM inequality, which states that if $a_1, \cdots, a_n > 0$ and $0 < t_1, \cdots, t_n < 1$ satisfying $t_1+\cdots+t_n = 1$, then $t_1a_1+\cdots+t_na_n \geq a_1^{t_1}\cdots a_n^{t_n}$ with equality if and only if $a_1 = \cdots = a_n$.

---

(4) Prove Hölder's inequality, which states that if $p, q > 1$, $\frac{1}{p}+\frac{1}{q} = 1$ and $a_1, \ldots, a_n, b_1, \ldots, b_n$ are real numbers, then

$$\Big|\sum_{i=1}^n a_ib_i\Big| \leq \big(\sum_{i+1}^n |a_i|^p\big)^{1/p}\big(\sum_{i=1}^n |b_i|^q\big)^{1/q}.$$

(Note the case $p=q=2$ is the Cauchy-Schwarz inequality.)

Solution. Let $A = \sum_{i=1}^n |a_i|^p$ and $B = \sum_{i=1}^n |b_i|^q$. If A or B is 0, then either all a_i's or all b_i's are 0, which will make both sides of the inequality 0. So we need only

consider the case $A \neq 0$ and $B \neq 0$. Let $t_1 = \frac{1}{p}$ and $t_2 = \frac{1}{q}$, then $0 < t_1, t_2 < 1$ and $t_1+t_2 = 1$. Let $x_i = \frac{|a_i|^p}{A}$ and $y_i = \frac{|b_i|^q}{B}$, then $\sum_{i=1}^n x_i = 1$ and $\sum_{i=1}^n y_i = 1$. Since $f(x) = e^x$ is strictly convex on $(-\infty,\infty)$, by the generalized Jensen's inequality,

$$x_i^{1/p}y_i^{1/q} = f(t_1\ln x_i + t_2\ln y_i) \leq t_1 f(\ln x_i) + t_2 f(\ln y_i) = \frac{x_i}{p}+\frac{y_i}{q}.$$

Adding these for $i = 1, \ldots, n$, we get

$$\sum_{i=1}^n \frac{|a_i||b_i|}{A^{1/p}B^{1/q}} = \sum_{i=1}^n x_i^{1/p}y_i^{1/q} \leq \frac{1}{p}\sum_{i=1}^n x_i + \frac{1}{q}\sum_{i=1}^n y_i = 1.$$

Therefore, $\Big|\sum_{i=1}^n a_ib_i\Big| \leq \sum_{i=1}^n |a_i||b_i| \leq A^{1/p}B^{1/q} = \big(\sum_{i=1}^n |a_i|^p\big)^{1/p}\big(\sum_{i=1}^n |b_i|^q\big)^{1/q}$.

Next we will introduce a generalization of Jensen's inequality.

Definition. If $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ satisfy the conditions

$$x_1 \geq x_2 \geq \cdots \geq x_n, \quad y_1 \geq y_2 \geq \cdots \geq y_n,$$

$$x_1 \geq y_1, \quad x_1+x_2 \geq y_1+y_2, \quad \ldots, \quad x_1+\cdots+x_{n-1} \geq y_1+\cdots+y_{n-1}$$

and

$$x_1+\cdots+x_n = y_1+\cdots+y_n,$$

then we say $(x_1, x_2, \ldots, x_n)$ *majorizes* $(y_1, y_2, \ldots, y_n)$ and write

$$(x_1, x_2, \ldots, x_n) \succ (y_1, y_2, \ldots, y_n).$$

Majorization Inequality. *If the function f is convex on the interval $I = [a,b]$ and $(x_1, x_2, \ldots, x_n) \succ (y_1, y_2, \ldots, y_n)$ for $x_i, y_i \in I$, then*

$$f(x_1)+f(x_2)+\cdots+f(x_n) \geq f(y_1)+f(y_2)+\cdots+f(y_n).$$

For strictly convex functions, equality holds if and only if $x_i = y_i$ for $i = 1, 2, \ldots, n$. The statements for concave functions can be obtained by reversing inequality signs.

Examples. (5) For an acute triangle ABC, show that

$$1 \le \cos A + \cos B + \cos C \le \frac{3}{2}$$

and determine when equality holds.

Solution. Without loss of generality, assume $A \ge B \ge C$. Then $A \ge \pi/3$ and $C \le \pi/3$. Since

$$\frac{\pi}{2} \ge A \ge \frac{\pi}{3}, \quad \pi \ge A + B (= \pi - C) \ge \frac{2\pi}{3},$$

we have $(\pi/2, \pi/2, 0) \succ (A, B, C) \succ (\pi/3, \pi/3, \pi/3)$. Since $f(x) = \cos x$ is strictly concave on $I = [0, \pi/2]$, by the majorization inequality,

$$\begin{aligned} 1 &= f(\frac{\pi}{2}) + f(\frac{\pi}{2}) + f(0) \\ &\le f(A) + f(B) + f(C) = \cos A + \cos B + \cos C \\ &\le f(\frac{\pi}{3}) + f(\frac{\pi}{3}) + f(\frac{\pi}{3}) = \frac{3}{2}. \end{aligned}$$

For the first inequality, equality cannot hold (as two of the angles cannot both be right angles). For the second inequality, equality holds if and only if the triangle is equilateral.

---

(6) If $x_1 \ge x_2 \ge \cdots \ge x_n$, then $(x_1, x_2, \ldots, x_n) \succ (x, x, \ldots, x)$, where x is the arithmetic mean of $x_1, x_2, \ldots, x_n$. (Applying this to the majorization inequality, we get Jensen's inequality.)

Solution. For $k = 1, 2, \ldots, n-1$, we have to show $x_1 + \cdots + x_k \ge kx$. Since

$$(n-k)(x_1 + \cdots + x_k) \ge (n-k)kx_k \ge k(n-k)x_{k+1} \ge k(x_{k+1} + \cdots x_n),$$

so $(n-k)(x_1 + \cdots + x_k) \ge k(x_{k+1} + \cdots + x_n)$. Adding $k(x_1 + \cdots + x_k)$ to both sides, we get $n(x_1 + \cdots + x_k) \ge k(x_1 + \cdots + x_n) = knx$. Therefore, $x_1 + \cdots + x_k \ge kx$.

---

(7) Find the maximum of $a^{12} + b^{12} + c^{12}$ if $-1 \le a, b, c \le 1$ and $a + b + c = -\frac{1}{2}$.

Solution. Note the continuous function $f(x) = x^{12}$ is convex on $[-1, 1]$ because $f''(x) = 132x^{10} \ge 0$ on $(-1, 1)$. If $1 \ge a \ge b \ge c \ge -1$ and $a + b + c = -\frac{1}{2}$, then we claim that $(1, -\frac{1}{2}, -1) \succ (a, b, c)$. This is because

$$1 \ge a \quad \text{and} \quad \frac{1}{2} = 1 - \frac{1}{2} \ge -c - \frac{1}{2} = a + b.$$

So by the majorization inequality,

$$a^{12} + b^{12} + c^{12} = f(a) + f(a) + f(c) \le f(1) + f(-\frac{1}{2}) + f(-1) = 2 + \frac{1}{2^{12}}.$$

The maximum value $2 + \frac{1}{2^{12}}$ is attained when $a = 1$, $b = -\frac{1}{2}$ and $c = -1$.

In some problems, the functions we need to consider may not be convex or concave on the entire intervals! Here is an example.

(8) (1999 IMO) Let n be a fixed integer, with $n \ge 2$.

(a) Determine the least constant C such that the inequality

$$\sum_{1 \le i < j \le n} x_i x_j (x_i^2 + x_j^2) \le C (\sum_{1 \le i \le n} x_i)^4$$

holds for all real numbers $x_1, x_2, \ldots, x_n \ge 0$.

(b) For this constant C, determine when equality holds.

Solution. Consider the case $n = 2$ first. Let $x_1 = m + h$ and $x_2 = m - h$, (i.e. $m = (x_1 + x_2)/2$ and $h = (x_1 - x_2)/2$), then

$$x_1 x_2 (x_1^2 + x_2^2) = 2(m^4 - h^4) \le 2m^4 = \frac{1}{8}(x_1 + x_2)^4$$

with equality if and only if $h = 0$, i.e. $x_1 = x_2$.

For the case $n > 2$, let $a_i = x_i/(x_1 + \cdots + x_n)$ for $i = 1, \ldots, n$, then $a_1 + \cdots + a_n = 1$. So $a_i \in [0, 1]$. In terms of a_i's, the inequality to be proved becomes

$$\sum_{1 \le i < j \le n} a_i a_j (a_i^2 + a_j^2) \le C.$$

The left side can be expanded and regrouped to give

$$\sum_{i=1}^{n} a_i^3 (a_1 + \cdots + a_{i-1} + a_{i+1} + \cdots + a_n) = \sum_{i=1}^{n} a_i^3 (1 - a_i).$$

Unfortunately, $f(x) = x^3(1-x) = x^3 - x^4$ is strictly convex only on $[0, \frac{1}{2}]$ as

$$f''(x) = 6x - 12x^2 = 6x(1-2x) > 0 \text{ on } (0, \frac{1}{2}).$$

Since the inequality is symmetric in the a_i's, we may assume $a_1 \ge a_2 \ge \cdots \ge a_n$.

If $a_1 \le \frac{1}{2}$, then since $(\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0) \succ (a_1, a_2, \ldots, a_n)$, by the majorization inequality,

$$f(a_1) + f(a_2) + \cdots + f(a_n) \le f(\frac{1}{2}) + f(\frac{1}{2}) + f(0) + \cdots + f(0) = \frac{1}{8}.$$

If $a_1 > \frac{1}{2}$, then $1 - a_1, a_2, \ldots, a_n \in [0, \frac{1}{2})$. Since $(1 - a_1, 0, \ldots, 0) \succ (a_2, \ldots, a_n)$, by the majorization inequality and case $n = 2$, we have

$$\begin{aligned} f(a_1) + f(a_2) + \cdots + f(a_n) &\le f(a_1) + f(1 - a_1) + f(0) + \cdots + f(0) \\ &= f(a_1) + f(1 - a_1) \le \frac{1}{8}. \end{aligned}$$

Equality holds if and only if two of the variables are equal and the other $n - 2$ variables all equal 0.

Exercises

1. Use the definition of convex function to give a proof of Jensen's inequality for the case $n = 4$. Then use the case $n = 4$ to prove the case $n = 3$.

2. Use Jensen's inequality to prove $a^a b^b c^c \ge (abc)^{(a+b+c)/3}$ for all $a, b, c > 0$.

3. Let $x_1, \ldots, x_n \in [0, 1]$ and $a_1, \ldots, a_n > 0$ be such that $a_1 + \cdots + a_n = 1$. Prove that
$$\sum_{i=1}^{n} \frac{a_i}{1 + x_i} \le \frac{1}{1 + x_1^{a_1} \cdots x_n^{a_n}}$$
and determine when equality holds.

4. Use Hölder's inequality to prove that if $a, b, c, d > 0$ and $c^2 + d^2 = (a^2 + b^2)^3$, then $\frac{a^3}{c} + \frac{b^3}{d} \ge 1$.

*5. Let P be a point inside triangle ABC such that $\angle PAB = \angle PBC = \angle PCA = \alpha$. Show that $\alpha \le 30^\circ$. (*Hint*: Apply sine law to three triangles.)

6. Suppose acute triangle ABC has two angles less than or equal to 60°. Show that
$$\sin \frac{A}{2} \sin \frac{B}{2} \sin \frac{C}{2} \ge \sin \frac{\pi}{4} \sin \frac{\pi}{6} \sin \frac{\pi}{12}.$$

7. Redo example (8) using AM-GM inequality.

8. Let $x, y, z > 1$, $xyz = 4096$ and $\max(x, y, z) \le 32$. Find the maximum and minimum of $x + y + z$.

9. Prove that if $a, b \ge 0$, then
$$\sqrt[3]{a + \sqrt[3]{a}} + \sqrt[3]{b + \sqrt[3]{b}} \le \sqrt[3]{a + \sqrt[3]{b}} + \sqrt[3]{b + \sqrt[3]{a}}.$$

10. (1998 Balkan Math Olympiad, case $n = 5$) Let $0 \le a_1 \le a_2 \le a_3 \le a_4 \le a_5$. Prove that
$$\sqrt[5]{a_1} - \sqrt[5]{a_2} + \sqrt[5]{a_3} - \sqrt[5]{a_4} + \sqrt[5]{a_5} \le \sqrt[5]{a_1 - a_2 + a_3 - a_4 + a_5}.$$

7. Mathematical Games (Part I)

An *invariant* is a quantity that does not change. A *monovariant* is a quantity that keeps on increasing or keeps on decreasing. In some mathematical games, winning often comes from understanding the invariants or the monovariants that are controlling the games.

Examples. (1) (1974 Kiev Math Olympiad) Numbers 1, 2, 3, . . . , 1974 are written on a board. You are allowed to replace any two of these numbers by one number, which is either the sum or the difference of these numbers. Show that after 1973 times of performing this operation, the only number left on the board cannot be 0.

Solution. There are 987 odd numbers on the board in the beginning. Every time the operation is performed, the number of odd numbers left either stays the same (when the numbers taken out are not both odd) or decreases by two (when the numbers taken out are both odd). So the number of odd numbers left on the board after each operation is always odd. So when there is one number left, it must be odd, hence it cannot be 0.

---

(2) In an 8×8 board, there are 32 white pieces and 32 black pieces, one piece in each square. If a player can change all the white pieces to black and all the black pieces to white in any row or column in a single move, then is it possible that after finitely many moves, there will be exactly one black piece left on the board?

Solution. No. If there are exactly k black pieces in a row or column before a move is made to that row or column, then after the move, the number of black pieces in the row or column will become $8 - k$, a change of $(8 - k) - k = 8 - 2k$ black pieces to the board. Since $8 - 2k$ is even, the parity of the number of black pieces stays the same before and after the move. Since at the start, there are 32 black pieces, there cannot be 1 black piece left at any time.

---

(3) Four x's and five o's are written around the circle in an arbitrary order. If two consecutive symbols are the same, then insert a new x between them, otherwise insert a new o between them. Remove the old x's and o's. Keep on repeating this operation. Is it possible to get nine o's?

Solution. If we let $x = 1$ and $o = -1$, then note that consecutive symbols are replaced by their product. If we consider the product P of all nine values before and after each operation, we will see that the new P is the square of the old P. Hence, P will always equal 1 after an operation. So nine o's yielding $P = -1$ can never happen.

---

(4) There are three piles of stones numbering 19, 8 and 9, respectively. You are allowed to choose two piles and transfer one stone from each of these two piles to the third piles. After several of these operations, is it possible that each of the three piles has 12 stones?

Solution. No. Let the number of stones in the three piles be a, b and c, respectively. Consider (mod 3) of these numbers. In the beginning, they are $1, 2, 0$. After one operation, they become $0, 1, 2$ no matter which two piles have stones transfer to the third pile. So the remainders are always $0, 1, 2$ in some order. Therefore, all piles having 12 stones are impossible.

---

(5) Two boys play the following game with two piles of candies. In the first pile, there are 12 candies and in the second pile, there are 13 candies. Each boy takes turn to make a move consisting of eating two candies from one of the piles or transfering a candy from the first pile to the second. The boy who cannot make a move loses. Show that the boy who played second cannot lose. Can he win?

Solution. Consider S to be the number of candies in the second pile minus the first. Initially, $S = 13 - 12 = 1$. After each move, S increases or decreases by 2. So S (mod 4) has the pattern $1, 3, 1, 3, \ldots$. Every time after the boy who played first made a move , S (mod 4) would always be 3. Now a boy loses if and only if there are no candies left in the first pile and one candy left in the second pile, then $S = 1 - 0 = 1$. So the boy who played second can always make a move, hence cannot lose.

Since either the total number of candies decreases or the number of candies in the first pile decreases, so eventually the game must stop, so the boy who played second must win.

---

(6) Each member of a club has at most three enemies in the club. (Here enemies are mutual.) Show that the members can be divided into two groups so that each member in each group has at most one enemy in the group.

Solution. In the beginning, randomly divide the members into two groups. Consider the number S of pairs of enemies in the same group. If a member has at least two enemies in the same group, then the member has at most one enemy in the other group. Transferring the member to the other group, we will decrease S by at least one. Since S is a nonnegative integer, it cannot be decreased forever. So after finitely many transfers, each member can have at most one enemy in the same group.

Remarks. This method of proving is known as *the method of infinite descent.* It showed that you cannot always decrease a quantity when it can only have finitely many possible values.

---

(7) (1961 All-Russian Math Olympiad) Real numbers are written in an $m \times n$ table. It is permissible to reverse the signs of all the numbers in any row or column. Prove that after a number of these operations, we can make the sum of the numbers along each line (row or column) nonnegative.

Solution. Let S be the sum of all the mn numbers in the table. Note that after an operation, each number stays the name or turns to its negative. Hence there are at most 2^{mn} tables. So S can only have finitely many possible values. To make the sum of the numbers in each line nonnegative, just look for a line whose numbers have a negative sum. If no such line exists, then we are done. Otherwise, reverse the sign of all the numbers in the line. Then S increases. Since S has finitely many possible values, S can increase finitely many times. So eventually the sum of the numbers in every line must be nonnegative.

---

(8) Given $2n$ points in the plane with no three of them collinear. Show that they can be divided into n pairs such that the n segments joining each pair do not intersect.

Solution. In the beginning randomly pair the points and join the segments. Let S be the sum of the lengths of the segments. (Note that since there are finitely many ways of connecting these $2n$ points by n segments, there are finitely many possible

values of S.) If two segments AB and CD intersect at O, then replace pairs AB and CD by AC and BD. Since $AB+CD = AO+OB+CO+OD > AC+BD$ by the triangle inequality, whenever there is an intersection, doing this replacement will always decrease S. Since there are only finitely many possible values for S, so eventually there will not be any intersection.

Exercises

1. Every number from 1 to 1,000,000 is replaced by the sum of its digits. The resulting numbers are repeatedly subjected to the same operation until all the numbers have one digit. Will the number of ones in the end be greater or less than the number of twos? (*Hint*: Show that the sum of digits of n is congruent to n (mod 9).)

2. (1989 Hungarian Math Olympiad) In the vertices of a square, we placed some matches. Intially there is one match at one vertex and no match at the other three vertices. In one move, it is allowed to remove any number of matches at one vertex and placed at the two adjacent vertices a total of twice the number of matches removed. After finitely many moves, can the number of matches be 1, 9, 8, 9 counting clockwise or counterclockwise at the four vertices?

*3. In each square of an 8×8 table an integer is written. We can choose an arbitrary 3×3 or 4×4 subtable and increase all the numbers in it by one. Is it possible to obtain numbers divisible by three in all squares of the 8×8 table after finitely many such operations? (*Hint*: Mark some squares so that every 3×3 or 4×4 subtable will cover a multiple of 3 marked squares.)

*4. The numbers $1, 2, \ldots, n$ are arranged in some order on a line. We can exchange any two adjacent numbers. Prove that an odd number of such exchanges produces an arrangement necessarily different from the initial one.

5. Several numbers are written around a circle. If four consecutive numbers a, b, c, d satisfy $(a-d)(b-c) > 0$, we can exchange b and c. Prove that we can perform this operation only finitely many times.

6. Each face of a cube has a number written on it, and not all the numbers are the same. Each of the numbers is replaced by the arithmetic mean of the numbers

written on the four adjacent faces. Is it possible to obtain the initial numbers on the faces again after at least one such operations?

7. Finitely many squares of an infinite square grid drawn on white paper are painted black. At each moment in time $t = 1, 2, 3, \ldots$ each square takes the color of the majority of the following squares: the square itself and its top and right-hand neighbors. Prove that some time later there will be no black square at all.

8. In the plane, there are n points, no three collinear, and n lines, no two are parallel. Prove that we can drop a perpendicular from each point to one of the lines, one perpendicular per line, such that no two perpendiculars intersect.

8. Mathematical Games (Part II)

There are many mathematical games involving strategies to win or to defend. These games may involve number theoretical properties or geometrical decompositions or combinatorial reasonings. Some games may go on forever, while some games come to a stop eventually. A winning strategy is a scheme that allows the player to make moves to win the game no matter how the opponent plays. A defensive strategy cuts off the opponent's routes to winning. The following examples illustrate some standard techniques.

Examples. (1) There is a table with a square top. Two players take turn putting a dollar coin on the table. The player who cannot do so loses the game. Show that the first player can always win.

Solution. The first player puts a coin at the center. If the second player can make a move, the first player can put a coin at the position symmetrically opposite the position where the second player placed his coin with respect to the center of the table. Since the area of the available space is decreasing, the game must end eventually. The first player will win.

---

(2) (*Bachet's Game*) Initially, there are n checkers on the table, where $n > 0$. Two persons take turn to remove at least 1 and at most k checkers each time from the table. The last person who can remove any checker wins the game. For what values of n will the first person have a winning strategy? For what values of n will the second person have a winning strategy?

Solution. By testing simple cases of n, we can easily see that if n is *not* a multiple of $k+1$ in the beginning, then the first person has a winning strategy, otherwise the second person has a winning strategy.

To prove this, let n be the numbers of checkers on the table. If $n = (k+1)q + r$ with $0 < r < k+1$, then the first person can win by removing r checkers each time. (Note $r > 0$ every time at the first person's turn since in the beginning it is so and the second person starts with a multiple of $k+1$ checkers each time and can only remove 1 to k checkers.)

However, if n is a multiple of $k+1$ in the beginning, then no matter how many checkers the first person takes, the second person can now win by removing r checkers every time.

(3) (*Game of Nim*) There are 3 piles of checkers on the table. The first, second and third piles have x, y and z checkers respectively in the beginning, where $x, y, z > 0$. Two persons take turn choosing one of the three piles and removing at least one to all checkers in that pile each time from the table. The last person who can remove any checker wins the game. Who has a winning strategy?

Solution. In base 2 representations, let

$$x = (a_1a_2\cdots a_n)_2,\ y = (b_1b_2\cdots b_n)_2,\ z = (c_1c_2\cdots c_n)_2,\ N = (d_1d_2\cdots d_n)_2,$$

where $d_i \equiv a_i + b_i + c_i \pmod 2$. (Here we may have to add zeros on the left of the base 2 representations of x, y, z to ensure they have to the same number of digits.) The first person has a winning strategy if and only if N is not 0, i.e. not all d_i's are 0.

To see this, suppose N is not 0. The winning strategy is to remove checkers so N becomes 0. When the d_i's are not all zeros, look at the smallest i such that $d_i = 1$, then one of a_i, b_i, c_i equals 1, say $a_i = 1$. Then remove checkers from the first pile so that x has $(e_ie_{i+1}\cdots e_n)_2$ checkers left, where

$$e_j = \begin{cases} a_j & \text{if } d_j = 0 \\ 1 - a_j & \text{if } d_j = 1. \end{cases}$$

(For example, if $x = (1000)_2$ and $N = (1001)_2$, then change x to $(0001)_2$.) Note $e_i = 0$ after the move so that the new x is less than the old x. Also, after the move, N becomes 0 as all the $d_j = 1$ before the move will become 0 by the definition of e_j. So the first person can always make a move. The second person will always have $N = 0$ at his turn and making any move will result in at least one d_i not 0, i.e. $N \neq 0$. As the number of checkers is decreasing, eventually the second person cannot make a move and will lose the game.

---

(4) Twenty girls are sitting around a table and are playing a game with n cards. Initially, one girl holds all the cards. In each turn, if at least one girl holds at least two cards, one of these girls must pass a card to each of her two neighbors. The game ends if and only if each girl is holding at most one card.

(a) Prove that if $n \geq 20$, then the game cannot end.

(b) Prove that if $n < 20$, then the game must end eventually.

Solution. (a) If $n > 20$, then by pigeonhole principle, there is at least one girl holding at least two cards every moment. So the game cannot end.

If $n = 20$, then label the girls $G_1, G_2, \ldots, G_{20}$ in the clockwise direction and let G_1 be the girl holding all the cards initially. Define the current value of a card to be i if it is being held by G_i. Let S be the total value of the cards. Initially, $S = 20$.

Consider before and after G_i passes a card to each of her neighbors. If $i = 1$ then S increases by $-1 - 1 + 2 + 20 = 20$. If $1 < i < 20$, then S does not change because $-i - i + (i-1) + (i+1) = 0$. If $i = 20$, then S decreases by 20 because $-20 - 20 + 1 + 19 = -20$. So before and after moves, S is always a multiple of 20. Assume the game will end. Then each girl holds a card and $S = 1 + 2 + \cdots + 20 = 210$, which is not a multiple of 20, a contradiction. So the game cannot end.

(b) To see the game must end if $n < 20$, let's have the two girls sign the card when it is the first time one of them passes card to the other. Whenever one girl passes a card to her neighbor, let's require the girl to use the signed card between the pair if available. So a signed card will be stuck between the pair who signed it. If $n < 20$, there will be a pair of neighbors who never signed any card, hence never exchange any card.

If the game can go on forever, record the number of times each girl passed cards. Since the game can go on forever, not every girl passed card finitely many time. Also, since there are $n < 20$ cards and 20 pairs of girls sitting next to each other, there is a pair of girls who do not sign any card, hence have no exchange. Moving clockwise one girl at a time from this pair, eventually there is a pair G_i and G_{i+1} such that G_i passed cards finitely many times and G_{i+1} passed cards infinitely many times. This is clearly impossible since G_i will eventually stopped passing cards and would keep on receiving cards from G_{i+1}.

---

(5) (1996 Irish Math Olympiad) On a 5×9 rectangular chessboard, the following game is played. Initially, a number of discs are randomly placed on some of the squares, no square containing more than one disc. A turn consists of moving all of the discs subject to the following rules:

(i) each disc may be moved one square up, down, left or right;

(ii) if a disc moves up or down on one turn, it must move left or right on the next turn, and vice versa;

(iii) at the end of each turn, no square can contain two or more discs.

The game stops if it becomes impossible to complete another turn. Prove that if initially 33 discs are placed on the board, the game must eventually stop. Prove also that it is possible to place 32 discs on the board so that the game can continue forever.

Solution. If 32 discs are placed in the lower right 4×8 rectangle, they can all move up, left, down, right, repeatedly. To show that a game with 33 discs must stop eventually, label the board as shown below:

1	2	1	2	1	2	1	2	1
2	3	2	3	2	3	2	3	2
1	2	1	2	1	2	1	2	1
2	3	2	3	2	3	2	3	2
1	2	1	2	1	2	1	2	1

Note that there are eight 3's. A disc on 1 goes to a 3 after two moves, a disc on 2 goes to a 1 or 3 immediately, and a disc on 3 goes to a 2 immediately. Thus if k discs start on 1 and $k > 8$, the game stops because there are not enough 3's to accommodate these discs. Thus we assume $k \leq 8$, in which case there are at most sixteen discs with squares on 1's or 3's at the start, and so at least seventeen on 2's. Of these seventeen, at most eight can move onto 3's after one move, so at least nine end up on 1's. These discs will not all be able to move onto 3's two moves later. So the game will stop.

---

(6) (1995 Israeli Math Olympiad) Two players play a game on an infinite board that consists of 1×1 squares. Player I chooses a square and marks it with an O. Then, player II chooses another square and marks it with an X. They play until one of the players marks a row or a column of 5 consecutive squares, and this player wins the game. If no player can achieve this, the game is a tie. Show that player

II can prevent player I from winning.

	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
…	1	2	3	3	1	2	3	3	…
…	1	2	4	4	1	2	4	4	…
…	3	3	1	2	3	3	1	2	…
…	4	4	1	2	4	4	1	2	…
…	1	2	3	3	1	2	3	3	…
…	1	2	4	4	1	2	4	4	…
…	3	3	1	2	3	3	1	2	…
…	4	4	1	2	4	4	1	2	…
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Solution. Label the squares as shown above. Note that each number occurs in a pair. The 1's and the 2's are in vertical pairs and the 3's and the 4's are in horizontal pairs. Whenever player I marks a square, player II will mark the other square in the pair. Since any 5 consecutive vertical or horizontal squares must contain a pair of the same numbers, so player I cannot win.

---

(7) (1999 USAMO) The Y2K Game is played on a 1×2000 grid as follow. Two players in turn write either an S or an O in an empty square. The first player who produces three consecutive boxes that spell SOS wins. If all boxes are filled without producing SOS then the game is a draw. Prove that the second player has a winning strategy.

Solution. Call an empty square *bad* if playing an S or an O in that square will let the next player gets SOS in the next move.

<u>*Key Observation*</u>: A square is bad if and only if it is in a block of 4 consecutive squares of the form S**S, where * denotes an empty square.

(*Proof.* Clearly, the empty squares in S**S are bad. Conversely, if a square is bad, then playing an O there will allow an SOS in the next move by the other player. Thus the bad square must have an S on one side and an empty square on the other side. Playing an S there will also lose the game in the next move, which means there must be another S on the other side of the empty square.)

Now the second player's winning strategy is as follow: after the first player made a move, play an S at least 4 squares away from either end of the grid and from the first player's first move. On the second move, the second player will play an S three squares away from the second player's first move so that the squares in between are empty. (If the second move of the first player is next to or one square away from the first move of the second player, then the second player will place the second S on the other side.) After the second move of the second player, there are 2 bad squares on the board. So eventually somebody will fill these squares and the game will not be a draw.

On any subsequent move, when the second player plays, there will be an odd number of empty squares and an even number of bad squares, so the second player can always play a square that is not bad.

---

(8) (1993 IMO) On an infinite chessboard, a game is played as follow. At the start, n^2 pieces are arranged on the chessboard in an $n \times n$ block of adjoining squares, one piece in each square. A move in the game is a jump in a horizontal or vertical direction over an adjacent occupied square to an unoccupied square immediately beyond. The piece that has been jumped over is then removed. Find those values of n for which the game can end with only one piece remaining on the board.

Solution. Consider the pieces placed at the *lattice points* $\mathbb{Z}^2 = \{(x, y) : x, y \in \mathbb{Z}\}$. For $k = 0, 1, 2$, let $C_k = \{(x, y) \in \mathbb{Z}^2 : x + y \equiv k (\mathrm{mod}\ 3)\}$. Let a_k be the number of pieces placed at lattice points in C_k.

A horizontal move takes a piece at (x, y) to an unoccupied point $(x \pm 2, y)$ jumping over a piece at $(x \pm 1, y)$. After the move, each a_k goes up or down by 1. So each a_k changes parity. If n is divisible by 3, then $a_0 = a_1 = a_2 = n^2/3$ in the beginning. Hence at all time, the a_k's are of the same parity. So the game cannot end with one piece left causing two a_k's 0 and the remaining 1.

If n is not divisible by 3, then the game can end. We show this by induction. For $n = 1$ or 2, this is easily seen. For $n \geq 4$, we introduce an L-trick to reduce the $n \times n$ square pieces to $(n - 3) \times (n - 3)$ square pieces.

Consider pieces at $(0, 0), (0, 1), (0, 2), (1, 0)$. The moves $(1, 0) \mapsto (-1, 0)$, $(0, 2) \mapsto (0, 0)$, $(-1, 0) \mapsto (1, 0)$ remove 3 consecutive pieces in a column and

leave the fourth piece at its original lattice point. We can apply this trick repeatedly to the $3 \times (n-3)$ pieces on the bottom left part of the $n \times n$ square from left to right, then the $(n-3) \times 3$ pieces on the right side from bottom to top and finally the 3×3 pieces on top right part from right to left. This will leave $(n - 3) \times (n - 3)$ pieces. Therefore, the $n \times n$ case follows from the $(n - 3) \times (n - 3)$ case, completing the induction.

Exercises

1. In Bachet's game, if the rule changes to each person can remove 1 or 2 or 4 or 8 or 16 or ... (a power of 2) checkers each time, who can win? (Of course, the answer depends on the value of n.)

2. In Bachet's game, if the rule changes to each person can remove one or a prime number of checkers each time, who can win?

3. Initially there is a chip at the corner of an $n \times n$ board. A and B alternatively move the chip one square to the left, right, up or down. They may not move it to a square already visited. The loser is the one who cannot move.

   Who wins if n is even? Who wins if n is odd? Who wins if the chip starts on a square, which is a neighbor to a corner square?

4. Start with two piles of p and q chips, respectively. A and B move alternately. A move consists in removing any pile and splitting the other piles into two piles (of not necessarily equal number of chips). The loser is the one who cannot make move any more. Who wins? (*Hint*: Depends on the parities of p and q.)

5. A and B alternately color squares of a 4×4 chessboard. The loser is the one who first completes a colored 2×2 subsquare. Who can force a win? Can B force a draw?

6. A and B alternately move a knight on a 1994×1994 chessboard. A makes only horizontal moves $(x, y) \mapsto (x \pm 2, y \pm 1)$, B makes only vertical moves $(x, y) \mapsto (x \pm 1, y \pm 2)$. A starts by choosing a square and making a move. Visiting a square for a second time is not permitted. The loser is the one who cannot move. Prove that A has a winning strategy.

9. Coordinate Geometry

When we do a geometry problem, we should first look at the given facts and the conclusion. If all these involve intersection points, feet of perpendiculars, parallel lines, then there is a good chance we can solve the problem by coordinate geometry. However, if they involve two or more circles, angle bisectors and areas of triangles, then sometimes it is still possible to solve the problem by choosing a good place to put the origin and the x-axis. Below we will give some examples. *It is important to stay away from messy computations!*

Example 1. (1995 IMO) Let A, B, C and D be four distinct points on a line, in that order. The circles with diameters AC and BD intersect at the points X and Y. The line XY meets BC at the point Z. Let P be a point on the line XY different from Z. The line CP intersects the circle with diameter AC at the points C and M, and the line BP intersects the circle with diameter BD at the points B and N. Prove that the lines AM, DN and XY are concurrent.

(*Remarks.* Quite obvious we should set the origin at Z. Although the figure is not symmetric with respect to line XY, there are pairs such as M, N and A, D and B, C that are symmetric in *roles*! So we work on the left half of the figure, the computations will be similar for the right half.)

Solution. (*Due to Mok Tze Tao, 1995 Hong Kong Team Member*) Set the origin at Z and the x-axis on line AD. Let the coordinates of the circumcenters of triangles AMC and BND be $(x_1, 0)$ and $(x_2, 0)$, and the circumradii be r_1 and r_2, respectively. Then the coordinates of A and C are $(x_1 - r_1, 0)$ and $(x_1 + r_1, 0)$, respectively. Let the coordinates of P be $(0, y_0)$. Since $AM \perp CP$ and the slope of CP is $-\dfrac{y_0}{x_1 + r_1}$, the equation of AM works out to be $(x_1 + r_1)x - y_0 y = x_1^2 - r_1^2$. Let Q be the intersection of AM with XY, then Q has coordinates $(0, \dfrac{r_1^2 - x_1^2}{y_0})$. Similarly, let Q' be the intersection of DN with XY, then Q' has coordinates $(0, \dfrac{r_2^2 - x_2^2}{y_0})$. Since $r_1^2 - x_1^2 = ZX^2 = r_2^2 - x_2^2$, so $Q = Q'$.

Example 2. (1998 APMO) Let ABC be a triangle and D the foot of the altitude from A. Let E and F be on a line passing through D such that AE is perpendicular to BE, AF is perpendicular to CF, and E and F are different from D. Let M and N be the midpoints of the line segments BC and EF, respectively. Prove that AN is perpendicular to NM.

(*Remarks.* We can set the origin at D and the x-axis on line BC. Then computing the coordinates of E and F will be a bit messy. A better choice is to set the line through D, E, F horizontal.)

Solution. (*Due to Cheung Pok Man, 1998 Hong Kong Team Member*) Set the origin at A and the x-axis parallel to line EF. Let the coordinates of D, E, F be $(d, b), (e, b), (f, b)$, respectively. The case $b = 0$ leads to $D = E$, which is not allowed. So we may assume $b \neq 0$. Since $BE \perp AE$ and the slope of AE is b/e, so the equation of line BE works out to be $ex + by = e^2 + b^2$. Similarly, since $CF \perp AF$ and $BC \perp AD$, the equations of lines CF and BC are $fx + by = f^2 + b^2$ and $dx + by = d^2 + b^2$, respectively. Solving the equations for BF and BC, we find B has coordinate $(d + e, b - \dfrac{de}{b})$. Similarly, C has coordinate $(d + f, b - \dfrac{df}{b})$. Then M has coordinate $(d + \dfrac{e+f}{2}, b - \dfrac{de+df}{2b})$ and N has coordinate $(\dfrac{e+f}{2}, b)$. So the slope of AN is $\dfrac{2b}{e+f}$ and the slope of MN is $-\dfrac{e+f}{2b}$. Therefore, $AN \perp MN$.

Example 3. (2000 IMO) Two circles Γ_1 and Γ_2 intersect at M and N. Let ℓ be the common tangent to Γ_1 and Γ_2 so that M is closer to ℓ than N is. Let ℓ touch Γ_1 at A and Γ_2 at B. Let the line through M parallel to ℓ meet the circle Γ_1 again at C and the circle Γ_2 again at D. Lines CA and DB meet at E; lines AN and CD meet at P; lines BN and CD meet at Q. Show that $EP = EQ$.

(*Remarks.* Here if we set the x-axis on the line through the centers of the circles, then the equation of the line $\ell = AB$ will be complicated. So it is better to have line AB on the x-axis.)

Solution. Set the origin at A and the x-axis on line AB. Let B, M have coordinates $(b, 0), (s, t)$, respectively. Let the centers O_1, O_2 of Γ_1, Γ_2 be at $(0, r_1), (b, r_2)$, respectively. Then C, D have coordinates $(-s, t), (2b - s, t)$, respectively. Since

AB, CD are parallel, $CD = 2b = 2AB$ implies A, B are midpoints of CE, DE, respectively. So E is at $(s, -t)$. We see $EM \perp CD$.

To get $EP = EQ$, it is now left to show M is the midpoint of segment PQ. Since $O_1O_2 \perp MN$ and the slope of O_1O_2 is $\frac{r_2 - r_1}{b}$, the equation of line MN is $bx + (r_2 - r_1)y = bs + (r_2 - r_1)t$. (This line should pass through the midpoint of segment AB.) Since $O_2M = r_2$ and $O_1M = r_1$, we get

$$(b - s)^2 + (r_2 - t)^2 = r_2^2 \quad \text{and} \quad s^2 + (r_1 - t)^2 = r_1^2.$$

Subtracting these equations, we get $\frac{b^2}{2} = bs + (r_2 - r_1)t$, which implies $(\frac{b}{2}, 0)$ is on line MN. Since PQ, AB are parallel and line MN intersects AB at its midpoint, then M must be the midpoint of segment PQ. Together with $EM \perp PQ$, we get $EP = EQ$.

Example 4. (2000 APMO) Let ABC be a triangle. Let M and N be the points in which the median and the angle bisector, respectively, at A meet the side BC. Let Q and P be the points in which the perpendicular at N to NA meets MA and BA, respectively, and O the point in which the perpendicular at P to BA meets AN produced. Prove that QO is perpendicular to BC.

(*Remarks.* Here the equation of the angle bisector is a bit tricky to obtain unless it is the x-axis. In that case, the two sides of the angle is symmetric with respect to the x-axis.)

Solution. (*Due to Wong Chun Wai, 2000 Hong Kong Team Member*) Set the origin at N and the x-axis on line NO. Let the equation of line AB be $y = ax + b$, then A has coordinates $(-b/a, 0)$ and P has coordinates $(0, b)$. The equations of lines AC and PO are $y = -ax - b$ and $y = -\frac{1}{a}x + b$, respectively. Then O has coordinates $(ab, 0)$. Let the equation of BC be $y = cx$. Then B has coordinates $(\frac{b}{c-a}, \frac{bc}{c-a})$, C has coordinates $(-\frac{b}{c+a}, -\frac{bc}{c+a})$, M has coordinates $(\frac{ab}{c^2-a^2}, \frac{abc}{c^2-a^2})$ and Q has coordinates $(0, \frac{ab}{c})$. Then BC has slope c and QO has slope $-\frac{1}{c}$. Therefore, $QO \perp BC$.

Example 5. (1998 IMO) In the convex quadrilateral $ABCD$, the diagonals AC and BD are perpendicular and the opposite sides AB and DC are not parallel. Suppose that the point P, where the perpendicular bisectors of AB and DC meet, is inside $ABCD$. Prove that $ABCD$ is a cyclic quadrilateral if and only if the triangles ABP and CDP have equal areas.

(*Remarks.* The area of a triangle can be computed by taking the half length of the cross product. A natural candidate for the origin is P and having the diagonals parallel to the axes will be helpful.)

Solution. (*Due to Leung Wing Chung, 1998 Hong Kong Team Member*) Set the origin at P and the x-axis parallel to line AC. Then the equations of lines AC and BD are $y = p$ and $x = q$, respectively. Let $AP = BP = r$ and $CP = DP = s$. Then the coordinates of A, B, C, D are

$$(-\sqrt{r^2 - p^2}, p),\ (q, \sqrt{r^2 - q^2}),\ (\sqrt{s^2 - p^2}, p),\ (q, -\sqrt{s^2 - q^2}),$$

respectively. Using the determinant formula for finding the area of a triangle, we see that the areas of triangles ABP and CDP are equal if and only if

$$\frac{1}{2}\begin{vmatrix} -\sqrt{r^2 - p^2} & p \\ q & \sqrt{r^2 - q^2} \end{vmatrix} = \frac{1}{2}\begin{vmatrix} \sqrt{s^2 - p^2} & p \\ q & -\sqrt{s^2 - q^2} \end{vmatrix},$$

which after cancelling $\frac{1}{2}$ on both sides is equivalent to

$$-\sqrt{r^2 - p^2}\sqrt{r^2 - q^2} - pq = -\sqrt{s^2 - p^2}\sqrt{s^2 - q^2} - pq.$$

Since the function $f(x) = -\sqrt{x^2 - p^2}\sqrt{x^2 - q^2} - pq$ is strictly decreasing when $x \geq |p|$ and $|q|$, equality of areas hold if and only if $r = s$, which is equivalent to A, B, C, D concyclic.

After seeing these examples, we would like to remind the readers that there are pure geometric proofs to each of the problems. For examples (1) and (3), there are proofs that only take a few lines. We encouage the readers to discover these simple proofs.

Although in the opinions of many people, a pure geometric proof is better and more beautiful than a coordinate geometric proof, we should point out that

sometimes the coordinate geometric proofs may be prefered when there are many cases. For example (2), the different possible orderings of the points D, E, F on the line can all happen as some pictures will show. The coordinate geometric proofs above cover all cases.

Exercises

1. (1994 Canadian Math Olympiad) Let ABC be an acute triangle. Let D be on side BC such that $AD \perp BC$. Let H be a point on segment AD different from A and D. Let line BH intersect side AC at E and line CH intersect side AB at F. Prove that $\angle EDA = \angle FDA$.

2. Let E be a point inside triangle ABC such that $\angle ABE = \angle ACE$. Let F and G be the feet of the perpendiculars from E to the internal and external bisectors, respectively, of angle BAC. Prove that the line FG passes through the midpoint of BC.

3. Let circle Γ_2 lie inside circle Γ_1 and the two circles do not intersect. For any point A on Γ_1, let B and C be the two points on Γ_1 such that AB and AC are tangent to Γ_2. Prove that BC is tangent to Γ_2 if and only if $OI^2 = R^2 - 2Rr$, where O, I are the centers of Γ_1, Γ_2 and R, r are the radii of Γ_1, Γ_2, respectively. (*Hint*: Set the origin at A and the x-axis on AI.)

4. (1999 IMO) Two circles Γ_1 and Γ_2 are contained inside the circle Γ and are tangent to Γ at distinct points M and N, respectively. Γ_1 passes through the center of Γ_2. The line passing through the two points of intersection of Γ_1 and Γ_2 meet Γ at A and B. The lines MA and MB meet Γ_1 at C and D, respectively. Prove that CD is tangent to Γ_2. (*Hint*: Set the orign at M, x-axis on MO_1, $t = \angle OO_1O_2$, where O, O_1, O_2 are the centers of $\Gamma, \Gamma_1, \Gamma_2$, respectively.)

10. Vector Geometry

A *vector* $\overrightarrow{XY}$ is an object having a magnitude (the length XY) and a direction (from X to Y). Two vectors are considered the same if and only if they have the same magnitudes and directions. A vector $\overrightarrow{OX}$ from the origin O to a point X is called a *position vector*. For convenience, often a position vector $\overrightarrow{OX}$ will simply be denoted by X, when the position of the origin is understood, so that the vector $\overrightarrow{XY} = \overrightarrow{OY} - \overrightarrow{OX}$ will simply be $Y - X$. The length of the position vector $\overrightarrow{OX} = X$ will be denoted by $|X|$. We have the triangle inequality $|X + Y| \le |X| + |Y|$, with equality if and only if $X = tY$ for some $t \ge 0$. Also, $|cX| = |c||X|$ for number c.

For a point P on the line XY, in terms of position vectors, $P = tX + (1-t)Y$ for some real number t. If P is on the segment XY, then $t = \dfrac{PY}{XY} \in [0, 1]$.

Examples. (1) (*1980 Leningrad High School Math Olympiad*) Call a segment in a convex quadrilateral a *midline* if it joins the midpoints of opposite sides. Show that if the sum of the midlines of a quadrilateral is equal to its semiperimeter, then the quadrilateral is a parallelogram.

Solution. Let $ABCD$ be such a convex quadrilateral. Set the origin at A. The sum of the lengths of the midlines is $\dfrac{|B+C-D| + |D+C-B|}{2}$ and the semiperimeter is $\dfrac{|B| + |C-D| + |D| + |C-B|}{2}$. So

$$|B+C-D| + |D+C-B| = |B| + |C-D| + |D| + |C-B|.$$

By triangle inequality, $|B| + |C-D| \ge |B+C-D|$, with equality if and only if $B = t(C-D)$ (or $AB \| CD$). Similarly, $|D| + |C-B| \ge |D+C-B|$, with equality if and only if $AD \| BC$. For the equation to be true, both triangle inequalities must be equalities. In that case, $ABCD$ is a parallelogram.

---

(2) (*Crux Problem 2333*) D and E are points on sides AC and AB of triangle ABC, respectively. Also, DE is not parallel to CB. Suppose F and G are points of BC and ED, respectively, such that $BF : FC = EG : GD = BE : CD$. Show that GF is parallel to the angle bisector of $\angle BAC$.

Solution. Set the origin at A. Then $E = pB$ and $D = qC$ for some $p, q \in (0, 1)$. Let $t = \dfrac{BF}{FC}$, then $F = \dfrac{tC + B}{t+1}$ and $G = \dfrac{tD + E}{t+1} = \dfrac{tqC + pB}{t+1}$.

Since $BE = tCD$, so $(1-p)|B| = t(1-q)|C|$. Thus,

$$F - G = \frac{t(1-q)}{t+1}C + \frac{1-p}{t+1}B = \frac{(1-p)|B|}{t+1}\Big(\frac{C}{|C|} + \frac{B}{|B|}\Big).$$

This is parallel to $\dfrac{C}{|C|} + \dfrac{B}{|B|}$, which is in the direction of the angle bisector of $\angle BAC$.

The *dot product* of two vectors X and Y is the number $X \cdot Y = |X||Y|\cos\theta$, where θ is the angle between the vectors. Dot product has the following properties:

(1) $X \cdot Y = Y \cdot X$, $(X+Y)\cdot Z = X \cdot Z + Y \cdot Z$ and $(cX)\cdot Y = c(X \cdot Y)$.

(2) $|X|^2 = X \cdot X$, $|X \cdot Y| \le |X||Y|$ and $OX \perp OY$ if and only if $X \cdot Y = 0$.

Example. (3) (*1975 USAMO*) Let A, B, C, D denote four points in space and AB the distance between A and B, and so on. Show that

$$AC^2 + BD^2 + AD^2 + BC^2 \ge AB^2 + CD^2.$$

Solution. Set the origin at A. The inequality to be proved is

$$C\cdot C + (B-D)\cdot(B-D) + D\cdot D + (B-C)\cdot(B-C) \ge B\cdot B + (C-D)\cdot(C-D).$$

After expansion and regrouping, this is the same as $(B-C-D)\cdot(B-C-D) \ge 0$, with equality if and only if $B - C = D = D - A$, i.e. $BCAD$ is a parallelogram.

For a triangle ABC, the position vectors of its centroid is $G = \dfrac{A+B+C}{3}$. If we take the circumcenter O as the origin, then the position of the orthocenter is $H = A + B + C$ as $\overrightarrow{OH} = 3\overrightarrow{OG}$. Now for the incenter I, let a, b, c be the side lengths and AI intersect BC at D. Since $BD : CD = c : b$ and $DI : AI = \dfrac{ca}{b+c} : c = a : b+c$, so $D = \dfrac{bB+cC}{b+c}$ and $I = \dfrac{aA+bB+cC}{a+b+c}$.

Examples. (4) (*2nd Balkan Math Olympiad*) Let O be the center of the circle through the points A, B, C, and let D be the midpoint of AB. Let E be the centroid of triangle ACD. Prove that the line CD is perpendicular to line OE if and only if $AB = AC$.

Solution. Set the origin at O. Then

$$D = \frac{A+B}{2},\ E = \frac{A+C+D}{3} = \frac{3A+B+2C}{6},\ D - C = \frac{A+B-2C}{2}.$$

Hence $CD \perp OE$ if and only if $(A+B-2C)\cdot(3A+B+2C) = 0$. Since $A\cdot A = B\cdot B = C\cdot C$, this is equivalent to $A\cdot(B-C) = A\cdot B - A\cdot C = 0$, which is the same as $OA \perp BC$, i.e. $AB = AC$.

---

(5) (*1990 IMO Unused Problem, Proposed by France*) Given $\triangle ABC$ with no side equal to another side, let G, I and H be its centroid, incenter and orthocenter, respectively. Prove that $\angle GIH > 90^\circ$.

Solution. Set the origin at the circumcernter. Then

$$H = A + B + C, \quad G = \frac{A+B+C}{3}, \quad I = \frac{aA+bB+cC}{a+b+c}.$$

We need to show $(G-I)\cdot(H-I) = G\cdot H + I\cdot I - I\cdot(G+H) < 0$. Now $A\cdot A = B\cdot B = C\cdot C = R^2$ and $2B\cdot C = B\cdot B + C\cdot C - (B-C)\cdot(B-C) = 2R^2 - a^2, \ldots$. Hence,

$$G\cdot H = \frac{(A+B+C)\cdot(A+B+C)}{3} = 3R^2 - \frac{a^2+b^2+c^2}{3},$$

$$I\cdot I = \frac{(aA+bB+cC)\cdot(aA+bB+cC)}{(a+b+c)^2} = R^2 - \frac{abc}{a+b+c},$$

$$I\cdot(G+H) = \frac{4(aA+bB+cC)\cdot(A+B+C)}{3(a+b+c)} = 4R^2 - \frac{2[a^2(b+c)+b^2(c+a)+c^2(a+b)]}{3(a+b+c)}.$$

Thus, it is equivalent to proving $(a+b+c)(a^2+b^2+c^2) + 3abc > 2[a^2(b+c)+b^2(c+a)+c^2(a+b)]$, which after expansion and regrouping will become $a(a-b)(a-c)+b(b-c)(b-a)+c(c-a)(c-b) > 0$. To obtain this inequality, without loss of generality, assume $a > b > c > 0$. Then $a(a-b)(a-c) > b(a-b)(b-c)$ so that the sum of the first two terms is positive. As the third term is positive, the above inequality is true.

The *cross product* of two vectors X and Y is a vector $X \times Y$ having magnitude $|X||Y|\sin\theta$, where θ is the angle between the vectors, and direction perpendicular to the plane of X and Y satisfying the right hand rule. Cross product has the following properties:

(1) $X \times Y = -Y \times X$, $(X+Y) \times Z = X \times Z + Y \times Z$ and $(cX) \times Y = c(X \times Y)$.

(2) $\dfrac{|X \times Y|}{2}$ is the area of triangle XOY. When $X, Y \neq O$, $X \times Y = 0$ if and only if X, O, Y are collinear.

Examples. (6) (*1984 Annual Greek High School Competition*) Let $A_1A_2A_3A_4A_5A_6$ be a convex hexagon having its opposite sides parallel. Prove that triangles $A_1A_3A_5$ and $A_2A_4A_6$ have equal areas.

Solution. Set the orign at any point. As the opposite sides are parallel, $(A_1 - A_2) \times (A_4 - A_5) = 0$, $(A_3 - A_2) \times (A_5 - A_6) = 0$ and $(A_3 - A_4) \times (A_6 - A_1) = 0$. Expanding these equations and adding them, we get $A_1 \times A_3 + A_3 \times A_5 + A_5 \times A_1 = A_2 \times A_4 + A_4 \times A_6 + A_6 \times A_2$. Now

$$[A_1A_3A_5] = \frac{|(A_1 - A_3) \times (A_1 - A_5)|}{2} = \frac{|A_1 \times A_3 + A_3 \times A_5 + A_5 \times A_1|}{2}.$$

Similarly,

$$[A_2A_4A_6] = \frac{|A_2 \times A_4 + A_4 \times A_6 + A_6 \times A_2|}{2}.$$

So $[A_1A_3A_5] = [A_2A_4A_6]$.

---

(7) (*1996 Balkan Math Olympiad*) Let $ABCDE$ be a convex pentagon and let M, N, P, Q, R be the midpoints of sides AB, BC, CD, DE, EA, respectively. If the segments AP, BQ, CR, DM have a common point, show that this point also lies on EN.

Solution. Set the origin at the common point. Since, A, P and the origin are collinear,

$$0 = A \times P = A \times \left(\frac{C+D}{2}\right) = \frac{A \times C + A \times D}{2}.$$

So $A \times C = D \times A$. Similarly, $B \times D = E \times B$, $C \times E = A \times C$, $D \times A = B \times D$. Then $E \times B = C \times E$. So $E \times N = E \times \left(\frac{B+C}{2}\right) = 0$, which implies E, N and the origin are collinear.

(8) (*16th Austrian Math Olympiad*) A line intersects the sides (or sides produced) BC, CA, AB of triangle ABC in the points A_1, B_1, C_1, respectively. The points A_2, B_2, C_2 are symmetric to A_1, B_1, C_1 with respect to the midpoints of BC, CA, AB, respectively. Prove that A_2, B_2, C_2 are collinear.

Solution. Set the origin at a vertex, say C. Then $A_1 = c_1B$, $B_1 = c_2A$, $C_1 = A + c_3(B - A)$ for some constants c_1, c_2, c_3. Since A_1, B_1, C_1 are collinear,

$$0 = (B_1 - A_1) \times (C_1 - A_1) = (c_1 - c_1c_2 - c_1c_3 + c_2c_3)A \times B.$$

Since $A_2 = B - A_1 = (1 - c_1)B$, $B_2 = A - B_1 = (1 - c_2)A$ and $C_2 = (A + B) - C_1 = c_3A + (1 - c_3)B$, so A_2, B_2, C_2 are collinear if and only if

$$0 = (B_2 - A_2) \times (C_2 - A_2) = (c_1 - c_1c_2 - c_1c_3 + c_2c_3)A \times B,$$

which is true.

Two vectors X and Y are *linearly independent* if the equation $aX + bY = 0$ has exactly one solution, namely $a = b = 0$. This is the case if and only if $X, Y \neq O$ and X, O, Y are not collinear. If $aX + bY + cZ = 0$ for some a, b, c not all zeros and $a + b + c = 0$, then X, Y, Z are collinear because, say $a \neq 0$, we have $b + c = -a$ and $X = -\frac{b}{a}Y - \frac{c}{a}Z = \underbrace{\frac{b}{b+c}}_{t} Y + \underbrace{\frac{c}{b+c}}_{1-t} Z$ is on the line YZ.

Examples. (9) (*Pappus' Theorem*) A line, which passes through points A_1, A_2, A_3, intersects another line, which passes through points B_1, B_2, B_3, at a point O. Let lines A_2B_3, A_3B_2 intersect at P_1; lines A_1B_3, A_3B_1 intersect at P_2 and lines A_1B_2, A_2B_1 intersect at P_3. Show that P_1, P_2, P_3 are collinear.

Solution. Set the origin at O. Let A and B be unit vectors along the lines through A_i's and B_i's, respectively. Then $A_i = r_iA$ and $B_i = s_iB$ for some constants r_i, s_i. Now

$$P_3 = tA_1 + (1-t)B_2 = tr_1A + (1-t)s_2B \quad \text{for some } t$$

and

$$P_3 = uA_2 + (1-u)B_1 = ur_2A + (1-u)s_1B \quad \text{for some } u.$$

Subtracting, we find $0 = (tr_1 - ur_2)A + [(1-t)s_2 - (1-u)s_1]B$. Since A and B are linearly independent, we must have $tr_1 - ur_2 = 0$ and $(1-t)s_2 - (1-u)s_1 = 0$. After solving for t and u, we find $P_3 = \dfrac{r_1r_2(s_1 - s_2)A + s_1s_2(r_1 - r_2)B}{r_1s_1 - r_2s_2}$. Similarly,

$$P_1 = \frac{r_2r_3(s_2 - s_3)A + s_2s_3(r_2 - r_3)B}{r_2s_2 - r_3s_3}, \quad P_2 = \frac{r_3r_1(s_3 - s_1)A + s_3s_1(r_3 - r_1)B}{r_3s_3 - r_1s_1}.$$

Note that $r_1s_1(r_2s_2 - r_3s_3)P_1 + r_2s_2(r_3s_3 - r_1s_1)P_2 + r_3s_3(r_1s_1 - r_2s_2)P_3 = 0$ and the sum of the coefficients is also 0. So P_1, P_2, P_3 are collinear.

---

(10) (*Crux Problem 2457*) In quadrilateral $ABCD$, we have $\angle A + \angle B = 2\alpha < 180^\circ$, and $BC = AD$. Construct isosceles triangles DCI, ACJ and DBK, where I, J and K are on the other side of CD from A, such that

$$\angle ICD = \angle IDC = \angle JAC = \angle JCA = \angle KDB = \angle KBD = \alpha.$$

(a) Show that I, J and K are collinear.

(b) Establish how they are distributed on the line.

Solution. Let $f(\overrightarrow{XY})$ be rotation of $\overrightarrow{XY}$ about the origin through angle $180^\circ - 2\alpha \neq 0$. Then f is linear (i.e. $f(\overrightarrow{WX} + \overrightarrow{YZ}) = f(\overrightarrow{WX}) + f(\overrightarrow{YZ})$ and $f(c\overrightarrow{XY}) = cf(\overrightarrow{XY})$) and $f(\overrightarrow{XY}) = \overrightarrow{XY}$ implies $\overrightarrow{XY} = 0$.

Since $\angle JAC = \angle JCA = \alpha$, so $\angle AJC = 180^\circ - 2\alpha$, which implies $f(\overrightarrow{JA}) = \overrightarrow{JC}$. Similarly, $f(\overrightarrow{ID}) = \overrightarrow{IC}$ and $f(\overrightarrow{KD}) = \overrightarrow{KB}$. Since $\angle A + \angle B = 2\alpha$, so $f(\overrightarrow{DA}) = \overrightarrow{CB}$. Now by linearity,

$$f(\overrightarrow{IJ}) = f(\overrightarrow{ID} + \overrightarrow{DA} + \overrightarrow{AJ}) = \overrightarrow{IC} + \overrightarrow{CB} - \overrightarrow{JC} = \overrightarrow{IB} + \overrightarrow{CJ},$$

$$f(\overrightarrow{IK}) = f(\overrightarrow{ID} + \overrightarrow{DK}) = \overrightarrow{IC} - \overrightarrow{KB} = \overrightarrow{IC} + \overrightarrow{BK}.$$

Then $f(\overrightarrow{IJ} + \overrightarrow{IK}) = \overrightarrow{IB} + \overrightarrow{CJ} + \overrightarrow{IC} + \overrightarrow{BK} = \overrightarrow{IJ} + \overrightarrow{IK}$. This implies $\overrightarrow{IJ} + \overrightarrow{IK} = 0$, which means I, J, K are collinear and I is the midpoint of segment JK.

Exercises

1. (*1997 Hungarian Math Olympiad*) Let R be the circumradius of triangle ABC, and let G and H be its centroid and orthocenter, respectively. Let F be the midpoint of GH. Show that $AF^2 + BF^2 + CF^2 = 3R^2$.

2. Let ABC be any triangle. Two squares $BAEP$ and $ACRD$ are constructed externally to ABC. Let M and N be the midpoints of BC and ED, respectively. Show that $AM \perp ED$ and $AN \perp BC$.

3. (*1996 Vietnamese Math Olympiad*) Let $ABCD$ be a tetrahedron with $AB = AC = CD$ and circumcenter O. Let G be the centroid of triangle ACD, let E be the midpoint of BG, and let F be the midpoint of AE. Prove that OF is perpendicular to BG if and only if OD is perpendicular to AC.

4. (*2001 IMO Unused Problem*) Let ABC be a triangle with centroid G. Determine, with proof, the position of the point P in the plane of ABC such that $AP \cdot AG + BP \cdot BG + CP \cdot CG$ is a minimum, and express this minimum value in terms of the side lengths of ABC. (*Hint*: Set the origin at G.)

5. (*Pedoe's Inequality*) Let a, b, c and a_1, b_1, c_1 be the side lengths of triangles ABC and $A_1B_1C_1$, respectively. Supposes their areas are S and S_1, respectively. Prove that

$$a_1^2(-a^2 + b^2 + c^2) + b_1^2(a^2 - b^2 + c^2) + c_1^2(a^2 + b^2 - c^2) \geq 16SS_1.$$

6. (*Desargue's Theorem*) For two triangles ABC and $A_1B_1C_1$, let lines BC, B_1C_1 intersect at P; lines CA, C_1A_1 intersect at Q and lines AB, A_1B_1 intersect at R. Prove that lines AA_1, BB_1, CC_1 are concurrent if and only if P, Q, R are collinear.

7. (*2004 APMO*) Let O be the circumcenter and H be the orthocentre of an acute triangle ABC. Prove that the area of one of the triangles AOH, BOH and COH is equal to the sum of the areas of the other two.

11. Pell's Equation

Let d be a positive integer that is not a square. The equation $x^2 - dy^2 = 1$ with variables x, y over integers is called *Pell's Equation.* It was Euler who attributed the equation to John Pell (1611-1685), although Brahmagupta (7th century), Bhaskara (12th century) and Fermat had studied the equation in details earlier.

A solution (x, y) of Pell's equation is called *positive* if both x and y are positive integers. Hence, positive solutions correspond to the lattice points in the first quadrant that lie on the hyperbola $x^2 - dy^2 = 1$. A positive solution (x_1, y_1) is called the *least positive solution* (or *fundamental solution*) if it satisfies $x_1 < x$ and $y_1 < y$ for every other positive solution (x, y). (As the hyperbola $x^2 - dy^2 = 1$ is strictly increasing in the first quadrant, the conditions for being least are the same as requiring $x_1 + y_1\sqrt{d} < x + y\sqrt{d}$.)

Theorem. *Pell's equation $x^2 - dy^2 = 1$ has infinitely many positive solutions. If (x_1, y_1) is the least positive solution, then for $n = 1, 2, 3, \ldots$, define $x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n$. The pairs (x_n, y_n) are all the positive solutions of the Pell's equation. The x_n's and y_n's are strictly increasing to infinity and satisfy the recurrence relations $x_{n+2} = 2x_1x_{n+1} - x_n$ and $y_{n+2} = 2x_1y_{n+1} - y_n$.*

We will comment on the proof. The least positive solution is obtained by writing $\sqrt{d}$ as a simple continued fraction. It turns out

$$\sqrt{d} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}},$$

where $a_0 = [\sqrt{d}]$ and $a_1, a_2, \ldots$ is a periodic positive integer sequence. The continued fraction will be denoted by $\langle a_0, a_1, a_2, \ldots\rangle$. The *k-th convergent* of $\langle a_0, a_1, a_2, \ldots\rangle$ is the number $\dfrac{p_k}{q_k} = \langle a_0, a_1, a_2, \ldots, a_k\rangle$ with p_k, q_k relatively prime. Let $a_1, a_2, \ldots, a_m$ be the period for $\sqrt{d}$. The least positive solution of Pell's equation turns out to be

$$(x_1, y_1) = \begin{cases} (p_{m-1}, q_{m-1}) & \text{if } m \text{ is even} \\ (p_{2m-1}, q_{2m-1}) & \text{if } m \text{ is odd} \end{cases}.$$

For example, $\sqrt{3} = \langle 1, 1, 2, 1, 2, \ldots\rangle$ and so $m = 2$, then $\langle 1, 1\rangle = \dfrac{2}{1}$. We check $2^2 - 3 \cdot 1^2 = 1$ and clearly, $(2, 1)$ is the least positive solution of $x^2 - 3y^2 = 1$. Next, $\sqrt{2} = \langle 1, 2, 2, \ldots\rangle$ and so $m = 1$, then $\langle 1, 2\rangle = \dfrac{3}{2}$. We check $3^2 - 2 \cdot 2^2 = 1$ and again clearly $(3, 2)$ is the least positive solution of $x^2 - 2y^2 = 1$.

Next, if there is a positive solution (x, y) such that $x_n + y_n\sqrt{d} < x + y\sqrt{d} < x_{n+1} + y_{n+1}\sqrt{d}$, then consider $u + v\sqrt{d} = (x + y\sqrt{d})/(x_n + y_n\sqrt{d})$. We will get $u + v\sqrt{d} < x_1 + y_1\sqrt{d}$ and $u - v\sqrt{d} = (x - y\sqrt{d})/(x_n - y_n\sqrt{d})$ so that $u^2 - dv^2 = (u - v\sqrt{d})(u + v\sqrt{d}) = 1$, contradicting (x_1, y_1) being the least positive solution.

To obtain the recurrence relations, note that $(x_1 + y_1\sqrt{d})^2 = x_1^2 + dy_1^2 + 2x_1y_1\sqrt{d} = 2x_1^2 - 1 + 2x_1y_1\sqrt{d} = 2x_1(x_1 + y_1\sqrt{d}) - 1$. So

$$\begin{aligned} x_{n+2} + y_{n+2}\sqrt{d} &= (x_1 + y_1\sqrt{d})^2(x_1 + y_1\sqrt{d})^n \\ &= 2x_1(x_1 + y_1\sqrt{d})^{n+1} - (x_1 + y_1\sqrt{d})^n \\ &= 2x_1x_{n+1} - x_n + (2x_1y_{n+1} - y_n)\sqrt{d}. \end{aligned}$$

The related equation $x^2 - dy^2 = -1$ may not have a solution, for example, $x^2 - 3y^2 = -1$ cannot hold as $x^2 - 3y^2 \equiv x^2 + y^2 \not\equiv -1 \pmod 4$. However, if d is a prime and $d \equiv 1 \pmod 4$, then a theorem of Lagrange asserts that it will have a solution. In general, if $x^2 - dy^2 = -1$ has a least positive solution (x_1, y_1), then all its positive solutions are pairs (x, y), where $x + y\sqrt{d} = (x_1 + y_1\sqrt{d})^{2n-1}$ for some positive integer n.

In passing, we remark that some k-th convergent numbers are special. If the length m of the period for $\sqrt{d}$ is even, then $x^2 - dy^2 = 1$ has $(x_n, y_n) = (p_{nm-1}, q_{nm-1})$ as all its positive solutions, but $x^2 - dy^2 = -1$ has no integer solution. If m is odd, then $x^2 - dy^2 = 1$ has (p_{jm-1}, y_{jm-1}) with j even as all its positive solutions and $x^2 - dy^2 = -1$ has (p_{jm-1}, q_{jm-1}) with j odd as all its positive solutions.

Example 1. Prove that there are infinitely many triples of consecutive integers each of which is a sum of two squares.

Solution. The first such triple is $8 = 2^2 + 2^2, 9 = 3^2 + 0^2, 10 = 3^2 + 1^2$, which suggests we consider triples $x^2 - 1, x^2, x^2 + 1$. Since $x^2 - 2y^2 = 1$ has infinitely

many positive solutions (x, y), we see that $x^2 - 1 = y^2 + y^2$, $x^2 = x^2 + 0^2$ and $x^2 + 1$ satisfy the requirement and there are infinitely many such triples.

Example 2. Find all triangles whoses sides are consecutive integers and areas are also integers.

Solution. Let the sides be $z - 1, z, z + 1$. Then the semiperimeter $s = \dfrac{3z}{2}$ and the area is $A = \dfrac{z\sqrt{3(z^2-4)}}{4}$. If A is an integer, then z cannot be odd, say $z = 2x$, and $z^2 - 4 = 3w^2$. So $4x^2 - 4 = 3w^2$, which implies w is even, say $w = 2y$. Then $x^2 - 3y^2 = 1$, which has $(x_1, y_1) = (2, 1)$ as the least positive solution. So all positive solutions are (x_n, y_n), where $x_n + y_n\sqrt{3} = (2 + \sqrt{3})^n$. Now $x_n - y_n\sqrt{3} = (2 - \sqrt{3})^n$. Hence,

$$x_n = \frac{(2+\sqrt{3})^n + (2-\sqrt{3})^n}{2} \quad \text{and} \quad y_n = \frac{(2+\sqrt{3})^n - (2-\sqrt{3})^n}{2\sqrt{3}}.$$

The sides of the triangles are $2x_n - 1, 2x_n, 2x_n + 1$ and the areas are $A = 3x_n y_n$.

Example 3. Find all positive integers k, m such that $k < m$ and

$$1 + 2 + \cdots + k = (k + 1) + (k + 2) + \cdots + m.$$

Solution. Adding $1 + 2 + \cdots + k$ to both sides, we get $2k(k + 1) = m(m + 1)$, which can be rewritten as $(2m + 1)^2 - 2(2k + 1)^2 = -1$. Now the equation $x^2 - 2y^2 = -1$ has $(1, 1)$ as its least positive solution. So its positive solutions are pairs (x_n, y_n), where $x_n + y_n\sqrt{2} = (1 + \sqrt{2})^{2n-1}$. Then

$$x_n = \frac{(1+\sqrt{2})^{2n-1} + (1-\sqrt{2})^{2n-1}}{2} \quad \text{and} \quad y_n = \frac{(1+\sqrt{2})^{2n-1} - (1-\sqrt{2})^{2n-1}}{2\sqrt{2}}.$$

Since $x^2 - 2y^2 = -1$ implies x is odd, so x is of the form $2m + 1$. Then $y^2 = 2m^2 + 2m + 1$ implies y is odd, so y is of the form $2k + 1$. Then $(k, m) = \left(\dfrac{y_n - 1}{2}, \dfrac{x_n - 1}{2}\right)$ with $n = 2, 3, 4, \ldots$ are all the solutions.

Example 4. Prove that there are infinitely many positive integers n such that n^2+1 divides $n!$.

Solution. The equation $x^2 - 5y^2 = -1$ has $(2, 1)$ as the least positive solution. So it has infinitely many positive solutions. Consider those solutions with $y > 5$. Then $5 < y < 2y \le x$ as $4y^2 \le 5y^2 - 1 = x^2$. So $2(x^2 + 1) = 5 \cdot y \cdot 2y$ divides $x!$, which is more than we want.

Example 5. For the sequence $a_n = [\sqrt{n^2 + (n + 1)^2}]$, prove that there are infinitely many n's such that $a_n - a_{n-1} > 1$ and $a_{n+1} - a_n = 1$.

Solution. First consider the case $n^2 + (n + 1)^2 = y^2$, which can be rewritten as $(2n + 1)^2 - 2y^2 = -1$. As in example 3 above, $x^2 - 2y^2 = -1$ has infinitely many positive solutions and each x is odd, say $x = 2n + 1$ for some n. For these n's, $a_n = y$ and $a_{n-1} = [\sqrt{(n-1)^2 + n^2}] = [\sqrt{y^2 - 4n}]$. The equation $y^2 = n^2 + (n + 1)^2$ implies $n > 2$ and $a_{n-1} \le \sqrt{y^2 - 4n} < y - 1 = a_n - 1$. So $a_n - a_{n-1} > 1$ for these n's.

Also, for these n's, $a_{n+1} = [\sqrt{(n+1)^2 + (n+2)^2}] = [\sqrt{y^2 + 4n + 4}]$. As $n < y < 2n + 1$, we easily get $y + 1 < \sqrt{y^2 + 4n + 4} < y + 2$. So $a_{n+1} - a_n = (y + 1) - y = 1$.

Example 6. (American Math Monthly E2606, proposed by R. S. Luthar) Show that there are infinitely many integers n such that $2n + 1$ and $3n + 1$ are perfect squares, and that such n must be multiples of 40.

Solution. Consider $2n + 1 = u^2$ and $3n + 1 = v^2$. On one hand, $u^2 + v^2 \equiv 2 \pmod 5$ implies $u^2, v^2 \equiv 1 \pmod 5$, which means n is a multiple of 5.

On the other hand, we have $3u^2 - 2v^2 = 1$. Setting $u = x + 2y$ and $v = x + 3y$, the equation becomes $x^2 - 6y^2 = 1$. It has infinitely many positive solutions. Since $3u^2 - 2v^2 = 1$, u is odd, say $u = 2k + 1$. Then $n = 2k^2 + 2k$ is even. Since $3n + 1 = v^2$, so v is odd, say $v = 4m \pm 1$. Then $3n = 16m^2 \pm 8m$, which implies n is also a multiple of 8.

Example 7. Prove that the only positive integral solution of $5^a - 3^b = 2$ is $a = b = 1$.

Solution. Clearly, if a or b is 1, then the other one is 1, too. Suppose (a, b) is a solution with both $a, b > 1$. Considering (mod 4), we have $1-(-1)^b \equiv 2 (\text{mod } 4)$, which implies b is odd. Considering (mod 3), we have $(-1)^a \equiv 2 (\text{mod } 3)$, which implies a is odd.

Setting $x = 3^b + 1$ and $y = 3^{(b-1)/2}5^{(a-1)/2}$, we get $15y^2 = 3^b5^a = 3^b(3^b + 2) = (3^b+1)^2-1 = x^2-1$. So (x, y) is a positive solution of $x^2-15y^2 = 1$. The least positive solution is $(4, 1)$. Then $(x, y) = (x_n, y_n)$ for some positive integer n, where $x_n+y_n\sqrt{15} = (4+\sqrt{15})^n$. After examinining the first few y_n's, we observe that y_{3k} are the only terms that are divisible by 3. However, they also seem to be divisible by 7, hence cannot be of the form 3^c5^d.

To confirm this, we use the recurrence relations on y_n. Since $y_1 = 1$, $y_2 = 8$ and $y_{n+2} = 8y_{n+1} - y_n$, taking y_n (mod 3), we get the sequence 1, 2, 0, 1, 2, 0, ... and taking y_n (mod 7), we get 1, 1, 0, -1, -1, 0, 1, 1, 0, -1, -1, 0, ...

Therefore, no $y = y_n$ is of the form 3^c5^d and $a, b > 1$ cannot be solution to $5^a - 3^b = 2$.

Example 8. Show that the equation $a^2 + b^3 = c^4$ has infinitely many solutions.

Solution. We will use the identity $1^3 + 2^3 + \cdots + n^3 = \left(\frac{n(n+1)}{2}\right)^2$, which is a standard exercise of mathematical induction. From the identity, we get $\left(\frac{(n-1)n}{2}\right)^2 + n^3 = \left(\frac{n(n+1)}{2}\right)^2$ for $n > 1$. All we need to do now is to show there are infinitely many positive integers n such that $n(n+1)/2 = k^2$ for some positive integers k. Then $(a, b, c) = ((n-1)n/2, n, k)$ solves the problem.

Now $n(n+1)/2 = k^2$ can be rewritten as $(2n+1)^2 - 2(2k)^2 = 1$. We know $x^2 - 2y^2 = 1$ has infinitely many positive solutions. For any such (x, y), clearly x is odd, say $x = 2m + 1$. Then $y^2 = 2m^2 + 2m$ implies y is even. So any such (x, y) is of the form $(2n + 1, 2k)$. Therefore, there are infinitely many such n.

For a fixed nonzero integer N, as the case $N = -1$ shows, the generalized equation $x^2 - dy^2 = N$ may not have a solution. If it has a least positive solution (x_1, y_1), then $x^2 - dy^2 = N$ has infinitely many positive solutions given by (x_n, y_n), where

$$x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})(a + b\sqrt{d})^{n-1}$$

and (a, b) is the least positive solution of $x^2 - dy^2 = 1$. However, in general these do not give all positive solutions of $x^2 - dy^2 = N$ as the following example will show.

Example 9. Consider the equation $x^2 - 23y^2 = -7$. It has $(x_1, y_1) = (4, 1)$ as the least positive solution. The next two solutions are $(19, 4)$ and $(211, 44)$. Now the least positive solution of $x^2 - 23y^2 = 1$ is $(a, b) = (24, 5)$. Since $(4 + \sqrt{23})(24 + 5\sqrt{23}) = 211 + 44\sqrt{23}$, the solution $(19, 4)$ is skipped by the formula above.

In case $x^2 - dy^2 = N$ has positive solutions, how do we get them all? A solution (x, y) of $x^2 - dy^2 = N$ is called *primitive* if x and y (and N) are relatively prime. For $0 \le s < |N|$, we say the solution belong to class C_s if $x \equiv sy \pmod{|N|}$. As x, y are relatively prime to N, so is s. Hence, there are at most $\phi(|N|)$ classes of primitive solutions, where $\phi(k)$ is *Euler's ϕ-function* denoting the number of positive integers $m \le k$ that are relatively prime to k. Also, for such s, $(s^2-d)y^2 \equiv x^2 - dy^2 \equiv 0 \pmod{|N|}$ and y, N relatively prime imply $s^2 \equiv d \pmod{|N|}$.

Theorem. *Let (a_1, b_1) be a C_s primitive solutions of $x^2 - dy^2 = N$. A pair (a_2, b_2) is also a C_s primitive solution of $x^2 - dy^2 = N$ if and only if $a_2 + b_2\sqrt{d} = (a_1 + b_1\sqrt{d})(u + v\sqrt{d})$ for some solution (u, v) of $x^2 - dy^2 = 1$.*

Proof. If (a_2, b_2) is C_s primitive, define $u + v\sqrt{d} = (a_2 + b_2\sqrt{d})/(a_1 + b_1\sqrt{d})$. Then $u = (a_1a_2 - db_1b_2)/N$ and $v = (a_1b_2 - b_1a_2)/N$. Also, $u - v\sqrt{d} = (a_2 - b_2\sqrt{d})/(a_1 - b_1\sqrt{d})$. Multiplying these two equations, we get $u^2 - dv^2 = N/N = 1$.

To see u, v are integers, note $a_1a_2 - db_1b_2 \equiv (s^2 - d)b_1b_2 \equiv 0 \pmod{|N|}$, which implies u is an integer. Since $a_1b_2 - b_1a_2 \equiv sb_1b_2 - b_1sb_2 = 0 \pmod{|N|}$, v is also an integer.

For the converse, multiplying the equation with its conjugate shows (a_2, b_2) solves $x^2 - dy^2 = N$. From $a_2 = ua_1 + dvb_1$ and $b_2 = ub_1 + va_1$, we get $a_1 = ua_2 - dvb_2$ and $b_1 = ub_2 - va_2$. Hence, common divisors of a_2, b_2 are also common divisors of a_1, b_1. So a_2, b_2 are relatively prime. Finally, $a_2 - sb_2 \equiv (usb_1 + dvb_1) - s(ub_1 + vsb_1) = (d - s^2)vb_1 \equiv 0 \pmod{|N|}$ concludes the proof.

Thus, all primitive solutions of $x^2 - dy^2 = N$ can be obtained by finding a solution (if any) in each class, then *multiply* them by solutions of $x^2 - dy^2 = 1$. For the nonprimitive solutions, we can factor the common divisors of a and b to reduce N.

Example 10. (1995 IMO proposal by USA leader T. Andreescu) Find the smallest positive integer n such that $19n + 1$ and $95n + 1$ are both integer squares.

Solution. Let $95n+1 = x^2$ and $19n+1 = y^2$, then $x^2-5y^2 = -4$. Now $\phi(4) = 2$ and $(1, 1)$, $(11, 5)$ are C_1, C_3 primitive solutions, respectively. As $(9, 4)$ is the least positive solution of $x^2 - 5y^2 = 1$ and $9 + 4\sqrt{5} = (2 + \sqrt{5})^2$, so the primitive positive solutions are pairs (x, y), where $x + y\sqrt{5} = (1 + \sqrt{5})(2 + \sqrt{5})^{2n-2}$ or $(11 + 5\sqrt{5})(2 + \sqrt{5})^{2n-2}$.

Since the common divisors of x, y divide 4, the nonprimitive positive solutions are the cases x and y are even. This reduces to considering $u^2 - 5v^2 = -1$, where we take $u = x/2$ and $v = y/2$. The least positive solution for $u^2 - 5v^2 = -1$ is $(2, 1)$. So $x + y\sqrt{5} = 2(u + v\sqrt{5}) = 2(2 + \sqrt{5})^{2n-1}$.

In attempt to combine these solutions, we look at the powers of $1 + \sqrt{5}$ coming from the least positive solutions $(1, 1)$. The powers are $1 + \sqrt{5}, 6 + 2\sqrt{5}, 16 + 8\sqrt{5} = 8(2 + \sqrt{5}), 56 + 24\sqrt{5}, 176 + 80\sqrt{5} = 16(11 + 5\sqrt{5}), \ldots$. Thus, the primitive positive solutions are (x, y) with $x + y\sqrt{5} = 2\big(\frac{1+\sqrt{5}}{2}\big)^{6n-5}$ or $2\big(\frac{1+\sqrt{5}}{2}\big)^{6n-1}$. The nonprimitive positive solutions are (x, y) with $x + y\sqrt{5} = 2\big(\frac{1+\sqrt{5}}{2}\big)^{6n-3}$. So the general positive solutions are (x, y) with

$$x + y\sqrt{5} = 2\big(\frac{1+\sqrt{5}}{2}\big)^{k} \quad \text{for odd } k.$$

Then

$$y = \frac{1}{\sqrt{5}}\Big(\big(\frac{1+\sqrt{5}}{2}\big)^{k} - \big(\frac{1-\sqrt{5}}{2}\big)^{k}\Big) = F_k,$$

where F_k is the k-th term of the famous *Fibonacci sequence.* Finally, $y^2 \equiv 1 \pmod{19}$ and k should be odd. The smallest such $y = F_{17} = 1597$, which leads to $n = (F_{17}^2 - 1)/19 = 134232$.

Comments: For the readers not familiar with the Fibonacci sequence, it is defined by $F_1 = 1$, $F_2 = 1$ and $F_{n+1} = F_n + F_{n-1}$ for $n > 1$. By math induction, we can check that they satisfy *Binet's formula* $F_n = (r_1^n - r_2^n)/\sqrt{5}$, where $r_1 = (1+\sqrt{5})/2$ and $r_2 = (1 - \sqrt{5})/2$ are the roots of the *characteristic equation* $x^2 = x + 1$. (Check cases $n = 1, 2$ and in the induction step, just use $r_i^{n+1} = r_i^n + r_i^{n-1}$.)