Problem 1: What is the smallest positive multiple of 11 whose decimal digits have an odd sum?

Solution 1: $209 = 11\cdot19$ is the smallest. Here is why: Because $10^N \equiv (-1)^N \mathbf{mod}\ 11$, an integer is divisible by 11 just when the alternating-signed sum of its digits is a multiple of 11. All two-digit multiples of 11 have even digit-sums. All three-digit multiples 1xy of 11 have even digit-sums $1 + x + y$ because $1 - x + y = 0$. Three-digit multiples 2xy of 11 have odd digit-sums $2 + x + y$ just when $2 - x + y = 11$; the only such instance is 209.

Prof. George Bergman suggested this problem. Brute force solves it by perusing the first 19 multiples of 11.

Problem 2: The number $x := (4 + \sqrt{17})^{714} = 3480003819317\ldots\bullet\ldots$ has infinitely many decimal digits after the decimal point $\bullet$, and fills some number N of decimal digits ahead of the point. What is the N^{th} decimal digit of x after the point? (N need not be known to solve this problem**!**)

Solution 2: The N^{th} digit is 7. You could compute $x \approx 3\mathbf{.}480003819317\ldots\cdot10^{649}$ easily using your calculator's logarithm key and hence infer that $N = 650$, but this is unnecessary. Whatever N may be, $x = 3\mathbf{.}48\ldots\cdot10^{N-1}$ and therefore $1/x = 2\mathbf{.}87\ldots\cdot10^{-N}$. The relevance of $1/x$ will be clear in a moment. First observe that $x = (4 + \sqrt{17})^{714} = J + K\sqrt{17}$ for some huge integers J and K which shall not be computed. Evidently the N^{th} digit of x after the point is the same as that of $K\sqrt{17}$ after the point. $1/x = (4 - \sqrt{17})^{714} = J - K\sqrt{17}$ because $4 - \sqrt{17} = -1/(4 + \sqrt{17})$. Therefore $K\sqrt{17} = J - 1/x$, and its N^{th} digit after the point is the same as that of

$$1 - 1/x = 1 - 2\mathbf{.}87\ldots\cdot10^{-N} = 0\mathbf{.}999\ldots999\underline{7}12\ldots\ .$$

Fast electronic computers detract from the entertainment value of this problem because nowadays they can calculate x to 1400 sig. dec. in under a second.

Problem 3: Represented in binary, $\sqrt{2} = 1\mathbf{.}0110101000001001111001100110011111111\ldots_{Two}$. Prove that at least $\sqrt{n-1}$ of its first n binary digits are 1's for every positive integer n.

Proof 3: Apparently the claim is true for $1 \le n \le 36$. Let ß(k) count the 1's in the binary representation of an integer $k \ge 0$. For example $ß(5) = ß(101._{Two}) = 2$.

Observe that $ß(J+K) \le ß(J) + ß(K)$; it can be proved by induction on the width of the larger of J and K as follows: The observation is easy to verify for one-bit integers; suppose it is true for all integers fewer than L bits wide ($L \ge 2$), and consider two integers J and K each L bits wide (perhaps including leading 0's). If J and K are not both odd, then no "carry" propagates out of the rightmost bit of J+K, and then its leftmost L–1 bits satisfy the induction hypothesis, so $ß(J+K) \le ß(J) + ß(K)$ in this case. If $J = K = 2^L-1$ then $ß(J+K) = ß(J) = L < ß(J) + ß(K)$ in this case. In the remaining case, $J = 2j+1$ and $K = 2k+1$ are both odd and, say, $k \le 2^{L-1} - 2$;

now j , k and k+1 are L–1 bits wide (perhaps including leading 0's) and $J+K = 2(j+k+1)$, so again $ß(J+K) = ß(j+k+1) \le ß(j) + ß(k+1) \le ß(j) + ß(k) + 1 = ß(J) - 1 + ß(K) < ß(J) + ß(K)$, as was claimed.

Next observe that $ß(J\cdot K) \le ß(J)\cdot ß(K)$ because the product $J\cdot K$ is the sum of $ß(J)$ numbers each a left-shifted copy of K and each with $ß(K)$ 1's . For example $4 = ß(3\cdot 5) = ß(3)\cdot ß(5) = 2\cdot 2$.

Finally let integer $K := \lfloor 2^{n-1}\cdot\sqrt{2} \rfloor$. Evidently $ß(K)$ is the number of 1's in the first n binary digits of $\sqrt{2}$. Now, $2^{n-1}\cdot\sqrt{2} - 1 < K < 2^{n-1}\cdot\sqrt{2}$, so $\lfloor 2^{2n-1} - 2^n\cdot\sqrt{2} + 2 \rfloor \le K^2 \le 2^{2n-1} - 1$. Break binary integers into two words $[\ldots][\ldots]._{Two}$ each n bits wide to see why all integers between $\lfloor 2^{2n-1} - 2^n\cdot\sqrt{2} + 2 \rfloor = [0111\ldots1110][10010\ldots]._{Two}$ and $2^{2n-1} - 1 = [0111\ldots1111][11111\ldots]._{Two}$ inclusive turn out to have at least n–1 binary 1's , so $n-1 \le ß(K^2) \le ß(K)^2$, whence follows the problem's claim that $ß(K) \ge \sqrt{n-1}$. Actually, $ß(K) \ge \sqrt{n}$ if $n \ge 3$; do you see why?

Boris Bukh found an estimate for $ß(K)$ bigger again: $ß(K) \ge \sqrt{2n + 1/4} - 1/2$ if $n \ge 3$. …

His estimate follows from the observation that, if $K = \sum_{j \text{ in } S} 2^j$ in which S is the set of indices $j \ge 0$ corresponding to the nonzero bits of integer K , then

$$K^2 = (\textstyle\sum_{j \text{ in } S} 2^j)\cdot(\sum_{i \text{ in } S} 2^i) = \sum_{j \text{ in } S} 2^{2j} + 2\sum_{i<j \text{ in } S} 2^{i+j} .$$

The last sum is over $(ß(K)-1)\cdot ß(K)/2$ pairs of indices (i, j) both in S with $i < j$. So, if $n \ge 3$,

$$n \le ß(K^2) \le ß(\textstyle\sum_{j \text{ in } S} 2^{2j}) + ß(2\sum_{i<j \text{ in } S} 2^{i+j}) \le ß(K) + (ß(K)-1)\cdot ß(K)/2 = (ß(K)+1)\cdot ß(K)/2 .$$

From $3 \le n \le (ß(K)+1)\cdot ß(K)/2$ follows $ß(K) \ge \sqrt{2n + 1/4} - 1/2$ as Boris claimed.

David H. Bailey suggested this problem. All numerical evidence available so far suggests that randomly selected bits in $\sqrt{2}$ are as likely to be 1's as 0's , but nobody has found a way to prove this yet, and not for lack of trying.

Problem 4: Let $(n) denote the sum of the digits in the decimal representation of any integer $n \ge 0$. For example $(5678) = 26 . Explain why $(8·n) ≥ $(n)/8 .

Solution 4: Because of an inequality $(m·n) ≤ $(m)·$(n) that will be proved below, we find that $(n) = $(1000·n) = $(8·n·125) ≤ $(8·n)·$(125) = $(8·n)·8, whence $(8·n) ≥ $(n)/8 .

A preliminary inequality $(m + n) ≤ $(m) + $(n) must be confirmed first; it will be proved by induction on the number k of digits in each of n and m : The inequality is obviously true when each of n and m has just one digit; then $(m + n) = $(m) + $(n) unless the sum m + n ≥ 10 , in which case $(m + n) = $(m) + $(n) – 9 . Now suppose the inequality is true when each of n and m has k digits including perhaps leading zeros inserted to expand one integer's representation to the same width as the other's. In particular $(m + 1) ≤ $(m) + 1 . Treat a sum of two (k+1)-digit numbers as the sum of $10\cdot m + \mu$ and $10\cdot n + \nu$ in which μ and ν are single digits, and each of m and n has k digits. Of course $(10·m) = $(m). There are two cases to consider:

- If $\mu+\nu \le 9$ then, as required to propagate the induction hypothesis, …
  $\$(10\cdot m+\mu + 10\cdot n+\nu) = \$(m+n) + \mu + \nu \le \$(m) + \$(n) + \mu + \nu = \$(10\cdot m+\mu) + \$(10\cdot n+\nu)$.
- If $\mu+\nu \ge 10$ then the induction hypothesis is propagated again by …
  $$\$(10\cdot m+\mu + 10\cdot n+\nu) = \$(m+n+1) + (\mu+\nu-10)$$
  $$\le \$(m)+\$(n)-9 + \mu + \nu \ < \ \$(10\cdot m+\mu) + \$(10\cdot n+\nu) .$$

Thus $\$(m+n) \le \$(m) + \$(n)$ has been confirmed for integers of every width $k \ge 1$. Finally the inequality $\$(m\cdot n) \le \$(m)\cdot\$(n)$ is confirmed first for single digits n by repeated addition, and then for any bigger integer n by decomposing it into a sum of digits multiplied by powers of 10 .

Problem 5: Suppose that the derivative $f'(x)$ of a real function $f(x)$ of a real variable exists for all x in some interval. Prove that f' maps this interval to an interval; in other words, that $f'(x)$ takes every value between any two that it takes at arguments x in this interval.

$f'(x)$ cannot be assumed continuous throughout the interval's interior. For example take $f(x) := x^2\cdot\sin(1/x)$ except $f(0) := 0$; its $f'(0) = 0$ and elsewhere $f'(x) = 2x\cdot\sin(1/x) - \cos(1/x)$, which is discontinuous at $x = 0$ because of infinitely rapid oscillation.

Proof 5: Suppose ξ and $\zeta > \xi$ lie in the interval in question, and that $f'(\xi) < \mu < f'(\zeta)$. (The other case $f'(\xi) > \mu > f'(\zeta)$ can be handled by substituting $-f(x)$ for $f(x)$.) Now consider the minimum value taken by $F(x) := f(x) - \mu\cdot x$ over $\xi \le x \le \zeta$. Since $F'(\xi) < 0 < F'(\zeta)$, that minimum value $F(\eta)$ must be taken at some η strictly between ξ and ζ , so $F'(\eta) = 0$, which implies $f'(\eta) = \mu$. This reasoning works for every μ strictly between $f'(\xi)$ and $f'(\zeta)$.

The property of f' proved here, taking every value between any two that are taken, is called "Darboux Continuity" or "the Intermediate Value Property". Though possessed by every continuous function, it does not imply continuity. Darboux Continuity can depart from ordinary continuity only at points of infinitely rapid oscillation.

Problem 6: Even in German, words cannot be arbitrarily long; so a language can have only finitely many words, surely fewer than the things and thoughts that a competent modern language must be able to distinguish. Consequently some words accrete more than one meaning; these must be distinguished by recourse to context. In a mathematical context many words, like

compact, degenerate, dependent, differentiate, field, group, ideal, limit, point, project, proof, ring, space, …

have technical meanings different from their more common meanings on the street. Some words, like *degree, order, singular, symmetry, …,* get overworked also in mathematics. We have to sensitize ourselves to the possibility of ambiguity in order to combat it first in what we read, and then in what we write. Ambiguity can be difficult to overcome; try not to make it worse.

Here is an example taken from Physics courses on Thermodynamics: Suppose the equations $z = z(y)$ and $y = y(x)$ can be solved for $x = x(z)$ satisfying $z \equiv z(y(x(z)))$; then, as if all the d…'s cancelled, $1 \equiv dz/dz = (dz/dy)(dy/dx)(dx/dz)$. Consider next a different situation in which an equation $f(x, y, z) = 0$ can be solved for any one of the variables x, y, z as a function of the other two satisfying, for example, $f(x, y, z(x,y)) \equiv 0$; now $(\partial z/\partial y)(\partial y/\partial x)(\partial x/\partial z) = -1$. Prove this last equation and explain why the ∂…'s don't cancel the way the d…'s did.

Solution 6: Here confusion is caused by ambiguities that arise when the names of variables are used also as names of functions. The ambiguity caused no trouble in the first instance, which should have read " $z = Z(y)$ and $y = Y(z)$ can be solved for $x = X(z)$ satisfying $z \equiv Z(Y(X(z)))$ and then $1 = dz/dz = Z'(Y(X(z)))\cdot Y'(X(z))\cdot X'(z)$." This follows from the chain rule, not from cancellation of d…'s . In the second instance three functions X, Y, Z are presumed to satisfy

$$f(x, y, Z(x,y)) \equiv 0 , \quad f(X(y, z), y, z) \equiv 0 \quad \text{and} \quad f(x, Y(z, x), z) \equiv 0$$

for all x, y and z in suitable regions of (x,y,z)-space. We can obtain all those three functions' derivatives implicitly from the three partial derivatives of f , which are

$$f_x(x,y,z) := \partial f(x,y,z)/\partial x , \quad f_y(x,y,z) := \partial f(x,y,z)/\partial y , \quad \text{and} \quad f_z(x,y,z) := \partial f(x,y,z)/\partial z .$$

To get $Z_y(x, y)$ differentiate the equation $f(x, y, Z(x,y)) \equiv 0$ with respect to y to get

$$f_x(x, y, Z(x,y))\cdot(0) + f_y(x, y, Z(x,y))\cdot(1) + f_z(x, y, Z(x,y))\cdot\partial Z(x,y)/\partial y = 0 ,$$

from which $Z_y(x, y) = -f_y(x, y, Z(x,y))/f_z(x, y, Z(x,y))$ follows. Similarly

$$Y_x(z, x) = -f_x(x, Y(z, x), z)/f_y(x, Y(z, x), z) \quad \text{and} \quad X_z(y, z) = -f_z(x, y, Z(x,y))/f_x(x, y, Z(x,y)) .$$

Consequently, at every point (x, y, z) on the surface where $f(x, y, z) = 0$, which implies that all three equations $x = X(y, z)$, $y = Y(z, x)$ and $z = Z(x, y)$ are satisfied simultaneously,

$$Z_y(x, y)\cdot Y_x(z, x)\cdot X_z(y, z) = (-f_y/f_z)(-f_x/f_y)(-f_z/f_x) = -1$$

provided all derivatives in question are finite. This is what " $(\partial z/\partial y)(\partial y/\partial x)(\partial x/\partial z) = -1$ " means.

The ∂…'s don't cancel because they stand for rather different infinitesimal increments. $\partial y/\partial x$ is the slope of the tangent to the curve of intersection of the surface $f(x, y, z) = 0$ with a plane on which $z = \text{constant}$. $\partial z/\partial y$ belongs to a different curve, the intersection with a plane on which $x = \text{constant}$. Another way to see what is happening is to inspect the differential relationship $0 = df(x, y, z) = f_x\cdot dx + f_y\cdot dy + f_z\cdot dz$ valid for all infinitesimal motions on the surface $f = 0$. Keeping z constant, so $dz = 0 = f_x\cdot dx + f_y\cdot dy$, leads to $Y_x = \partial y/\partial x = -f_x/f_y$, as we saw above, but for infinitesimal increments dy and dx along a different curve than for $Z_y = \partial z/\partial y = -f_y/f_z$.

Problem 7: Mathematical jargon includes a few words of the form "…jection". Do you know what they all mean? Assemble a list of such words each obtained by *adjecting* one of the prefixes

ab-, ad-, bi-, de-, dis-, e-, in-, inter-, intro-, ob-, pro-, re-, retro-, sub-, sur-

in front of "jection", and explain each word's meaning(s), mathematical or not, succinctly.

Solution 7: ject" comes from the Latin "*iacere*", meaning "to throw". Different prefixes tell which way to throw.

abjection: a state of misery or wretchedness, or low or contemptible status. Latin: "discarded".

adjection: added attachment (thing or act), usually in front. Latin: "thrown at"

bijection: a one-to-one (invertible) mathematical mapping of all of one set onto all of another.

dejection: a state of gloom. Medical: defecation, excrement. Latin: "cast down".

disjection: dismemberment, usually violent. Latin: "thrown asunder"

ejection: state or act of expulsion, or material expelled. Latin: "thrown outwards".

injection: state or act of introducing one thing into the interior of another, or the thing injected. A one-to-one mathematical mapping of all of one set onto at least part of another. Latin: "thrown in"

interjection: an exclamation usually emotional. Latin: "thrown between"

introjection: Psychiatry: unconscious incorporation of one person's characteristics by another.

objection: opposition, or reason for it, or statement of it. Latin: "thrown against"

projection: protrusion (thing or act); plan; prediction of trend. Psychology: attribution to another of one's own thoughts. The (mathematical) image in one space of an object in or through another space of higher dimension. Latin: "thrown forth"

rejection: refusal (thing or act). Medical: immunological resistance to foreign tissue. Latin: "thrown back"

retrojection: expulsion to or from the rear (thing or act). Latin: "thrown behind"

subjection: subjugation. Latin: "thrown under"

surjection: a possibly many-to-one mathematical map of all of one set onto all of another. Latin: "thrown upon"

Many old texts use "mapping *onto*" instead of "surjection", and "one-to-one mapping *into*" instead of "injection", often letting the context determine whether an unadorned "mapping into" is one-to-one.

Problem 8: Explain why 0 is a *Cluster Point* of the infinite sequence $\{\sqrt{n}\cdot\sin(n)\}_{n\ge1}$.
If a point-set has a *Cluster Point* its every open neighborhood contains infinitely many points of the set.

Solution 8: $|\sin(n)| = |\sin(n \;\mathbf{mod}\; \pi)| = |\sin(\pi\cdot«n/\pi»)|$ where $«x» := x - (\text{the integer nearest } x)$ for every real x, so $|«x»| \le 1/2$. This means our task is to explain why $\sqrt{n}\cdot|\sin(\pi\cdot«n/\pi»)|$ becomes arbitrarily tiny infinitely often as n runs through all positive integers. Since $«x»$ will be invoked only for irrational values x, the ambiguity that arises when $«x» = \pm1/2$ because x is a half-integer will not affect our explanation. Furthermore $«x-y» = «x» - «y»$ whenever $|«x» - «y»| < 1/2$; other cases that arise when $«x-y» = «x» - «y» - \mathrm{signum}(«x» - «y»)$ will not arise during our explanation.

$\sqrt{n}\cdot|\sin(n)| = \sqrt{n}\cdot|\sin(\pi\cdot«n/\pi»)| < \sqrt{n}\cdot\pi\cdot|«n/\pi»|$ because $|\sin(x)| < |x|$ for all real $x \ne 0$. This reduces our task to explaining why $\sqrt{n}\cdot|«n/\pi»|$ is tinier than any prescribed positive threshold ε, no matter how small, for infinitely many positive integers n . Towards this end our first step is to find one such n :

Given some such tiny positive threshold $\varepsilon < 1/2$, choose any big integer $N > 1/\varepsilon^2$, and let $K(N)$ be the value of integer k that minimizes $|«k/\pi»|$ subject to the constraint $0 < k \le N$. What comes next will explain why $0 < |«K(N)/\pi»| < 1/N$. The $N+1$ fractional parts $«0», «1/\pi»$, $«2/\pi», «3/\pi», \dots$ and $«N/\pi»$ are distinct (because π is irrational) and lie strictly between $-1/2$ and $1/2$, so some two of them must differ by less than $1/N$. Let the smallest difference be $|«j/\pi» - «i/\pi»| < 1/N$ for, say, $0 \le i < j \le N$. Now $0 < k := j - i \le N$ and, thus constrained, k minimizes $|«k/\pi»|$ because $0 < |«k/\pi»| = |«j/\pi - i/\pi»| = |«j/\pi» - «i/\pi»| < 1/N < 1/2$. This $k = K(N)$, and its minimized $|«k/\pi»|$ satisfies $0 < |«K(N)/\pi»| < 1/N < \varepsilon^2$. Consequently the integer $n := K(N)$ has $0 < \sqrt{n}\cdot|«n/\pi»| = \sqrt{K(N)}\cdot|«K(N)/\pi»| < \sqrt{N}/N = 1/\sqrt{N} < \varepsilon$ as desired.

Any one such integer $n := K(N)$ spawns infinitely many more, each bigger than the one before, as follows: Choose any integer $M > 1/|«n/\pi»| = 1/|«K(N)/\pi»| > N$. The minimizing positive $K(M) \le M$ has $0 < |«K(M)/\pi»| < 1/M < |«K(N)/\pi»| = (\text{minimum of } |«k/\pi»| \text{ for } 0 < k \le N)$; consequently $K(M) > K(N)$. Thus integer $m := K(M) > n$ and, like n , satisfies the desired inequality $0 < \sqrt{m}\cdot.|«m/\pi»| < \varepsilon$. When repeated, this process generates an increasing infinite sequence of integers $n_1 := n < n_2 := m < n_3 < n_4 < \dots$ each satisfying $0 < \sqrt{n_j}\cdot|«n_j/\pi»| < \varepsilon$ starting from any tiny positive threshold ε no matter how small. Each $\sqrt{n_j}\cdot|\sin(n_j)| < \pi\cdot\varepsilon$ too.

Letting $\varepsilon \to 0$ establishes that 0 is a cluster point of the sequence $\{\sqrt{n}\cdot\sin(n)\}_{n\ge1}$.

Some writers prefer "Point of Accumulation" to "Cluster-point".