USA TSTST 2022 Solutions

United States of America — TST Selection Test

Andrew Gu and Evan Chen

64nd IMO 2023 Norway and 12th EGMO 2023 Slovenia

Contents

§1 Solutions to Day 1

§1.1 Solution to TSTST 1, by Holden Mui

In terms of n, we wish find the smallest integer k for which $(0,1)^2 \setminus S$ is always a union of k open rectangles for every set $S \subset (0,1)^2$ of size n.

We claim the answer is $k = \boxed{2n+2}$.

The lower bound is given by picking

$$S = \{(s_1, s_1), (s_2, s_2), \dots, (s_n, s_n)\}$$

for some real numbers $0 < s_1 < s_2 < \cdots < s_n < 1$. Consider the $4n$ points

$$S' = S + \{(\varepsilon, 0), (0, \varepsilon), (-\varepsilon, 0), (0, -\varepsilon)\} \subset (0,1)^2$$

for some sufficiently small $\varepsilon > 0$. The four rectangles covering each of

$$(s_1 - \varepsilon, s_1), (s_1, s_1 - \varepsilon), (s_n + \varepsilon, s_n), (s_n, s_n + \varepsilon)$$

cannot cover any other points in S'; all other rectangles can only cover at most 2 points in S', giving a bound of

$$k \geq 4 + \frac{|S'| - 4}{2} = 2n + 2.$$

To prove that $2n+2$ rectangles are sufficient, assume that the number of distinct y-coordinates is at least the number of distinct x-coordinates. Let

$$0 = x_0 < x_1 < \cdots < x_m < x_{m+1} = 1,$$

where $x_1, \dots, x_m$ are the distinct x-coordinates of points in S, and let Y_i be the set of y-coordinates of points with x-coordinate x_i. For each $1 \leq i \leq m$, include the $|Y_i| + 1$ rectangles

$$(x_{i-1}, x_{i+1}) \times ((0,1) \setminus Y_i)$$

in the union, and also include $(0, x_1) \times (0,1)$ and $(x_m, 1) \times (0,1)$; this uses $m + n + 2$ rectangles. All remaining uncovered points are between pairs of points with the same y-coordinate and adjacent x-coordinates $\{x_i, x_{i+1}\}$. There are at most $n - m$ such pairs by the initial assumption, so covering the points between each pair with

$$(x_i, x_{i+1}) \times (y - \varepsilon, y + \varepsilon)$$

for some sufficiently small $\varepsilon > 0$ gives a total of

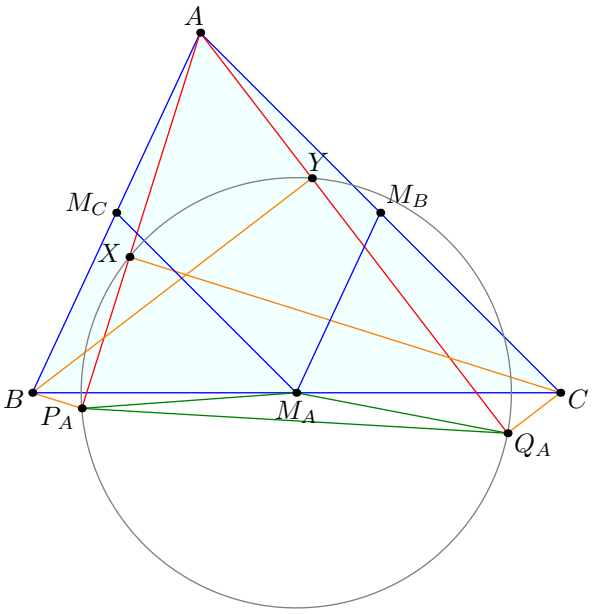
$$(m + n + 2) + (n - m) = 2n + 2$$

rectangles.

§1.2 Solution to TSTST 2, by Hongzhou Lin

We discard the points S_A and T_A since they are only there to direct the angles correctly in the problem statement.

First solution, by author Let X be the projection from C to AP_A, Y be the projection from B to AQ_A.



Claim — Line ℓ_A passes through M_A, the midpoint of BC. Also, quadrilateral P_AQ_AYX is cyclic with circumcenter M_A.

Proof. Since

$$AP_A \cdot AX = AB \cdot AC \cdot \cos\theta \cos(\angle A - \theta) = AQ_A \cdot AY,$$

it follows that P_A, Q_A, Y, X are concyclic by power of a point. Moreover, by projection, the perpendicular bisector of P_AX passes through M_A, similar for Q_AY, implying that M_A is the center of P_AQ_AYX. Hence ℓ_A passes through M_A. □

Claim — $\measuredangle(M_AM_C, \ell_A) = \measuredangle YP_AQ_A$.

Proof. Indeed, $\ell_A \perp P_AQ_A$, and $M_AM_C \perp P_AY$ (since $M_AP_A = M_AY$ from $(P_AQ_AY_AX)$ and $M_CP_A = M_CM_A = M_CY$ from the circle with diameter AB). Hence $\measuredangle(M_AM_C, \ell_A) = \measuredangle(P_AY, P_AQ_A) = \measuredangle YP_AQ_A$. □

Therefore,

$$\frac{\sin\measuredangle(M_AM_C, \ell_A)}{\sin\measuredangle(\ell_A, M_AM_B)} = \frac{\sin\measuredangle YP_AQ_A}{\sin\measuredangle P_AQ_AX} = \frac{YQ_A}{XP_A} = \frac{BC\sin(\angle C + \theta)}{BC\sin(\angle B + \theta)} = \frac{\sin(\angle C + \theta)}{\sin(\angle B + \theta)},$$

and we conclude by trig Ceva theorem.

Second solution via Jacobi, by Maxim Li Let D be the foot of the A-altitude. Note that line BC is the external angle bisector of $\angle P_ADQ_A$.

Claim — (DP_AQ_A) passes through the midpoint M_A of BC.

Proof. Perform $\sqrt{bc}$ inversion. Then the intersection of BC and (DP_AQ_A) maps to the second intersection of (ABC) and $(A'P_AQ_A)$, where A' is the antipode to A on (ABC), i.e. the center of spiral similarity from BC to P_AQ_A. Since $BP_A : CQ_A = AB : AC$, we see the center of spiral similarity is the intersection of the A-symmedian with (ABC), which is the image of M_A in the inversion. □

It follows that M_A lies on ℓ_A; we need to identify a second point. We'll use the circumcenter O_A of (DP_AQ_A). The perpendicular bisector of DP_A passes through M_C; indeed, we can easily show the angle it makes with M_CM_A is $90^\circ - \theta - C$, so $\angle O_AM_CM_A = 90 - \theta - C$, and then by analogous angle-chasing we can finish with Jacobi's theorem on $\triangle M_AM_BM_C$.

§1.3 Solution to TSTST 3, by Krit Boonsiriseth

Answer All N such that $t^2 \le N < t^2 + t$ for some positive integer t.

Solution 1 (local) If $t^2 \le N < t^2 + t$ then the sequence $s_n = tn + 1$ satisfies both conditions. It remains to show that no other values of N work.

Define $a_n \coloneqq s_n - s_{n-1}$, and let p be the minimal period of $\{a_n\}$. For each $k \in \mathbb{Z}_{\ge 0}$, let $f(k)$ be the integer such that

$$s_{f(k)} - s_k \le N < s_{f(k)+1} - s_k.$$

Note that $f(s_{n-1}) = s_n$ for all n. Since $\{a_n\}$ is periodic with period p, $f(k+p) = f(k)+p$ for all k, so $k \mapsto f(k) - k$ is periodic with period p. We also note that f is nondecreasing: if $k < k'$ but $f(k') < f(k)$ then

$$N < s_{f(k')+1} - s_{k'} < s_{f(k)} - s_k \le N,$$

which is absurd. We now claim that

$$\max_k(f(k) - k) < p + \min_k(f(k) - k).$$

Indeed, if $f(k') - k' \ge p + f(k) - k$ then we can shift k and k' so that $0 \le k - k' < p$, and it follows that $k \le k' \le f(k') < f(k)$, violating the fact that f is nondecreasing. Therefore $\max_k(f(k)-k) < p+\min_k(f(k)-k)$, so $f(k)-k$ is uniquely determined by its value modulo p. In particular, since $a_n = f(s_{n-1}) - s_{n-1}$, a_n is also uniquely determined by its value modulo p, so $\{a_n \bmod p\}$ also has minimal period p.

Now work in $\mathbb{Z}/p\mathbb{Z}$ and consider the sequence $s_0, f(s_0), f(f(s_0)), \ldots$. This sequence must be eventually periodic; suppose it has minimal period p', which must be at most p. Then, since

$$f^n(s_0) - f^{n-1}(s_0) = s_n - s_{n-1} = a_n,$$

and $\{a_n \bmod p\}$ has minimal period p, we must have $p' = p$. Therefore the directed graph G on $\mathbb{Z}/p\mathbb{Z}$ given by the edges $k \to f(k)$ is simply a p-cycle, which implies that the map $k \mapsto f(k)$ is a bijection on $\mathbb{Z}/p\mathbb{Z}$. Therefore, $f(k+1) \ne f(k)$ for all k (unless $p = 1$, but in this case the following holds anyways), hence

$$f(k) < f(k+1) < \cdots < f(k+p) = f(k) + p.$$

This implies $f(k+1) = f(k) + 1$ for all k, so $f(k) - k$ is constant, therefore $a_n = f(s_{n-1}) - s_{n-1}$ is also constant. Let $a_n \equiv t$. It follows that $t^2 \le N < t^2 + t$ as we wanted.

Solution 2 (global). Define $\{a_n\}$ and f as in the previous solution. We first show that $s_i \not\equiv s_j \pmod p$ for all $i < j < i + p$. Suppose the contrary, i.e. that $s_i \equiv s_j \pmod p$ for some i, j with $i < j < i + p$. Then $a_{s_i+k} = a_{s_j+k}$ for all $k \geq 0$, therefore $s_{s_i+k} - s_{s_i} = s_{s_j+k} - s_{s_j}$ for all $k \geq 0$, therefore

$$a_{i+1} = f(s_i) - s_i = f(s_j) - s_j = a_{j+1} \quad \text{and} \quad s_{i+1} = f(s_i) \equiv f(s_j) = s_{j+1} \pmod p.$$

Continuing this inductively, we obtain $a_{i+k} = a_{j+k}$ for all k, so $\{a_n\}$ has period $j - i < p$, which is a contradiction. Therefore $s_i \not\equiv s_j \pmod p$ for all $i < j < i + p$, and this implies that $\{s_i, \dots, s_{i+p-1}\}$ forms a complete residue system modulo p for all i. Consequently we must have $s_{i+p} \equiv s_i \pmod p$ for all i.

Let $T = s_p - s_0 = a_1 + \cdots + a_p$. Since $\{a_n\}$ is periodic with period p, and $\{i+1, \dots, i+kp\}$ contains exactly k values of each residue class modulo p,

$$s_{i+kp} - s_i = a_{i+1} + \cdots + a_{i+kp} = kT$$

for all i, k. Since $p \mid T$, it follows that $s_{s_p} - s_{s_0} = \frac{T}{p} \cdot T = \frac{T^2}{p}$. Summing up the inequalities

$$s_{s_n} - s_{s_{n-1}} \leq N < s_{s_n+1} - s_{s_{n-1}} = s_{s_n} - s_{s_{n-1}} + a_{s_n+1}$$

for $n \in \{1, \dots, p\}$ then implies

$$\frac{T^2}{p} = s_{s_p} - s_{s_0} \leq Np < \frac{T^2}{p} + a_{s_1+1} + a_{s_2+1} + \cdots + a_{s_p+1} = \frac{T^2}{p} + T,$$

where the last equality holds because $\{s_1 + 1, \dots, s_p + 1\}$ is a complete residue system modulo p. Dividing this by p yields $t^2 \leq N < t^2 + t$ for $t \coloneqq \frac{T}{p} \in \mathbb{Z}^+$.

Remark (Author comments)**.** There are some similarities between this problem and IMO 2009/3, mainly that they both involve terms of the form s_{s_n} and s_{s_n+1} and the sequence $s_0, s_1, \dots$ turns out to be an arithmetic progression. Other than this, I don't think knowing about IMO 2009/3 will be that useful on this problem, since in this problem the fact that $\{s_{n+1} - s_n\}$ is periodic is fundamentally important.

The motivation for this problem comes from the following scenario: assume we have boxes that can hold some things of total size $\leq N$, and a sequence of things of size $a_1, a_2, \dots$ (where $a_i \coloneqq s_{i+1} - s_i$). We then greedily pack the things in a sequence of boxes, 'closing' each box when it cannot fit the next thing. The number of things we put in each box gives a sequence $b_1, b_2, \dots$. This problem asks when we can have $\{a_n\} = \{b_n\}$, assuming that we start with a sequence $\{a_n\}$ that is periodic.

(Extra motivation: I first thought about this scenario when I was pasting some text repeatedly into the Notes app and noticed that the word at the end of lines are also (eventually) periodic.)

§2 Solutions to Day 2

§2.1 Solution to TSTST 4, by Merlijn Staps

Claim — If $a \mid b$ then $f(a) \mid f(b)$.

Proof. From applying the condition with $n = a$, we find that the set $S_a = \{n \geq 2 : a \mid f(n)\}$ is an arithmetic progression with common difference $f(a)$. Similarly, the set $S_b = \{n \geq 2 : b \mid f(n)\}$ is an arithmetic progression with common difference $f(b)$. From $a \mid b$ it follows that $S_b \subseteq S_a$. Because an arithmetic progression with common difference x can only be contained in an arithmetic progression with common difference y if $y \mid x$, we conclude $f(a) \mid f(b)$. □

In what follows, let $a \geq 2$ be any positive integer. Because $f(a)$ and $f(2a)$ are both divisible by $f(a)$, there are $a+1$ consecutive values of f of which at least two divisible by $f(a)$. It follows that $f(f(a)) \leq a$.

However, we also know that exactly one of $f(2)$, $f(3)$, …, $f(1+f(a))$ is divisible by a; let this be $f(t)$. Then we have $S_a = \{t, t+f(a), t+2f(a), \ldots\}$. Because $a \mid f(t) \mid f(2t)$, we know that $2t \in S_a$, so t is a multiple of $f(a)$. Because $2 \leq t \leq 1+f(a)$, and $f(a) \geq 2$ for $a \geq 2$, we conclude that we must have $t = f(a)$, so $f(f(a))$ is a multiple of a. Together with $f(f(a)) \leq a$, this yields $f(f(a)) = a$. Because $f(f(a)) = a$ also holds for $a = 1$ (from the given condition for $n = 1$ it immediately follows that $f(1) = 1$), we conclude that $f(f(a)) = a$ for all a, and hence f is a bijection.

Moreover, we now have that $f(a) \mid f(b)$ implies $f(f(a)) \mid f(f(b))$, i.e. $a \mid b$, so $a \mid b$ if and only if $f(a) \mid f(b)$. Together with the fact that f is a bijection, this implies that $f(n)$ has the same number of divisors of n. Let p be a prime. Then $f(p) = q$ must be a prime as well. If $q \neq p$, then from $f(p) \mid f(pq)$ and $f(q) \mid f(pq)$ it follows that $pq \mid f(pq)$, so $f(pq) = pq$ because $f(pq)$ and pq must have the same number of divisors. Therefore, for every prime number p we either have that $f(p) = p$ or $f(pf(p)) = pf(p)$. From here, it is easy to see that $f(n) = n$ for infinitely many n.

§2.2 Solution to TSTST 5, by Ray Li

The answer is 22. To prove the lower bound, note that there are $2022 \cdot 2021 + 2 \in (2^{21}, 2^{22})$ possible colorings. If Bob makes less than 22 queries, then he can only output 2^{21} possible colorings, which means he is wrong on some coloring.

Now we show Bob can always win in 22 queries. A key observation is that the set of Red points is convex, as is the set of Blue points, so if a set of points is all the same color, then their convex hull is all the same color.

Lemma

Let $B_0, \ldots, B_{k+1}$ be equally spaced points on a circular arc such that colors of B_0 and B_{k+1} differ and are known. Then it is possible to determine the colors of $B_1, \ldots, B_k$ in $\lceil \log_2 k \rceil$ queries.

Proof. There exists some $0 \leq i \leq k$ such that B_0, …, B_i are the same color and $B_{i+1}, \ldots, B_{k+1}$ are the same color. (If $i < j$ and B_0 and B_j were Red and B_i and B_{k+1} were Blue, then segment B_0B_j is Red and segment B_iB_{k+1} is Blue, but they intersect). Therefore we can binary search. □

Lemma

Let $B_0, \ldots, B_{k+1}$ be equally spaced points on a circular arc such that colors of B_0, $B_{\lceil k/2 \rceil}$, B_{k+1} are both Red and are known. Then at least one of the following holds: all of $B_1, \ldots, B_{\lceil k/2 \rceil}$ are Red or all of $B_{\lceil k/2 \rceil}, \ldots, B_k$ are Red. Furthermore, in one query we can determine which one of the cases holds.

Proof. The existence part holds for similar reason to previous lemma. To figure out which case, choose a point P such that all of $B_0, \ldots, B_{k+1}$ lie between rays PB_0 and $PB_{\lceil k/2 \rceil}$, and such that $B_1, \ldots, B_{\lceil k/2 \rceil - 1}$ lie inside triangle $PB_0B_{\lceil k/2 \rceil}$ and such that $B_{\lceil k/2 \rceil + 1}, \ldots, B_{k+1}$ lie outside (this point should always exist by looking around the intersections of lines $B_0B_1, B_{\lceil k/2 \rceil - 1}B_{\lceil k/2 \rceil}$). Then if P is Red, all the inside points are Red because they lie in the convex hull of Red points $P, B_0, B_{\lceil k/2 \rceil}$. If P is Blue, then all the outside points are Red: if B_i were Blue for $i > \lceil k/2 \rceil$, then the segment PB_i is Blue and intersect the segment $B_0B_{\lceil k/2 \rceil}$, which is Red, contradiction. □

Now the strategy is: Bob picks A_1. WLOG it is Red. Now suppose Bob does not know the colors of $\le 2^k - 1$ points $A_i, \ldots, A_j$ with $j - i + 1 \le 2^k - 1$ and knows the rest are Red. I claim Bob can win in $2k - 1$ queries. First, if $k = 1$, there is one point and he wins by querying the point, so the base case holds, so assume $k > 1$. Bob queries $A_{i + \lceil (j-i+1)/2 \rceil}$. If it is Blue, he finishes in $2 \log_2 \lceil (j - i + 1)/2 \rceil \le 2(k - 1)$ queries by the first lemma, for a total of $2k - 1$ queries. If it is Red, he can query one more point and learn some half of $A_i, \ldots, A_j$ that are Red by the second lemma, and then he has reduced it to the case with $\le 2^{k-1} - 1$ points in two queries, at which point we induct.

§2.3 Solution to TSTST 6, by Hongzhou Lin

First solution, by author Let $\odot OX_AY_A$ intersects AB, AC again at U, V. Then by Reim's theorem $UVCB$ are concyclic. Hence the radical axis of $\odot OX_AY_A$, $\odot OBC$ and $\odot(UVCB)$ are concurrent, i.e. OK_A, BC, UV are concurrent, Denote the intersection as K_A^*, which is indeed the inversion of $K_A$with respect to $\odot O$. (The inversion sends $\odot OBC$ to the line BC).

Let P_A, P_B, P_C be the circumcenters of $\triangle OBC, \triangle OCA, \triangle OAB$ respectively.

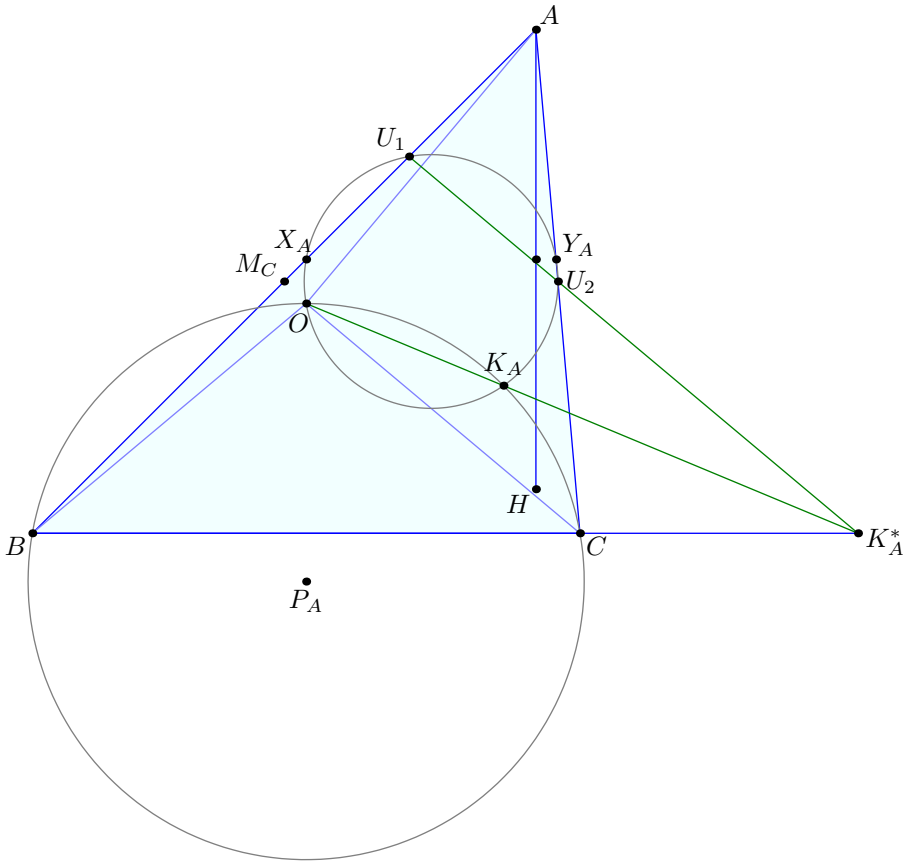
Claim — K_A^* coincides with the intersection of P_BP_C and BC.

Proof. Note that $d(O, BC) = 1/2AH = d(A, X_AY_A)$. This means the midpoint M_C of AB is equal distance to X_AY_A and the line through O parallel to BC. Together with $OM_C \perp AB$ implies that $\angle M_CX_AO = \angle B$. Hence $\angle UVO = \angle B = \angle AVU$. Similarly $\angle VUO = \angle AUV$, hence $\triangle AUV \simeq \triangle OUV$. In other words, UV is the perpendicular bisector of AO, which pass through P_B, P_C. Hence K_A^* is indeed $P_BP_C \cap BC$. □

Finally by Desargue's theorem, it suffices to show that AP_A, BP_B, CP_C are concurrent. Note that

$$d(P_A, AB) = P_AB \sin(90^\circ + \angle C - \angle A),$$
$$d(P_A, AC) = P_AC \sin(90^\circ + \angle B - \angle A).$$

Hence the symmetric product and trig Ceva finishes the proof.



Second solution, from Jeffrey Kwan Let O_A be the circumcenter of $\triangle AX_AY_A$. The key claim is that:

> **Claim —** $O_AX_AY_AO$ is cyclic.

Proof. Let DEF be the orthic triangle; we will show that $\triangle OX_AY_A \sim \triangle DEF$. Indeed, since AO and AD are isogonal, it suffices to note that

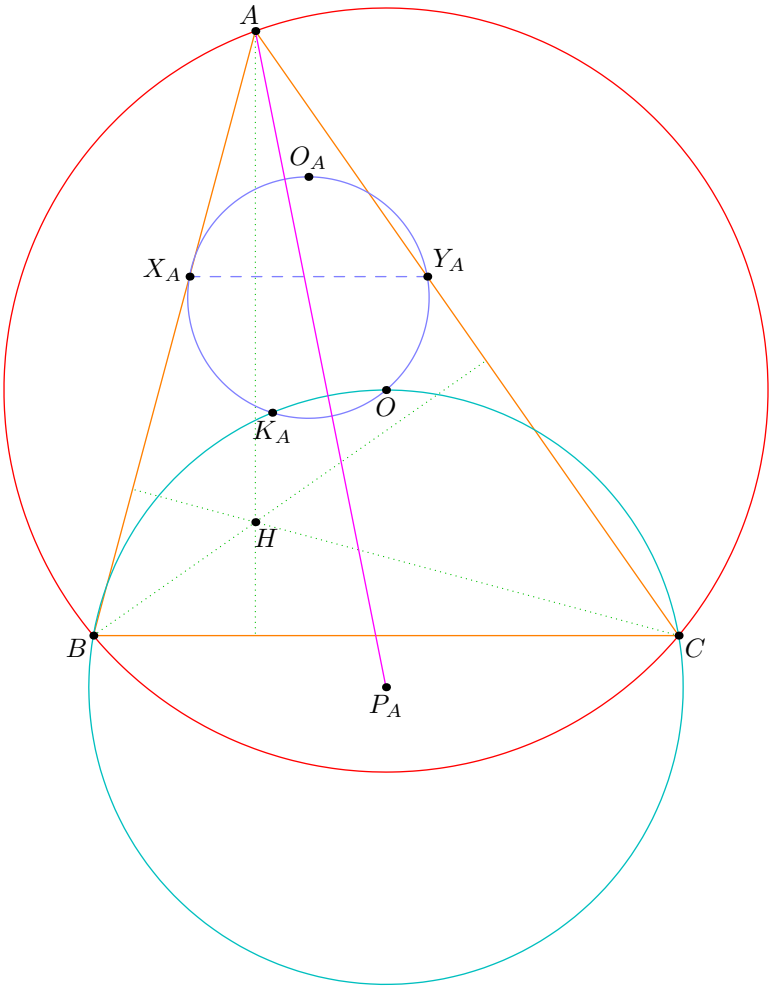
$$\frac{AX_A}{AB} = \frac{AH/2}{AD} = \frac{R\cos A}{AD},$$

and so

$$\frac{AO}{AD} = R \cdot \frac{AX_A}{AB \cdot R\cos A} = \frac{AX_A}{AE} = \frac{AY_A}{AF}.$$

Hence $\angle X_AOY_A = 180^\circ - 2\angle A = 180^\circ - \angle X_AO_AY_A$, which proves the claim. □

Let P_A be the circumcenter of $\triangle OBC$, and define P_B, P_C similarly. By the claim, A is the exsimilicenter of (OX_AY_A) and (OBC), so AP_A is the line between their two centers. In particular, AP_A is the perpendicular bisector of OK_A.



Claim — AP_A, BP_B, CP_C concur at T.

Proof. The key observation is that O is the incenter of $\triangle P_A P_B P_C$, and that A, B, C are the reflections of O across the sides of $\triangle P_A P_B P_C$. Hence $P_A A$, $P_B B$, $P_C C$ concur by Jacobi. □

Now T lies on the perpendicular bisectors of OK_A, OK_B, and OK_C. Hence $OK_A K_B K_C$ is cyclic with center T, as desired.

§3 Solutions to Day 3

§3.1 Solution to TSTST 7, by Merlijn Staps

Let the circle through A, B, and E intersect CD again at E', and let the circle through D, A, and F intersect BC again at F'. Now $ABEE'$ and $DAF'F$ are cyclic quadrilaterals with two parallel sides, so they are isosceles trapezoids. From $KA = KB$, it now follows that $KE = KE'$, whereas from $KA = KD$ it follows that $KF = KF'$.

Next, let the circle through A, B, and E intersect AC again at S. Then

$$\angle ASB = \angle AEB = \frac{1}{2}(\angle ADB + \angle ACB) = \frac{1}{2}(\angle ADB + \angle DAC) = \frac{1}{2}\angle AMB,$$

where M is the intersection of AC and BD. From $\angle ASB = \frac{1}{2}\angle AMB$, it follows that $MS = MB$, so S is the point on MC such that $MS = MB = MD$. By symmetry, the circle through A, D, and F also passes through S, and it follows that the line AS is the radical axis of the circles (ABE) and (ADF).

By power of a point, we now obtain

$$CE \cdot CE' = CS \cdot CA = CF \cdot CF',$$

from which it follows that E, F, E', and F' are concyclic. The segments EE' and FF' are not parallel, so their perpendicular bisectors only meet at one point, which is K. Hence $KE = KF$.

§3.2 Solution to TSTST 8, by Merlijn Staps

There are two families of functions that work: for each $\alpha \in \mathbb{R}$ the function $f(n) = \lfloor \alpha n \rfloor$, and for each $\alpha \in \mathbb{R}$ the function $f(n) = \lceil \alpha n \rceil - 1$. (For irrational α these two functions coincide.) It is straightforward to check that these functions indeed work; essentially, this follows from the identity

$$\left\lfloor \frac{\lfloor xn \rfloor}{n} \right\rfloor = \lfloor x \rfloor$$

which holds for all positive integers n and real numbers x.

We now show that every function that works must be of one of the above forms. Let f be a function that works, and define the sequence $a_1, a_2, \ldots$ by $a_n = f(n!)/n!$. Applying the give condition with $(n!, n+1)$ yields $a_{n+1} \in [a_n, a_n + \frac{1}{n!})$. It follows that the sequence $a_1, a_2, \ldots$ is non-decreasing and bounded from above by $a_1 + e$, so this sequence must converge to some limit α.

If there exists a k such that $a_k = \alpha$, then we have $a_\ell = \alpha$ for all $\ell > k$. For each positive integer m, there exists $\ell > k$ such that $m \mid \ell!$. Plugging in $mn = \ell!$, it then follows that

$$f(m) = \left\lfloor \frac{f(\ell!)}{\ell!/m} \right\rfloor = \lfloor \alpha m \rfloor$$

for all m, so f is of the desired form.

If there does not exist a k such that $a_k = \alpha$, we must have $a_k < \alpha$ for all k. For each positive integer m, we can now pick an ℓ such that $m \mid \ell!$ and $a_\ell = \alpha - x$ with x arbitrarily small. It then follows from plugging in $mn = \ell!$ that

$$f(m) = \left\lfloor \frac{f(\ell!)}{\ell!/m} \right\rfloor = \left\lfloor \frac{\ell!(\alpha - x)}{\ell!/m} \right\rfloor = \lfloor \alpha m - mx \rfloor .$$

If αm is an integer we can choose ℓ such that $mx < 1$, and it follows that $f(m) = \lceil \alpha m \rceil - 1$. If αm is not an integer we can choose ℓ such that $mx < \{\alpha m\}$, and it also follows that $f(m) = \lceil \alpha m \rceil - 1$. We conclude that in this case f is again of the desired form.

§3.3 Solution to TSTST 9, by Vincent Huang

For any positive integer n, define an (n, k)-good sequence to be a finite sequence of integers each between 1 and n inclusive satisfying the second property in the problem statement. The problems asks to show that, for all sufficiently large integers n, there is an (n, k)-good sequence of length at least $0.499n^2$.

Fix $k \geq 2$ and consider some prime power $n = p^m$ with $p > k + 1$. Consider some $0 < g < \frac{n}{k} - 1$ with $\gcd(g, n) = 1$ and let a be the smallest positive integer with $g^a \equiv \pm 1 \pmod{n}$.

> **Claim** (Main claim) — For k, n, g, a defined as above, there is an (n, k)-good sequence of length $a(n + 2) + 2$.

To prove the main claim, we need some results about the structure of $\mathbb{Z}/n\mathbb{Z}$. Specifically, we'll first show that any nontrivial arithmetic sequence is uniquely recoverable.

Lemma

Consider any arithmetic progression of length $i \leq k$ whose common difference is relatively prime to n, and let S be the set of residues it takes modulo n. Then there exists a unique integer $0 < d \leq \frac{n}{2}$ and a unique integer $0 \leq a < n$ such that

$$S = \{a, a + d, \ldots, a + (i - 1)d\}.$$

Proof of lemma. We'll split into cases, based on if i is odd or not.

Case 1: i is odd, so $i = 2j + 1$ for some j. Then the middle term of the arithmetic progression is the average of all residues in S, which we can uniquely identify as some u (and we know n is coprime to i, so it is possible to average the residues). We need to show that there is only one choice of d, up to $\pm$, so that $S = \{u - jd, u - (j - 1)d, \ldots, u + jd\}$.

Let X be the sum of squares of the residues in S, so we have

$$X \equiv (u - jd)^2 + (u - (j - 1)d)^2 + \cdots + (u + jd)^2 = (2j + 1)u^2 + d^2 \frac{j(j + 1)(2j + 1)}{3},$$

which therefore implies

$$3(X - (2j + 1)u^2)(j(j + 1)(2j + 1))^{-1} \equiv d^2,$$

thus identifying d uniquely up to sign as desired.

Case 2: i is even, so $i = 2j$ for some j. Once again we can compute the average u of the residues in S, and we need to show that there is only one choice of d, up to $\pm$, so that $S = \{u - (2j - 1)d, u - (2j - 3)d, \ldots, u + (2j - 1)d\}$. Once again we compute the sum of squares X of the residues in S, so that

$$X \equiv (u-(2j-1)d)^2+(u-(2j-3)d)^2+\cdots+(u+(2j-1)d)^2 = 2ju^2+\frac{(2j-1)2j(2j+1)}{3}$$

which therefore implies

$$3(X - 2ju^2)((2j - 1)2j(2j + 1))^{-1} \equiv d^2,$$

again identifying d uniquely up to sign as desired.

Thus we have shown that given the set of residues an arithmetic progression takes on modulo n, we can recover that progression up to sign. Here we have used the fact that given $d^2 \pmod{n}$, it is possible to recover d up to sign provided that n is of the form p^m with $p \neq 2$ and $\gcd(d, n) = 1$. □

Now, we will proceed by chaining many arithmetic sequences together.

Definition. For any integer l between 0 and $a-1$, inclusive, define C_l to be the sequence $0, g^l, g^l, 2g^l, 3g^l, \dots, (n-1)g^l, (n-1)g^l$ taken $\pmod n$. (This is just a sequence where the ith term is $(i-1)g^l$, except the terms $g^l, (n-1)g^l$ is repeated once.)

Definition. Consider the sequence S_n of residues mod n defined as follows:

- The first term of S_n is 0.
- For each $0 \le l < a$, the next $n+2$ terms of S_n are the terms of C_l in order.
- The next and final term of S_n is 0.

We claim that S_n constitutes a k-good string with respect to the alphabet of residues modulo n. We first make some initial observations about S_n.

Lemma

S_n has the following properties:

- S_n has length $a(n+2)+2$.
- If a contiguous subsequence of S_n of length $\le k$ contains two of the same residue $\pmod n$, those two residues occur consecutively in the subsequence.

Proof of lemma. The first property is clear since each C_l has length $n+2$, and there are a of them, along with the 0s at beginning and end.

To prove the second property, consider any contiguous subsequence $S_n[i : i+k-1]$ of length k which contains two of the same residue modulo n. If $S_n[i : i+k-1]$ is wholly contained within some C_l, it's clear that the only way $S_n[i : i+k-1]$ could repeat residues if it repeats one of the two consecutive values g^l, g^l or $(n-1)g^l, (n-1)g^l$, so assume that is not the case.

Now, it must be true that $S_n[i : i+k-1]$ consists of one contiguous subsequence of the form

$$(n-k_1)g^{l-1}, (n-(k_1-1))g^{l-1}, \dots, (n-1)g^{l-1}, (n-1)g^{l-1},$$

which are the portions of $S_n[i : i+k-1]$ contained in C_{l-1}, and then a second contiguous subsequence of the form

$$0, g^l, g^l, 2g^l, \dots, k_2 g^l,$$

which are the portions of $S_n[i : i+k-1]$ contained in C_l, and we obviously have $k_2+k_1 = k-3$. For $S_n[i : i+k-1]$ to contain two of the same residue in non-consecutive positions, there would have to exist some $0 < u \le k_1, 0 < v \le k_2$ with $(n-u)g^{l-1} \equiv vg^l \pmod n$, meaning that $u + gv \equiv 0 \pmod n$. But we know since $k_1 + k_2 < k$ that $0 < u + gv < k + kg < n$, so this is impossible, as desired. □

Now we can prove the main claim.

Proof of main claim. Consider any multiset M of $2 \le i \le k$ residues $\pmod n$ which corresponds to some unknown contiguous subsequence of S_n. We will show that it is possible to uniquely identify which contiguous subsequence M corresponds to, thereby showing that S_n has no twins of length i for each $2 \le i \le k$, and then the result will follow.

First suppose M contains some residue twice. By the last lemma there are only a few possible cases:

- M contains multiple copies of the residue 0. In this case we know M contains the beginning of S_n, so the corresponding contiguous subsequence is just the first i terms of S_n.

- M contains multiple copies of multiple residues. By the last lemma and the structure of S_n, we can easily see that M must contain two copies of $-g^{i-1}$ and two copies of g^i for some $0 \le i < a$ that can be identified uniquely, and M must contain portions of both C_{i-1}, C_i. It follows M's terms can be partitioned into two portions, the first one being
$$-i_1 g^{i-1}, -(i_1-1)g^{i-1}, \dots, -g^{i-1}, -g^{i-1},$$
and the second one being
$$0, g^i, g^i, 2g^i, \dots, i_2 g^i$$
for some i_1, i_2 with $i_1 + i_2 = i - 3$, and we just need to uniquely identify i_1, i_2. Luckily, by dividing the residues in M by g^{i-1}, we know we can partition M's terms into
$$-i_1, -(i_1-1), \dots, -1, -1$$
as well as
$$0, g, g, 2g, \dots, i_2 g.$$
Now since $i_2 g \le kg < n-k$ and $-i_1 \equiv n - i_1 \ge n - k$ it is easy to see that i_1, i_2 can be identified uniquely, as desired.

- M contains multiple copies of only one residue g^i, for some $0 \le i < a$ that can be identified uniquely. Then by the last lemma M must be located at the beginning of C_i and possibly contain the last few terms of C_{i-1}, so M must be of the form $g^i, g^i, 2g^i, \dots, i_1 g^i$, along with possibly the term 0 or the terms $0, -g^{i-1}$. So when we divide M by g^{i-1} we should be left with terms of the form $g, g, 2g, \dots, i_1 g$ along with possibly 0 or $0, -1$. Since $i_1 g \le kg < n-k$, we can easily disambiguate these cases and uniquely identify the contiguous subsequence corresponding to M.

- M contains multiple copies of only one residue $-g^i$, for some $0 \le i < a$ that can be identified uniquely. Then by the last lemma M must be located at the end of C_i and possibly the first terms of C_{i+1}, so M must be of the form $-g^i, -g^i, -2g^i, \dots, -i_1 g^i$, along with possibly the term 0 or the terms $0, g^{i+1}$. So when we divide M by g^{i-1} we should be left with terms of the form $-1, -1, -2, \dots, -i_1$, along with possibly 0 or $0, g$. Since $-i_1 \equiv n - i_1 \ge \frac{n}{2}$ and $g < \frac{n}{2}$, we can disambiguate these cases and uniquely identify the contiguous subsequence corresponding to M.

Thus in all cases where M contains a repeated residue, we can identify the unique contiguous subsequence of S_n corresponding to M.

When M does not contain a repeated residue, it follows that M cannot contain both of the g^i terms or $(n-1)g^i$ terms at the beginning or end of each C_i. It follows that M is either entirely contained in some C_i or contained in the union of the end of some C_i with the beginning of some C_{i+1}, meaning M corresponds to a contiguous subsequence of $(-g^i, 0, g^{i+1})$. In the first case, since each C_i is an arithmetic progression when the repeated terms are ignored, Lemma 1 implies that we can uniquely determine the location of M, and in the second case, it is easy to tell which contiguous subsequence of $(-g^i, 0, g^{i+1})$ corresponds to M.

Therefore, in all cases, for any multiset M corresponding to some contiguous subsequence of S_n of length $i \le k$, we can uniquely identify the contiguous subsequence, meaning S_n is k-good with respect to the alphabet of residues modulo n, as desired. $\square$

Now we will finish the problem. We observe the following.

Claim — Fix k and let $p > k+1$ be a prime. Then for $n = p^2$ we can find a (n,k)-good sequence of length $\frac{p(p-1)(p^2+2)}{2}$.

Proof of last claim. Let g be the smallest primitive root modulo $n = p^2$, so that $a = \frac{p(p-1)}{2}$. As long as we can show that $g < \frac{n}{k} - 1$, we can apply the previous claim to get the desired bound.

We will prove a stronger statement that $g < p$. Indeed, consider any primitive root $g_0 \pmod p$. Then $g_0 + ap$ has order $p-1$ modulo p, so its order modulo p^2 is divisible by $p-1$, hence $g_0 + ap$ is a primitive root modulo p^2 as long as $(g_0+ap)^{p-1} \not\equiv 1 \pmod{p^2}$. Now

$$(g_0+ap)^{p-1} = \sum_i g_0^{p-1-i}(ap)^i \binom{p-1}{i} \equiv g_0^{p-1} + g_0^{p-2}(ap) \pmod{p^2}.$$

In particular, of the values $g_0, g_0+p, \dots, g_0+p(p-1)$, only one has order $p-1$ and the rest are primitive roots.

So for each $0 < g_0 < p$ which is a primitive root modulo p, either g_0 is a primitive root modulo p^2 or g_0 has order $p-1$ but $g_0+p, g_0+2p, \dots, g_0+p(p-1)$ are all primitive roots. By considering all choices of g_0, we either find a primitive root $\pmod{p^2}$ which is between 0 and p, or we find that all residues $\pmod{p^2}$ of order $p-1$ are between 0 and p. But if $\operatorname{ord}_{p^2}(a) = p-1$ then $\operatorname{ord}_{p^2}(a^{-1}) = p-1$, and two residues between $0, p$ cannot be inverses modulo p^2 (because with the exception of 1, they cannot multiply to something $\geq p^2+1$), so there is always a primitive root between $0, p$ as desired. □

Now for arbitrarily large n we can choose $p < \sqrt{n}$ with $\frac{p}{\sqrt{n}}$ arbitrarily close to 1; by the previous claim, we can get an (n,k)-good sequence of length at least $\frac{p-1}{p} \cdot \frac{p^4}{2}$ for any constant, so for sufficiently large n, p we get (n,k)-good sequences of length $0.499n^2$.