UNM-PNM Statewide High School Mathematics Contest LVIII
Round 2 Solutions with Comments

Dear Students,
If you have suggestions about the Contest, or if you have a different solution to any of this year's second-round problems, please mail them to:

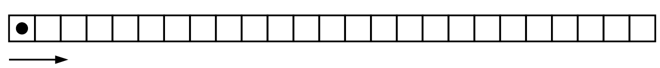
Professors James Degnan, Stephen Lau, and Cristina Pereyra
Department of Mathematics and Statistics
1 University of New Mexico
MSC01 1115
Albuquerque, NM 87131

or email them to `jamdeg@unm.edu`, `lau@unm.edu` and `crisp@unm.edu`. Please remember that you can find information about past contests at `https://mathcontest.unm.edu/`

We express our gratitude to Coach Sean Choi for running online sessions to review problems before the first and after the second rounds. We also thank Sean and Bill Cordwell for sharing their solutions with us. As you will see, we drew on students' work for many of the solutions presented below. Finally, thanks to all participants, their teachers, and families. You are an inspiration for us!

---

1. (a) Cora and Ernie are playing a new board game. The game board has 25 squares arranged in a 1×25 rectangle. They take turns moving a marker either 1 or 2 spaces forward, with Cora moving first. The marker is initially in a square at the end of the board, as shown, and must always be moved towards the other end. A player who cannot make a move that stays on the board loses. Which of the players can ensure victory?



(b) What if the game board is now 1×2026 and the marker can be moved 1, 2, or 3 spaces. Can one of the players ensure victory?

Solution 1. (a) The second player, Ernie, can always win by playing the opposite move played by the first player, Cora. If Cora moves 1, then Ernie moves 2, and if Cora moves 2, then Ernie moves 1. At the end of each Cora-then-Ernie round of play, the marker has moved 3 places, and it is Cora's turn. After 8 rounds the marker has moved $3 \times 8 = 24$ spaces and is now at square $1 + 24 = 25$, the end of the board. It is Cora's turn, but she can only move outside the board and loses the game. Therefore, Ernie has a winning strategy.

(b) The first player, Cora, has a winning strategy by first playing 1, and thereafter always playing $4-n$, where n is the number of moves played by the second player, Ernie, immediately beforehand. First, Cora plays 1, and the marker is now in square 2. So we can think of the game as starting with the marker in square 2, with Ernie making the first move. Now, if Ernie plays 1, then Cora plays 3; if Ernie plays 2, then Cora plays 2; if Ernie plays 3, then Cora plays 1. After each Ernie-then-Cora round of play the marker has moved 4 places at it is Ernie's turn. After 506 such rounds the marker has moved $4 \times 506 = 2024$ spaces, and it is now in square $2 + 2024 = 2026$, the end of the board. It is Ernie's turn, and he can only move outside the board and loses. Therefore, Cora has a winning strategy.

20	21	22	23	24	25
W	W	L	W	W	L

FIGURE 1. END OF BOARD FOR FIRST GAME.

2019	2020	2021	2022	2023	2024	2025	2026
W	W	W	L	W	W	W	L

FIGURE 2. END OF BOARD FOR SECOND GAME.

Remark 1. While the strategies described in the first solution are clear enough, how do we find them in the first place? A number of students solved the problem systematically along lines similar to the ones we describe in the next solution.

Solution 2. (a) We refer to Fig. 1 showing the end of the board for the first game. Any player whose move starts with the marker in square 25 cannot move, and so loses. The 25th square is then a losing square (and so labeled with an L). Since any player whose move starts with the marker in squares 23 and 24 can move the marker to square 25, this player can always force their opponent to start their next move with the marker in the losing 25th square. The squares 23 and 24 are winning squares (and so each is labeled with a W). Consider a player whose turn starts with the marker in square 22. They can only move the marker to squares 23 or 24, both of which are winning squares for their opponent. The 22nd square is a losing square. Our reasoning here shows that any player who starts their turn with the marker in squares 25,22,19,16,13,10,7,4,1 can be forced to lose (their opponent must always move to ensure that they start each turn with the marker in one of these squares). Since Cora starts in square 1, a losing square, Ernie can always win.

(b) Now see Fig. 2. The reasoning for this part is similar to that for part **(a)**. Now square 2026 is a losing square, but squares 2025,2024,2023 are winning squares (any player who begins their turn with the marker in one of these squares can force their opponent to start their next turn with the marker in square 2026). Therefore, we see that 2026,2022,2018,… are losing squares. Since $2026/4 = 506\frac{1}{2}$, that is $2026 \equiv 2 \pmod 4$, we see that square 2 is a losing square. Thus Cora has a winning strategy. Her initial step in this strategy is to force Ernie to start his first turn from square 2, followed by moves which ensure that Ernie always starts his turn on a losing square.

Remark 2. Some students claimed, for example, that Ernie *must* always win in the 25-square game. We have shown that Ernie, if he plays wisely, can always win. But he could also lose, were he to play heedlessly. For example, suppose initially Cora and Ernie each move the marker one square. Then each Cora-then-Ernie round advances the marker by 2 squares, at it is Cora's turn. After 11 such rounds, the marker is in square $1 + 2 \cdot 11 = 23$, and it's Cora's turn. She now moves the marker 2 squares, so that Ernie starts his next turn from the losing 25th square. In this scenario Ernie loses. Some students did correctly notice the following. If both Cora and Ernie always move 1 square (a combined round of 2 squares), or they both always move 2 squares (a combined round of 4 squares), then Ernie will win, respectively after 12 of the combined 2-square rounds and 6 of the combined 4-square rounds. Nonetheless, as the first scenario shows, Ernie can lose the 25-square game.

2. (a) Consider four cats that wish to occupy a condo. There are eight condos in a row. Each cat picks a condo at random to occupy independently of the other cats (and each condo can hold any number of cats). Find the probability that each cat chooses a different condo.

(b) Solve the problem for n cats and $2n$ condos for $n > 4$.

Solution. First recall that $k! = k \times (k-1) \times \cdots \times 2 \times 1$ (we say "k factorial") is the product of the first k positive integers. The factorial $k!$ represents the number of ways we can arrange k distinct objects in a sequence (the number of *permutations*). Factorials appear in the *binomial coefficient* (read "n choose k")

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}, \qquad n \geq k,$$

which counts the number of subsets with k distinct elements that can be chosen from n distinct elements.

(a) Think of the cats and the condos as distinct. The probability in question will arise as a fraction, with numerator equal to the number of ways the cats can arrange themselves distinctly and denominator equal to the number of ways the cats can choose condos without restriction. We turn to computation of the numerator first, observing that the

$$\begin{aligned} \text{numerator} &= \text{number of ways the cats can arrange themselves distinctly} \\ &= (\text{number of ways to choose 4 distinct condos from 8 condos}) \\ &\times (\text{number of ways 4 cats can arrange themselves in 4 distinct condos}). \end{aligned}$$

From our review in the preamble, the number of ways to choose 4 condos from 8 is $\binom{8}{4} = 70$. For each such choice, the cats can arrange themselves in $4! = 24$ ways. Therefore, there are $24 \times 70 = 1680$ ways for the cats to arrange themselves without sharing a condo, and our numerator is 1680.

Next, turn to the denominator. For unrestricted arrangement there are 8 choices for each of the four cats, or $8^4 = 4096$ ways that the cats can choose condos without restriction. Since each of these is also equally likely, the probability that they each choose a different condo is $70 \times 4!/8^4 = \frac{1680}{4096} = \frac{105}{256}$.

(b) For n cats and $2n$ condos, the same reasoning applies. There are now $\binom{2n}{n}$ ways to choose distinct condos, and $n!$ ways the cats can arrange themselves. Therefore, $\binom{2n}{n} \times n!$ is the number of ways the cats can find a distinct arrangement. However, now $(2n)^n$ is the number of possible arrangements, and the probability in question is $p = \binom{2n}{n} \times n!/(2n)^n$. Using the formula for the binomial coefficient from the preamble, this result may also be expressed as

$$p = \frac{(2n)!}{n!} \frac{1}{2^n n^n}.$$

Solution 2. Several students took the following first-principles approach. In particular, Owen Petersen, 12th grade, The ASK Academy, had a concise presentation. Here we think of that cats as choosing the condos sequentially. **(a)** The first cat to choose a condo will find an empty one with probability $1 = \frac{8}{8}$. The second cat to choose will then encounter 1 occupied condo, and 7 unoccupied ones. Therefore, the second cat will choose an unoccupied condo with probability $\frac{7}{8}$. By the same reasoning, the third cat will choose an unoccupied condo with probability $\frac{6}{8}$, and the fourth cat will choose an unoccupied condo with probability $\frac{5}{8}$. Therefore, the probability that all four cats choose unoccupied condos is the product

$$\tfrac{8}{8} \times \tfrac{7}{8} \times \tfrac{6}{8} \times \tfrac{5}{8} = \frac{7 \times 6 \times 5}{8^3} = \tfrac{210}{512} = \tfrac{105}{256}.$$

For **(b)** the reasoning is the same. The first cat chooses an unoccupied condo with probability $1 = \frac{2n}{2n}$, the second cat chooses an unoccupied condo with probability $\frac{2n-1}{2n}$, the third chooses an unoccupied condo with probability $\frac{2n-2}{2n}$, etc. Therefore, the probability that all cats will choose unoccupied condos is the product

$$p = \prod_{k=0}^{n-1} \frac{2n-k}{2n} = \frac{1}{(2n)^n} \prod_{k=0}^{n-1} (2n-k).$$

We have chosen the limits in the product based on part **(a)**, where $n = 4$ and probability that the fourth cat chooses an unoccupied condo is $\frac{5}{8}$. Alternatively, notice that the nth cat, when confronted with a scenario in which the previous $n-1$ cats have all chosen unoccupied condos, will see $2n$ condos with $2n - (n-1) = n+1$ distinct condos unoccupied by the previous $n-1$ cats. Therefore, the probability that the nth cat chooses an unoccupied condo is $\frac{n+1}{2n}$. The top limit in the above product is then correctly set as $n-1$.

To simplify the last expression, consider

$$\begin{aligned}\prod_{k=0}^{n-1} (2n-k) &= 2n \times (2n-1) \times \cdots \times (n+1) \\ &= \frac{2n \times (2n-1) \times \cdots \times (n+1) \times n \times (n-1) \times \cdots \times 2 \times 1}{n \times (n-1) \times \cdots \times 2 \times 1} \\ &= \frac{(2n)!}{n!}.\end{aligned}$$

Note that the intermediate expressions with $\cdots$ make sense with $n \geq 4$. With the last two equations, we find

$$p = \frac{(2n)!}{2^n n^n n!},$$

the same expression found before.

Remark. You might wonder how the probability p behaves in the $n \to \infty$ limit, although the limit is preposterous: we cannot have an infinite number of cats or condos! However, we might alternatively ask how p behaves for very large n, i.e. how does the expression behave *asymptotically*? To explore this question, we consider the asymptotics of the factorial provided by the famous Sterling's approximation: $k! \sim \sqrt{2\pi k} k^k e^{-k}$ as $k \to \infty$. Using this approximation, we find

$$p \sim \frac{\sqrt{4\pi n}(2n)^{2n} e^{-2n}}{\sqrt{2\pi n} n^n e^{-n}} \frac{1}{2^n n^n} = \sqrt{2} 2^n e^{-n} = \sqrt{2}(2/e)^n \to 0 \text{ as } n \to \infty.$$

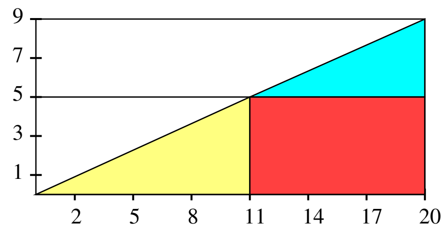
Here we have used $2 < e$. The exact expression for p cannot be used on a computer even for $n = 100$ (not so many cats), as the factorial $(2n)!$ will evaluate to `Inf` (see that last problem). However, our asymptotic approximation $\sqrt{2}(2/e)^n$ is useful well beyond 1000.

3. Consider the following figure which is drawn accurately to scale.

(a) Show that the right triangle whose vertices are $(0,0)$, $(20,0)$, and $(20,9)$ has area 90.

(b) Show that the triangle whose vertices are $(0,0)$, $(11,0)$, and $(11,5)$ has area 27.5. Show that the rectangle whose vertices are $(11,0)$, $(20,0)$, $(20,5)$, and $(11,5)$ has area 45. Finally, show that the triangle whose vertices are $(11,5)$, $(20,5)$, and $(20,9)$ has area 18. Whence the area of the shaded region is 90.5

(c) Explain why the results in **(a)** and **(b)** differ.



Solution. For **(a)** the area of the triangle with the described vertices is

$$A = \tfrac{1}{2}20 \cdot 9 = 90.$$

For **(b)** the area of the first triangle is $A_1 = \frac{1}{2}11 \cdot 5 = \frac{55}{2} = 27\frac{1}{2}$. The area of the rectangle is $A_2 = (20-11) \cdot 5 = 45$. The area of the second triangle is $A_3 = \frac{1}{2}(20-11) \cdot (9-5) = 18$. So the total area of the shaded region is $A_1 + A_2 + A_3 = 27\frac{1}{2} + 45 + 18 = 90\frac{1}{2}$. The result differs from that found in **(a)**. For **(c)** the slope of the diagonal line connecting $(0,0)$ and $(20,9)$ is

$$m = \frac{\text{rise}}{\text{run}} = \frac{9}{20}.$$

Therefore, as the diagonal passes through the origin, this line has the equation

$$y = \tfrac{9}{20}x.$$

A portion of this line is the hypotenuse of the triangle described in **(a)**. Notice that the point $(11, \frac{9}{20} \cdot 11) = (11, \frac{99}{20}) = (11, 4.95) \neq (11,5)$. So the hypotenuses of the yellow and cyan triangle do NOT together form a straight line. The shaded region is in fact NOT a right triangle, rather a quadrilateral, but it does contain the right triangle described in **(a)**.

Remark. Gavin Mitchell, 12th grade, Albuquerque Academy, noted that the composite region described in **(b)** is comprised of the right triangle in **(a)** and a second thin triangle with vertices $(0,0)$, $(20,9)$, and $(11,5)$. Take the segment connecting $(0,0)$ and $(20,9)$ as the base of this thin triangle. This base is also the hypotenuse of the right triangle from **(a)**, and it has length $\sqrt{481} = \sqrt{20^2 + 9^2}$. We know that the thin triangle must have area $\frac{1}{2}$, since this is the discrepancy between the areas for **(a)** and **(b)**. Therefore,

$$\tfrac{1}{2} = \tfrac{1}{2}\sqrt{481}h,$$

and the height of this second thin triangle must be $1/\sqrt{481}$. So this triangle as a long base and a short height. In fact, it is so thin that, to the eye, it appears as a straight line when drawn in the figure.

4. (a) How many positive integers strictly less than 2026 are multiples of 3 and 4 but not multiples of 5?

(b) How many positive integers strictly less than 2026 are multiples of 3 or 4 but not multiples of 5?

Solution 1. (a) A number is both a multiple of 3 and 4 if and only if it is a multiple of 12. The problem asks us to find the multiples of 12 that are not multiples of 5. Let us list the multiples of 12 that are strictly less than 2026:

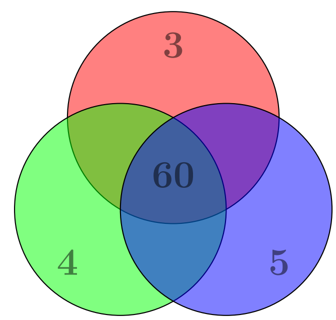
$$12, 24, 36, 48, 60, 72, \dots, 2004, 2016.$$

Note that $12 = 12 \times 1$ and $2016 = 12 \times 168$, so there are 168 positive integers that are multiples of 12 and simultaneously smaller than 2026. However, this includes the following multiples of 5 (and 12):

$$60, 120, 180, 240, \dots 1920, 1980.$$

Now, $1980 = 60 \times 33$. So there are 33 positive numbers that are multiples of 12 and 5 (equivalently, multiples of 60) which are smaller than 2026. There are $135 = 168 - 33$ numbers strictly less than 2026 that are both multiples of 3 and 4, but not multiples of 5.

To visualize the description above, consider the Venn diagram on the right showing the multiples of 12 (dark green overlap and darkest blue overlap) as the intersection of the multiples of 3 (red disk) and 4 (green disk). Here we always mean the multiples which are also strictly less than 2026. We remove the simultaneous multiples of 12 and 5 (multiples of 60, the darkest blue overlap). The leftover dark green overlap contains 135 integers, the multiples of 3 and 4 which are not multiples of 5.



(b) The multiples of 3 less than 2026 are

$$3, 6, 9, 12, 15, 18, \dots, 2013, 2016, 2019, 2022, 2025.$$

Note that $2025 = 3 \times 675$, so there are 675 multiples of 3 that are less than 2026. Similarly, the multiples of 4 less than 2026 are

$$4, 8, 12, 16, 20, \dots, 2012, 2016, 2020, 2024.$$

This time, $2024 = 4 \times 506$, so there are 506 multiples of 4 that are less than 2026. Each of these counts includes the numbers that are multiples of 3 and 4 and are strictly less than 2026, in other words the multiples of 12 less than 2026 are double counted. From part **(a)** we know there are 168 multiples of 12 that are smaller than 2026. So the number of multiples of 3 or 4 that are smaller than 2026 is $1013 = 675 + 506 - 168$.

From the count 1013, we must remove multiples of 5 and 3, and multiples of 5 and 4. However, we will then be twice removing from the count those multiples of 3, 4, and 5. So we should add them back. The multiples of 3 and 5 (multiples of 15) smaller than 2026 are

$$15, 30, 45, 60, 75, \dots, 1965, 1980, 1995, 2010, 2025.$$

Where $2025 = 15 \times 135$, so there are 135 multiples of 3 and 5 that are smaller than 2026. Likewise, the multiples of 4 and 5 (multiples of 20) smaller than 2026 are

$$20, 40, 60, 80, 100, \dots, 1960, 1980, 2000, 2020.$$

Here $2020 = 20 \times 101$, so there are 101 multiples of 4 and 5 that are smaller than 2026. Finally, the multiples of 3, 4, and 5 are the multiples of 60 that we already counted in part **(a)**. As we saw in the first part, there are 33 multiples of 60 less than 2026. Lo and behold, the number of multiples of 3 or 4 that are not multiples of 5, but are smaller than 2026, is

$$810 = 1013 - 135 - 101 + 33.$$

Solution 2. In set-theory language we have used the *inclusion-exclusion principle*: given finite sets D and E with union $D \cup E$, the number $|D \cup E|$ of elements in the union is $|D \cup E| = |D| + |E| - |D \cap E|$. Here, $|E|$ and $|D|$ are the number of elements in the finite sets E and D, respectively. By the principle of mathematical induction, this formula generalizes to any number of finite sets. For three finite sets D, E, F, the correct formula is

$$|D \cup E \cup F| = |D| + |E| + |F| - |D \cap E| - |E \cap F| - |F \cap D| + |D \cap E \cap F|.$$

Also, if $D \setminus E$ denotes the elements of D that are not in E, then $|D \setminus E| = |D| - |D \cap E|$.

Let A, B, C be the sets of multiples strictly less that 2026 of 3,4,5, respectively. Then by long division[1], we find $|A| = 675$, $|B| = 506$, and $|C| = 405$. Moreover, $A \cap B$, $A \cap C$, $B \cap C$ are then, respectively, the sets of multiples strictly less than 2026 of 12, 15, 20. Again, by long division we have $|A \cap B| = 168$, $|A \cap C| = 135$, $|B \cap C| = 101$. Finally, $A \cap B \cap C$ is the set of multiples strictly less than 2026 of 3 and 4 and 5, that is of 60. Long division gives $|A \cap B \cap C| = 33$.

For **(a)** the set of multiples of 3 and 4 is $A \cap B$, and the set of multiples of 3 and 4 that are not multiples of 5 is $(A \cap B) \setminus C$. Therefore, we examine

$$|(A \cap B) \setminus C| = |A \cap B| - |(A \cap B) \cap C| = 168 - 33 = 135.$$

For **(b)** the set of multiples of 3 or 4 is $A \cup B$ and the set of multiples of 3 or 4 that are not multiples of 5 is $(A \cup B) \setminus C$. We can compute its cardinality as follows:

$$\begin{aligned}
|(A \cup B) \setminus C| &= |A \cup B| - |(A \cup B) \cap C| \\
&= |A| + |B| - |A \cap B| - |(A \cap C) \cup (B \cap C)| \\
&= |A| + |B| - |A \cap B| - (|A \cap C| + |B \cap C| - |(A \cap C) \cap (B \cap C)|) \\
&= |A| + |B| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C| \\
&= 675 + 506 - 168 - 135 - 101 + 33 = 810.
\end{aligned}$$

Remark. From the calculation just performed,

$$|(A \cup B) \setminus C| + |C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Then with the formula for $|D \cup E \cup F|$ given in the preamble, we find

$$|(A \cup B) \setminus C| + |C| = |A \cup B \cup C|.$$

The last formula also follows from the identity $((A \cup B) \setminus C) \cup C = A \cup B \cup C$. Notice that $(A \cup B) \setminus C$ and C are disjoint sets, so that $|((A \cup B) \setminus C) \cap C| = 0$ and

$$|A \cup B \cup C| = |((A \cup B) \setminus C) \cup C| = |(A \cup B) \setminus C)| + |C|.$$

[1]Since our interest in each case are those multiples *strictly* less than 2026, we should, using floor notation, compute $|A| = \lfloor \frac{2025}{3} \rfloor$, $|B| = \lfloor \frac{2025}{4} \rfloor$, and $|C| = \lfloor \frac{2025}{5} \rfloor$. However, as Hiro Jau, 12th grade, La Cueva High School, pointed out, 2026 is not divisible by 3, 4, or 5. Thus, we may instead compute $|A| = \lfloor \frac{2026}{3} \rfloor$, $|B| = \lfloor \frac{2026}{4} \rfloor$, and $|C| = \lfloor \frac{2026}{5} \rfloor$. The same comment pertains to the other encountered long divisions.

5. Consider the following *logic table*, with entries T (True) and F (False).

p	q	$p \to q$
T	T	T
T	F	F
F	T	T
F	F	T

(a) Based on the table, what is the probability that the k-times nested statement

$$\underbrace{(((\cdots(}_{k \text{ times}} s_0 \to s_1) \to \cdots s_{k-1}) \to s_k)$$

is T? Here $k \geq 1$ and each of s_0, s_1, ..., s_k are either T or F.

(b) What is probability that the infinitely nested statement is true?

Solution 1. Recall that in the first round of the contest you were asked to consider the case $k = 2$. We now generalize to any $k \geq 1$. Note we are implicitly assuming for $k \geq 1$ that each of s_0, s_1, ..., s_k are either T or F with probability $\frac{1}{2}$.

(a) Let S_k represent the k-times nested statement, in particular $S_1 = (s_0 \to s_1)$, in which case the 2-times nested statement is $S_2 = (S_1 \to s_2) = ((s_0 \to s_1) \to s_2)$, and iterating, the $(k+1)$-times nested statement for $k \geq 1$ is $S_{k+1} = (S_k \to s_{k+1})$. We simply copy over the logic table accordingly. Further, let $P(S_k)$ be the probability that the described nested statement S_k is true, in which case $1 - P(S_k)$ is the probability that S_k is false. It follows from the table that

S_k	s_{k+1}	$S_{k+1} = (S_k \to s_{k+1})$
T	T	T
T	F	F
F	T	T
F	F	T

$$P(S_{k+1}) = \tfrac{1}{2}P(S_k) + (1 - P(S_k)) = 1 - \tfrac{1}{2}P(S_k), \text{ with } P(S_1) = \tfrac{3}{4}.$$

The statement about S_1, namely the implication $(s_0 \to s_1)$ or equivalently $(p \to q)$, follows from the original table given in the problem statement; we see that $(p \to q)$ is true with probability $\frac{3}{4}$.

The above recursion can be re-indexed as $P(S_k) = 1 - \frac{1}{2}P(S_{k-1})$, and then for $k \geq 3$ we have

$$P(S_k) = 1 - \tfrac{1}{2}P(S_{k-1}) = 1 - \tfrac{1}{2}\Big(1 - \tfrac{1}{2}P(S_{k-2})\Big) = 1 - \tfrac{1}{2} + \tfrac{1}{4}P(S_{k-2}).$$

This process can be repeated. Indeed, starting with $P(S_k)$ we can invoke the recursion $k-1$ times to conclude that

$$\begin{aligned}
P(S_k) &= 1 - \tfrac{1}{2} + \tfrac{1}{4}P(S_{k-2}) \\
&= 1 - \tfrac{1}{2} + \tfrac{1}{4} - \tfrac{1}{8}P(S_{k-3}) \\
&= 1 - \tfrac{1}{2} + \tfrac{1}{4} - \tfrac{1}{8} + \cdots + \big(-\tfrac{1}{2}\big)^{j-1} + \big(-\tfrac{1}{2}\big)^{j} P(S_{k-j}) \quad \text{for} \quad j \leq k-1 \\
&= 1 - \tfrac{1}{2} + \tfrac{1}{4} - \tfrac{1}{8} + \cdots + \big(-\tfrac{1}{2}\big)^{k-2} + \big(-\tfrac{1}{2}\big)^{k-1} P(S_1) \\
&= 1 - \tfrac{1}{2} + \tfrac{1}{4} - \tfrac{1}{8} + \cdots + \big(-\tfrac{1}{2}\big)^{k-2} + \big(-\tfrac{1}{2}\big)^{k-1} \tfrac{3}{4}.
\end{aligned}$$

To recognize the last expression as a finite geometric sum, first make the substitution $\frac{3}{4} = 1 - \frac{1}{2} + \frac{1}{4}$, in order to reach

$$\begin{aligned}
P(S_k) &= 1 - \tfrac{1}{2} + \tfrac{1}{4} - \tfrac{1}{8} + \cdots + \big(-\tfrac{1}{2}\big)^{k-2} + \big(-\tfrac{1}{2}\big)^{k-1} - \big(-\tfrac{1}{2}\big)^{k} + \big(-\tfrac{1}{2}\big)^{k+1} \\
&= \sum_{j=0}^{k+1} \big(-\tfrac{1}{2}\big)^{j}.
\end{aligned}$$

At this point, we use the familiar geometric sum formula $1 + a + a^2 + \cdots + a^n = \sum_{j=0}^{n} a^j = (1 - a^{n+1})/(1 - a)$ with $a = -\frac{1}{2}$ and $n = k + 1$, thereby finding

$$P(S_k) = \frac{1 - \big(-\frac{1}{2}\big)^{k+2}}{1 + \frac{1}{2}} = \tfrac{2}{3} - \tfrac{2}{3}\big(-\tfrac{1}{2}\big)^{k+2} = \tfrac{2}{3} + \tfrac{1}{3}\big(-\tfrac{1}{2}\big)^{k+1}.$$

Let us verify this solution by induction. First, check the base case,

$$P(S_1) = \tfrac{2}{3} + \tfrac{1}{3}(-\tfrac{1}{2})^2 = \tfrac{2}{3} + \tfrac{1}{12} = \tfrac{9}{12} \overset{\checkmark}{=} \tfrac{3}{4}.$$

Next, consider the induction step,

$$\begin{aligned} 1 - \tfrac{1}{2}P(S_k) &= \tfrac{2}{3} - \tfrac{1}{6}(-\tfrac{1}{2})^{k+1} \\ &= \tfrac{2}{3} + \tfrac{1}{3}(-\tfrac{1}{2})^{k+2} \\ &\overset{\checkmark}{=} P(S_{k+1}). \end{aligned}$$

For added confidence, check $P(S_2) = \frac{2}{3} + \frac{1}{3}(-\frac{1}{2})^3 = \frac{1}{3}(2 - \frac{1}{8}) = \frac{5}{8}$, the result from Round 1.

(b) The expression $(-\frac{1}{2})^{k+1}$ oscillates about and approaches 0 as k becomes large. This suggests that $P(S_k) \to \frac{2}{3}$ as $k \to \infty$. Indeed, notice that $\frac{2}{3}$ solves the "fixed-point equation" $P = 1 - \frac{1}{2}P$. Subtraction of the identity $\frac{2}{3} = 1 - \frac{1}{2} \cdot \frac{2}{3}$ from the recursion relation $P(S_{k+1}) = 1 - \frac{1}{2}P(S_k)$ yields

$$P(S_{k+1}) - \tfrac{2}{3} = -\tfrac{1}{2}\big(P(S_k) - \tfrac{2}{3}\big).$$

Taking absolute value on both sides, we get

$$\big|P(S_{k+1}) - \tfrac{2}{3}\big| = \tfrac{1}{2}\big|P(S_k) - \tfrac{2}{3}\big|,$$

and then by induction

$$\big|P(S_k) - \tfrac{2}{3}\big| = \big(\tfrac{1}{2}\big)^{k-1}\big|P(S_1) - \tfrac{2}{3}\big| = \big(\tfrac{1}{2}\big)^{k-1}\tfrac{1}{12} = \tfrac{1}{3}\big(\tfrac{1}{2}\big)^{k+1},$$

where we have used $P(S_1) = \frac{3}{4}$. This argument proves the assertion $P(S_k) \to \frac{2}{3}$ as $k \to \infty$, since the right-hand side $\frac{1}{3}\big(\frac{1}{2}\big)^{k+1} \to 0$ as $k \to \infty$.

Solution 2. Here we focus on **(a)**, as the solution for **(b)** is the same. The last solution found the following linear, first-order, inhomogeneous difference equation (or recursion relation).

$$P_{k+1} + \tfrac{1}{2}P_k = 1, \tag{1}$$

where P_k is an abbreviation for $P(S_k)$. This equation defines a sequence $\{P_1, P_2, P_3, \cdots\}$ of numbers. For our problem $P_1 = P(S_1) = \frac{3}{4}$, but for the moment we want to view P_1 as a free parameter, the "start value" which, once chosen, determines the sequence. The recursion (1) is *linear* because the sequence terms P_k appear neither with any powers nor in the argument of a nonlinear function. It is *first-order* since it relates only successive terms in the sequence. It is *inhomogeneous* due the 1 on the right-hand side; this term, the *inhomogeneity*, is different from the other terms; it does not involve the sequence terms. From (1) we obtain the corresponding *homogeneous equation* simply by replacing the inhomogeneity 1 by 0, giving

$$p_{k+1} + \tfrac{1}{2}p_k = 0. \tag{2}$$

Here we have also replaced P_k by p_k to emphasize that the homogeneous equation will, of course, define a different sequence, even were we to assume $p_1 = P_1$. Our solution will make use of the following principle:

The general solution to an inhomogeneous linear difference equation has the following form: the general solution to the associated homogeneous equation plus any particular solution to the inhomogeneous equation.

We shall not establish this statement, but will later confirm its correctness for our problem.

Our first task is to find a particular solution (any solving sequence) to the inhomogeneous recursion (1). As in **Solution 1**, we might notice that $P = \frac{2}{3}$ is the solution to the "fixed-point equation" $P = 1 - \frac{1}{2}P$, implying that the constant sequence $\{\frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \cdots\}$ obeys (1) and can serve as our particular solution. Without this observation about the "fixed-point" $\frac{2}{3}$, we might have started with $P(S_1) = \frac{3}{4}$, and then used the difference equation to compute $P(S_2) = 1 - \frac{3}{8} = \frac{5}{8}$, $P(S_3) = 1 - \frac{5}{16} = \frac{11}{16}$, $P(S_4) = 1 - \frac{11}{32} = \frac{21}{32}$, $P(S_5) = 1 - \frac{21}{64} = \frac{43}{64}$, and so on. Can we recognize the pattern? It looks like the limit of the sequence is $\frac{2}{3}$, and we might again consider the sequence $\{\frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \cdots\}$.

To find the general solution to the homogeneous recursion (2), make the *Ansatz*

$$p_k = \alpha r^k,$$

where α is an arbitrary constant. Substitution of this *Ansatz* into (2) yields the *root condition* $r^{k+1} = -\frac{1}{2}r^k$. Therefore, either $r = 0$ or $r = -\frac{1}{2}$. The case $r = 0$ yields the zero sequence, but this case is also captured by $\alpha = 0$. The general solution to (2) is then

$$p_k = \alpha\big(-\tfrac{1}{2}\big)^k.$$

Consequently, the general solution to (1) is

$$P_k = \tfrac{2}{3} + \alpha\big(-\tfrac{1}{2}\big)^k,$$

To get the particular solution with start condition P_1, we choose $\alpha = \frac{4}{3} - 2P_1$, so that

$$P_k = \tfrac{2}{3} + \big(\tfrac{4}{3} - 2P_1\big)\big(-\tfrac{1}{2}\big)^k. \tag{3}$$

For the logic table problem we need the initial condition $P_1 = \frac{3}{4}$, leading to

$$P_k = \tfrac{2}{3} + \tfrac{1}{3}\big(-\tfrac{1}{2}\big)^{k+1},$$

the same result found before in **Solution 1**.

Remark. We establish the following. *Claim: The general solution to the inhomogeneous recursion* (1) *is* (3)*, and so the general solution is parameterized by the start condition* P_1.

To prove the claim, assume that both $\{P_1, P_2, P_3, \cdots\}$ and $\{Q_1, Q_2, Q_3, \cdots\}$ solve (1). Then $\{P_1 - Q_1, P_2 - Q_2, P_3 - Q_3, \cdots\}$ obeys the homogeneous recursion (2), that is $P_{k+1} - Q_{k+1} = -\frac{1}{2}(P_k - Q_k)$ for $k = 1, 2, 3, \ldots$. It then follows by induction that

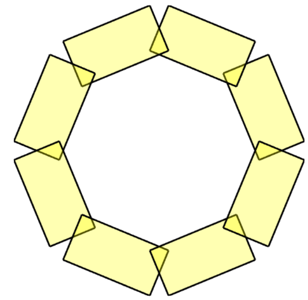
$$P_k - Q_k = \big(-\tfrac{1}{2}\big)^{k-1}(P_1 - Q_1).$$

This shows that a solution to (1) is uniquely determined by its start condition. Indeed, if $P_1 = Q_1$, then the last equation gives $P_k = Q_k$ for all $k = 1, 2, \ldots$. Moreover, if we now view $\{P_1, P_2, P_3, \cdots\}$ as any solution, but $\{Q_1, Q_2, Q_3, \cdots\}$ as the particular solution $\{\frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \cdots\}$, then the last equation becomes

$$P_k - \tfrac{2}{3} = \big(-\tfrac{1}{2}\big)^{k-1}(P_1 - \tfrac{2}{3}).$$

Rearrangement then gives (3).

6. Salvador has a paint brush of width 4 centimeters. Each thick "segment" made with the brush is in fact a rectangle of width 4 centimeters and some number of centimeters in length. Salvador paints eight legs of an "octagon", each a 4 centimeter by 10 centimeter rectangle as shown in the figure. The short sides of the rectangles have midpoints which coincide. What is the area of the genuine octagon formed inside the painted rectangles (the inside area which is not painted)?



Solution. We thank Sean Choi for the idea behind this problem. Zooming in on the intersection of two thick "segments", we recognize this intersection as a symmetric kite with two sides of length 2 and two sides of length ξ. The inner white octagon then has sides of length $\ell = 10 - 2\xi$. A regular octagon of side length ℓ has area (we can derive this in a moment)

$$A = 2(1+\sqrt{2})\ell^2.$$

The task in then to compute ξ. Each interior angle of an octagon is 135° or $\frac{3}{4}\pi$. From the diagram, we therefore see that the vertex angle of any kite between the two sides of length ξ is also $\frac{3}{4}\pi$. Each kite can be divided into two right triangles, each with side angles $\frac{1}{8}\pi$ (with opposite side ξ) and $\frac{3}{8}\pi$ (with opposite side 2). Incidentally, this shows that the vertex of each kite between the two sides of length 2 has angle 45° or $\frac{1}{4}\pi$. Therefore,

$$\xi = 2\tan(\tfrac{1}{8}\pi) = 2\sqrt{\frac{1+\cos(\frac{3}{4}\pi)}{1-\cos(\frac{3}{4}\pi)}} = 2\sqrt{\frac{\sqrt{2}-1}{\sqrt{2}+1}} = 2\sqrt{2}-2.$$

The side length of the inner octagon is therefore $10 - 4\sqrt{2} + 4 = 2(7-2\sqrt{2})$. Its area is then

$$A = 8(1+\sqrt{2})(7-2\sqrt{2})^2 = 8(1+\sqrt{2})\big(57-28\sqrt{2}\big) = \boxed{8\big(1+29\sqrt{2}\big)}.$$

To complete the solution, we derive the area formula for an octagon quoted above, reusing our result for $\tan(\frac{1}{8}\pi)$. A unit-side octagon is comprised of 8 isosceles triangles, where the central vertex has angle 45°, that is $\frac{1}{4}\pi$, and opposite side 1. See the left-hand side of Fig. 3. Therefore, its area is that of 16 right triangles, where one vertex has angle $\frac{1}{8}\pi$ and opposite side $\frac{1}{2}$. The other side then has length

$$\frac{\frac{1}{2}}{\tan(\frac{1}{8}\pi)} = \frac{\frac{1}{2}}{\sqrt{2}-1} = \tfrac{1}{2}\sqrt{2} + \tfrac{1}{2},$$

where we have used the result $\tan(\frac{1}{8}\pi) = \sqrt{2}-1$ found above. The area of one right triangle is then

$$\tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\left(\tfrac{1}{2}\sqrt{2}+\tfrac{1}{2}\right) = \tfrac{1}{8}\big(\sqrt{2}+1\big).$$

Multiplication by 16 gives $2(\sqrt{2}+1)$ for the area of an octagon with side length 1. Dimensional analysis then yields the quoted formula. See below for another derivation of this area formula.

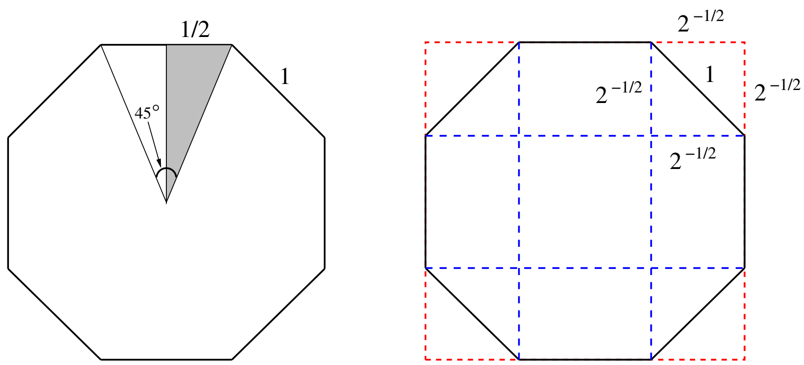


Figure 3. Octagons of unit side-length. Due to limitations of our software for producing figures, in the figure we use $2^{-1/2}$ for $\sqrt{\frac{1}{2}}$.

Remark. Here are two other strategies for computing the area of an octagon with unit side-length, referring to the right-hand side of Fig. 3. The outer red square in the right figure has area $(1 + 2 \cdot \sqrt{\frac{1}{2}})^2 = 3 + 2\sqrt{2}$. To find the area of the inscribed octagon, we subtract the areas of four triangles, each of area $\frac{1}{2} \cdot \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2}} = \frac{1}{4}$. So the area of the octagon is $3 + 2\sqrt{2} - 1 = 2(1 + \sqrt{2})$, as found earlier. Alternatively, we could view the area of the octagon as the area of four triangles (each with area $\frac{1}{4}$), four rectangles (each with area $1 \cdot \sqrt{\frac{1}{2}}$), and one square (with area 1); see the blue partitions. The total area is then $4 \cdot \frac{1}{4} + 4\sqrt{\frac{1}{2}} + 1 = 2(1 + \sqrt{2})$, as before.

Second solution. Several students made the following argument. In particular, Daniel Choi, 8th grade, Bosque School, made this argument with a precise sketch. To compute the distance ξ, refer to Fig. 4. In the figure the large right triangle formed from the yellow-shaded kite and the small grey-shaded triangle is 45°-90°-45°, with two sides of length 2. The hypotenuse then has length $2\sqrt{2}$. The smaller triangle is also 45°-90°-45°, and it has sides (the blue segments) of length $\xi = 2\sqrt{2} - 2$. The side length of the octagon is then $10 - 2\xi = 14 - 4\sqrt{2}$, as found before. To finish from here, we proceed as in the last solution.

Third solution. Argument due to Inler Lu, 9th grade, La Cueva High School. Here we compute the distance ξ described in the first solution using the equation of a line, rather than trigonometry.

Consider Fig. 5 and let the intersecting midpoint of the two rectangles (the solid dot) have coordinates $(x, y) = (0, 0)$, so that the lower, right corner of the top rectangle has coordinates $(0, -2)$. The blue dotted line which passes through the midpoint has equation $y = x$. (Actually, we need a spot of trigonometry here to convince ourselves that this line has unit slope. We can confirm this fact by verifying that the angle marked in green is 45°.) Since the distance between the midpoint and where the blue and red-dotted lines intersect is 2, we know that the intersection point of the red and blue dotted lines has coordinates $(-\sqrt{2}, -\sqrt{2})$. With these coordinates, and the observation that the red-dotted line has slope -1, we write down the equation of the red-dotted line: $y + \sqrt{2} = -(x + \sqrt{2})$ or $y = -x - 2\sqrt{2}$. The

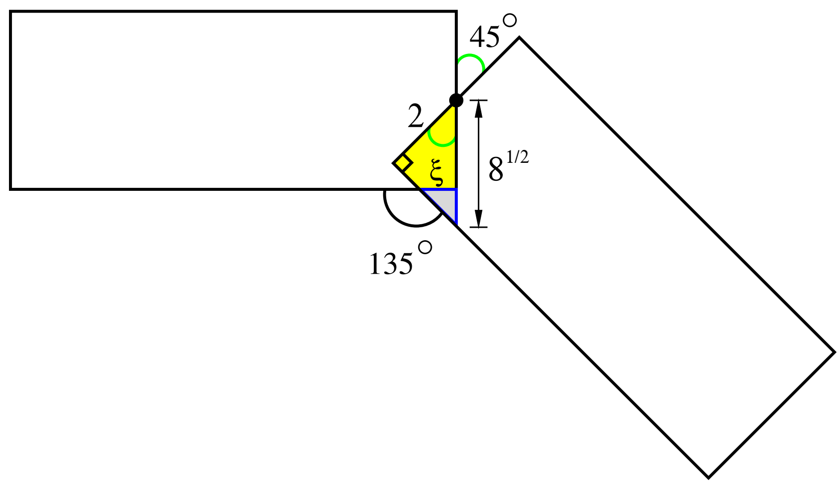


FIGURE 4. Due to the limitations of our software for making figures, here $8^{1/2}$ is used for $2\sqrt{2}$.

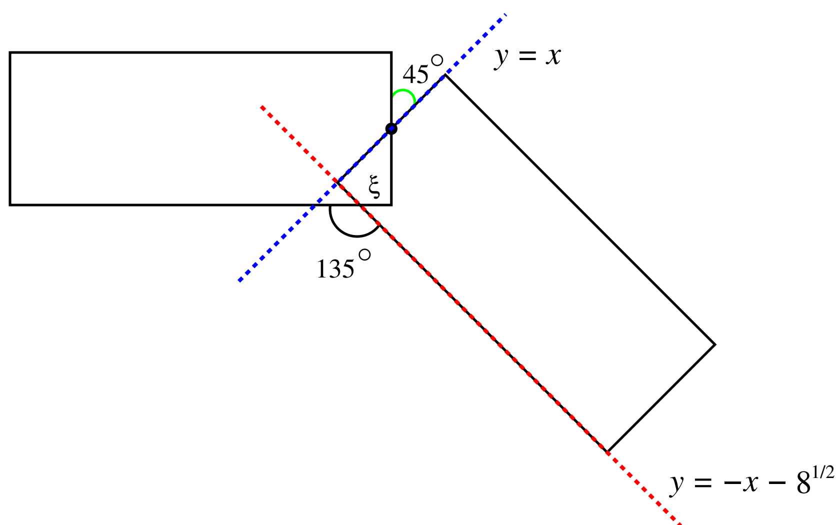


FIGURE 5. Due to the limitations of our software for making figures, here $8^{1/2}$ is used for $2\sqrt{2}$.

red-dotted line crosses the bottom of the top rectangle when $-2 = y = -x - 2\sqrt{2}$, that is at the point with coordinates $(-2(\sqrt{2} - 1), -2)$. This shows that $\xi = 2\sqrt{2} - 2$, as found in the first solution.

Fourth solution. Argument due to GRACE HSIEH, 11th grade, La Cueva High School. This approach does not require the precise area formula for an octagon, only that the formula depends on the square of the octagon's side length. We proceed as before in computing $\xi = 2\sqrt{2} - 2$. In addition to the inner octagon of interest, we also have an outer octagon whose sides have length $10 + 2\xi$. The area of this outer octagon is the area of the inner octagon plus the area of 8 trapezoids, each with height 4 and base lengths $10 - 2\xi$ and

$10 + 2\xi$. So each trapezoid has area 40, and

$$A_{\text{outer octagon}} = A_{\text{inner octagon}} + 320.$$

However, we also know by scaling that

$$\begin{aligned}
A_{\text{outer octagon}} &= \left(\frac{10 + 2\xi}{10 - 2\xi}\right)^2 A_{\text{inner octagon}} \\
&= \left(\frac{6 + 4\sqrt{2}}{14 - 4\sqrt{2}}\right)^2 A_{\text{inner octagon}} \\
&= \left(\frac{(6 + 4\sqrt{2})(14 + 4\sqrt{2})}{164}\right)^2 A_{\text{inner octagon}} \\
&= \left(\frac{116 + 80\sqrt{2}}{164}\right)^2 A_{\text{inner octagon}} \\
&= \left(\frac{29 + 20\sqrt{2}}{41}\right)^2 A_{\text{inner octagon}}
\end{aligned}$$

With the last two equations we find

$$A_{\text{inner octagon}} + 320 = \left(\frac{29 + 20\sqrt{2}}{41}\right)^2 A_{\text{inner octagon}},$$

and upon rearrangement

$$\frac{(29 + 20\sqrt{2})^2 - 41^2}{41^2} A_{\text{inner octagon}} = 320.$$

This gives that the area in question is

$$A = \frac{41^2 \cdot 320}{(29 + 20\sqrt{2})^2 - 41^2}.$$

Some work is needed to recognize this as the expression found in the first solution. In fact, the denominator in this expression reduces to $40(29\sqrt{2} - 1)$, so that

$$A = \frac{41^2 \cdot 320(1 + 29\sqrt{2})}{40(2 \cdot 29^2 - 1)} = \frac{41^2 \cdot 8(1 + 29\sqrt{2})}{(2 \cdot 29^2 - 1)}.$$

Finally, we check that $2 \cdot 29^2 - 1 = 2 \cdot 841 - 1 = 1681 = 41^2$, and so arrive at the same simple expression found earlier.

7. Suppose u and v are real numbers with $u \geq 1$, and consider the following expression.

$$f(u,v) = uv + \sqrt{(u^2-1)(v^2+1)} + \sqrt{\left(uv + \sqrt{(u^2-1)(v^2+1)}\right)^2 + 1}$$

(a) Write this expression as $f(u,v) = g(u)h(v)$, where $g(u)$ is a function solely of u and $h(v)$ is a function solely of v. That is, separate variables.

(b) Find a solution to the nonlinear system of equations

$$f(u,v) = 1, \qquad u^2 + v^2 + uv = \tfrac{7}{8}.$$

That is, find specific values of u and v which obey both equations simultaneously. Is your solution unique? *Hint*: first express the solutions to $f(u,v) = 1$ in terms of a single parameter $t = h(v)$.

Solution 1. (a) The expression $f(u,v)$ is reducible to

$$\begin{aligned}
&uv + \sqrt{(u^2-1)(v^2+1)} + \sqrt{\left(uv + \sqrt{(u^2-1)(v^2+1)}\right)^2 + 1} \\
&\quad = uv + \sqrt{(u^2-1)(v^2+1)} + \sqrt{u^2v^2 + 2uv\sqrt{(u^2-1)(v^2+1)} + (u^2-1)(v^2+1) + 1} \\
&\quad = uv + \sqrt{(u^2-1)(v^2+1)} + \sqrt{2u^2v^2 + u^2 - v^2 + 2uv\sqrt{(u^2-1)(v^2+1)}} \\
&\quad = uv + \sqrt{(u^2-1)(v^2+1)} + \sqrt{\left(u\sqrt{v^2+1} + v\sqrt{u^2-1}\right)^2} \\
&\quad = uv + \sqrt{(u^2-1)(v^2+1)} + u\sqrt{v^2+1} + v\sqrt{u^2-1} \\
&\quad = \left(v + \sqrt{v^2+1}\right)\left(u + \sqrt{u^2-1}\right).
\end{aligned}$$

To conclude that $u\sqrt{v^2+1} + v\sqrt{u^2-1}$ is positive and thereby get the second-to-last equality, we have used the following facts: (i) $\sqrt{v^2+1} > -v$ for any real v (positive or negative) and (ii) $u > \sqrt{u^2-1}$ when $u \geq 1$. Together (i) and (ii) imply that $u\sqrt{v^2+1} > -v\sqrt{u^2-1}$.

For **(b)** to achieve $\left(v + \sqrt{v^2+1}\right)\left(u + \sqrt{u^2-1}\right) = 1$, take

$$v + \sqrt{v^2+1} = t, \qquad u + \sqrt{u^2-1} = t^{-1}, \qquad t \in (0,1].$$

To find the restriction on t, notice that $v + \sqrt{v^2+1} > 0$ for $-\infty < v < \infty$, showing that $t > 0$. However, $u + \sqrt{u^2-1} \geq 1$ for $u \geq 1$, showing that $t^{-1} \geq 1$ or $t \leq 1$. From the last set of equations we readily find

$$v = \tfrac{1}{2}(t - t^{-1}), \qquad u = \tfrac{1}{2}(t + t^{-1}).$$

We now plug these parameterized expressions for u and v into the second equation, finding

$$\tfrac{7}{8} = u^2 + v^2 + uv = \tfrac{1}{4}(t^2 + 2 + t^{-2}) + \tfrac{1}{4}(t^2 - 2 + t^{-2}) + \tfrac{1}{4}(t^2 - t^{-2}) = \tfrac{3}{4}t^2 + \tfrac{1}{4}t^{-2}.$$

Multiplication of the last equation by $8t^2$ yields the equation $6t^4 - 7t^2 + 2 = 0$. From here

$$t^2 = \tfrac{1}{12}\left(7 \pm \sqrt{49-48}\right) = \tfrac{1}{12}\left(7 \pm 1\right) = \tfrac{2}{3}, \tfrac{1}{2}.$$

Since t should not be negative,

$$t = \sqrt{\tfrac{2}{3}}, \sqrt{\tfrac{1}{2}}.$$

These t-values determine values for u and v. In the first case we have

$$u = \tfrac{3}{4}\sqrt{2} \simeq 1.060660171779821, \qquad v = -\tfrac{1}{4}\sqrt{2} \simeq -0.3535533905932738.$$

In the second case we have

$$u = \tfrac{5}{12}\sqrt{6} \simeq 1.020620726159658, \qquad v = -\tfrac{1}{12}\sqrt{6} \simeq -0.2041241452319315.$$

The solution is not unique. It would be quite a feat to have found these solutions during the live test, and, without a calculator, one could not have found these decimal approximations!

Solution 2(a). Argument due to Laszlo Zolyomi, 12th grade, with *other* listed as school. Here is a quick way to get the correct separation of variables for part **(a)**. Based on the given expression for $f(u,v)$, the postulated separation of variables, and the allowed values for u, we test

$$\begin{aligned} f(1,v) &= g(1)h(v) = v + \sqrt{v^2+1} \\ f(u,0) &= g(u)h(0) = \sqrt{u^2-1} + \sqrt{(u^2-1)+1} = u + \sqrt{u^2-1}. \end{aligned}$$

We have immediately found the correct separation of variables, with $g(u) = u + \sqrt{u^2-1}$ and $h(v) = v + \sqrt{v^2+1}$, and might just move on from here to part **(b)**. Nonetheless, with time available, we might further check that

$$g(u)h(v) = uv + u\sqrt{v^2+1} + v\sqrt{u^2=1} + \sqrt{(u^2-1)(v^2+1)}.$$

To show that the right-hand side of the last equation agrees with the given expression for $f(u,v)$, follow the same calculations documented in **Solution 1**.

Solution 3(b). Argument due to Andrew Madsen, 10th grade, La Cueva High School. Here we find the solutions of the coupled system of equations stated in **(b)**, but without establishing the separation of variables requested in **(a)**. First, notice that

$$f(u,v) = w + \sqrt{w^2+1}, \quad w = uv + \sqrt{(u^2-1)(v^2+1)}.$$

Also defining $q(u,v) = w - \sqrt{w^2+1}$, we see that $f(u,v)q(u,v) = w^2-(w^2+1) = -1$. We also see that $f(u,v) - q(u,v) = 2\sqrt{w^2+1}$. When $f(u,v) = 1$, the equation $f(u,v)q(u,v) = -1$ implies $q(u,v) = -1$. Then, from $2 = 1-(-1) = f(u,v) - q(u,v) = 2\sqrt{w^2+1}$, we have $1 = \sqrt{w^2+1}$. We conclude that $w = 0$, that is $uv = -\sqrt{(u^2-1)(v^2+1)}$. Notice then that v must be negative, as we are told $u \geq 1$. Upon squaring the last equation, we arrive at $u^2v^2 = u^2v^2 + u^2 - v^2 - 1$ or $u^2 - v^2 = 1$, the equation for a hyperbola! Hmm. Since $u \geq 1$, we then find $u = \sqrt{1+v^2}$. Substitute this result into the second equation from **(b)**, thereby finding

$$1 + 2v^2 + v\sqrt{1+v^2} = \tfrac{7}{8}.$$

Rearrangement of this equation then yields

$$v\sqrt{1+v^2} = -\tfrac{1}{8} - 2v^2 \Longrightarrow v^2(1+v^2) = \tfrac{1}{64} + \tfrac{1}{2}v^2 + 4v^4 \Longrightarrow \tfrac{1}{64} - \tfrac{1}{2}v^2 + 3v^4 = 0.$$

We have then arrived at

$$192v^4 - 32v^2 + 1 = 0. \tag{4}$$

The quadratic formula determines that the roots of the last equation are

$$v^2 = \tfrac{16}{192} \pm \tfrac{1}{2\cdot 192}\sqrt{32^2 - 4\cdot 192} = \tfrac{1}{12} \pm \tfrac{1}{24} = \tfrac{1}{8}, \tfrac{1}{24}.$$

From here we recover precisely the values found above:

$$v = -\tfrac{1}{4}\sqrt{2}, -\tfrac{1}{12}\sqrt{6},$$

always choosing, as we must, the negative root. These values then determine corresponding u values. For example, if $v = -\frac{1}{12}\sqrt{6}$, then

$$u = \sqrt{1+v^2} = \sqrt{1+\tfrac{1}{24}} = \sqrt{\tfrac{25}{24}} = \tfrac{5}{12}\sqrt{6},$$

as found in the first solution. Likewise, $v = -\frac{1}{4}\sqrt{2}$ determines that $u = \frac{3}{4}\sqrt{2}$.

Remark. These comments require famililarity with logarithmic and exponential functions. Consider the logarithmic representations of the inverse hyperbolic trigonometric functions:

$$\sinh^{-1} v = \ln\left(v + \sqrt{v^2+1}\right), \qquad \cosh^{-1} u = \ln\left(u + \sqrt{u^2-1}\right), \tag{5}$$

where the domains of validity are $-\infty < v < \infty$ and $1 \le u \le \infty$. These formulas may seem exotic, but they stem directly from the definitions (in terms of the natural exponential function) of the hyperbolic trigonometric functions themselves, and these definitions[2] are, in fact, memorable. The identity established in **(a)** from **Solution 1** can be written as

$$\left(v + \sqrt{v^2+1}\right)\left(u + \sqrt{u^2-1}\right) = w + \sqrt{w^2+1}, \qquad w = uv + \sqrt{(u^2-1)(v^2+1)}.$$

The (natural) logarithm of this identity yields the following addition formula.[3]

$$\boxed{\sinh^{-1} v + \cosh^{-1} u = \sinh^{-1}\left(uv + \sqrt{(u^2-1)(v^2+1)}\right)}$$

For **(a)** we see that

$$\ln f(u,v) = \sinh^{-1}\left(uv + \sqrt{(u^2-1)(v^2+1)}\right).$$

Then we recover the claimed separation of variables upon exponentiation of the boxed equation. For **(b)** we use the boxed equation to write the logarithm of $f(u,v) = 1$ as

$$\sinh^{-1} v + \cosh^{-1} u = 0. \tag{6}$$

Figure 6 depicts graphical solution of (6). Since the arccosh term is non-negative and defined for $u \ge 1$, for the above equation to hold the arcsinh term must be non-positive with $v \le 0$. Equation (6) implicitly defines u as a function of v, but, with the formula for cosh in terms of the exponential function, we find the explicit dependence

$$u = \frac{1}{2}\left(e^{\sinh^{-1} v} + e^{-\sinh^{-1} v}\right) = \frac{1}{2}\left(v + \sqrt{v^2+1} + \frac{1}{v + \sqrt{v^2+1}}\right). \tag{7}$$

This is precisely the formula $u = \frac{1}{2}(t + t^{-1})$ found in the first solution, where $t = h(v)$.

---

[2]Namely, $\sinh x := \frac{1}{2}(e^x - e^{-x})$, and $\cosh x := \frac{1}{2}(e^x + e^{-x})$. If you solve the quadratic equation $y = \frac{1}{2}(e^x - e^{-x})$ for e^x in terms of y you will get $e^x = y + \sqrt{y^2+1}$, taking the natural logarithm on both sides, you get $x = \ln(y + \sqrt{y^2+1})$; do the same for $\cosh x$ and you will get the second equation in (5). If you know how to manipulate complex numbers and are aware of Euler's formula $e^{ix} = \cos x + i \sin x$, then you can deduce the following formulas for the ordinary sine and cosine in terms of complex exponentials: $\sin x = \frac{1}{2i}(e^{ix} - e^{-ix})$ and $\cos x = \frac{1}{2}(e^{ix} + e^{-ix})$. With this understanding, the definitions of the hyperbolic trigonometric functions are not far-fetched.

[3]Here we are using that for $a, b > 0$, $\ln(ab) = \ln a + \ln b$.

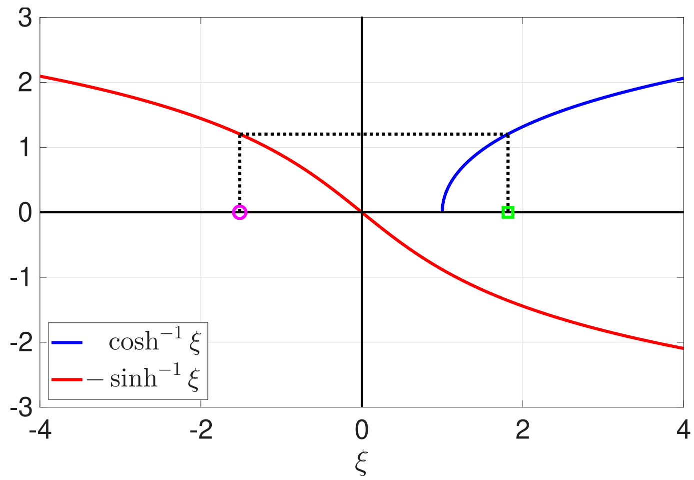


FIGURE 6. GEOMETRY OF SOLUTIONS TO (6). Each horizontal line, with ordinate value $\ln(1/t)$ for $t \in (0,1]$, defines a (u,v) pair which obeys the equation. In the figure the magenta circle corresponds to $v = \frac{1}{2}(t - t^{-1})$ and the green square to $u = \frac{1}{2}(t + t^{-1})$, both for $t = 0.3$. From the standpoint of this diagram, the parameter choice $s = 1/t$, with $s \in [1,\infty)$, would have been cleaner. Then $v = -\frac{1}{2}(s - s^{-1})$ and $u = \frac{1}{2}(s + s^{-1})$, with the graphical solutions determined by ordinate values in the range $0 \le \ln s < \infty$.

With the above functional dependence $u = u(v)$, we write the second equation $u^2 + v^2 + uv = \frac{7}{8}$ solely in terms of v. First, manipulation of the last formula yields

$$\begin{aligned} 4(u^2+v^2) &= 6v^2 + 3 + 2v\sqrt{v^2+1} + \frac{1}{2v^2+1+2v\sqrt{v^2+1}} \\ &= 6v^2 + 4 + 2v\sqrt{v^2+1} - \frac{2v^2+2v\sqrt{v^2+1}}{2v^2+1+2v\sqrt{v^2+1}} \\ &= 6v^2 + 4 + 2v\sqrt{v^2+1} - \frac{2v}{v+\sqrt{v^2+1}}. \end{aligned}$$

Second, with (7) we find

$$4uv = 2v^2 + 2v\sqrt{v^2+1} + \frac{2v}{v+\sqrt{v^2+1}}.$$

Addition of the last two results yields

$$4(u^2+v^2+uv) = 8v^2 + 4 + 4v\sqrt{v^2+1}.$$

The equation $8(u^2+v^2+uv) - 7 = 0$ then becomes

$$16v^2 + 8v\sqrt{v^2+1} + 1 = 0.$$

With further algebra, we manipulate the last equation into

$$192v^4 - 32v^2 + 1 = 0,$$

precisely the equation (4) encountered above. From here, the solution continues the same way as described before in **Solution 3(b)**.

8. MATLAB (software invented at UNM by Cleve Moler) represents a computer number as

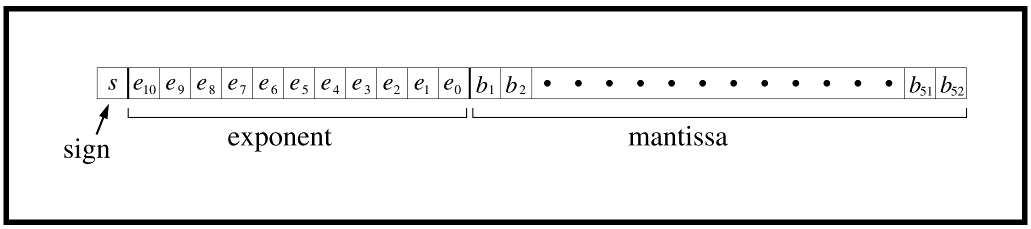
$$x = (-1)^s (1.b_1 b_2 b_3 \cdots b_{52})_2 \times 2^{F-1023}.$$

This representation corresponds to a 64-bit string in computer memory, as shown in the figure. Each "bit" is either a 0 or 1. The *sign* bit s determines whether the number is positive or negative. The *exponent* F is stored as an 11-bit string $e_{10}e_9e_8e_7e_6e_5e_4e_3e_2e_1e_0$, and corresponds to the number

$$F = (e_{10}e_9 \cdots e_0)_2 = e_{10}2^{10} + e_9 2^9 + \cdots + e_0 2^0.$$

The *true exponent* $F - 1023$ differs from F by an *exponential bias* of 1023. The *mantissa* is a 52-bit string $b_1 b_2 \cdots b_{52}$ which defines the number

$$(1.b_1 b_2 \cdots b_{52})_2 = 1 + \frac{b_1}{2} + \frac{b_2}{2^2} + \cdots + \frac{b_{52}}{2^{52}}.$$



(a) What is the allowed range for F (its smallest and largest possible values)?

(b) What is the allowed range for $(1.b_1 b_2 \cdots b_{52})_2$? Give compact answers.

(c) Suppose that x is stored in computer memory as the following string, with two zeros among 62 ones.

0111 1111 1110 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111

Give a compact expression for x.

Solution. For **(a)** the smallest F corresponds to eleven 0-bits, that is $0 = (00000000000)_2$. The largest F corresponds to eleven 1-bits, that is $2047 = (11111111111)_2$. We know that eleven 1-bits gives 2047, because adding $1 = 2^0$ to $2^{10} + 2^9 + \cdots + 2^2 + 2^1 + 2^0$ yields

$$\begin{aligned}\left(2^{10} + 2^9 + \cdots + 2^2 + 2^1 + 2^0\right) + 1 &= (\overbrace{11111111111}^{\text{eleven ones}})_2 + 1 \\ &= (1\underbrace{00000000000}_{\text{eleven zeros}})_2 \\ &= 2^{11},\end{aligned}$$

and $2^{11} = 2048 = 2047 + 1$. So the allowed range is all integers F obeying $0 \le F \le 2047$. For **(b)** if all bits in the mantissa are 0, then

$$(1.\underbrace{000 \cdots 00}_{\text{52 zeros}})_2 = 1.$$

If all bits in the mantissa are 1, then

$$(1.\underbrace{111\cdots 11}_{52\text{ ones}})_2 = \sum_{n=0}^{52} 2^{-n}.$$

The last summation is arguably a compact expression, but can simplify further with the summation formula for a finite geometric series. Indeed, we then have

$$(1.\underbrace{111\cdots 11}_{52\text{ ones}})_2 = = \frac{1-2^{-53}}{1-\frac{1}{2}} = 2 - 2^{-52}.$$

Therefore, the allowed range is $1 \leq (1.b_1b_2\cdots b_{52})_2 \leq 2 - 2^{-52}$. This answer is sufficient, but realize that $(1.b_1b_2\cdots b_{52})_2$ cannot take all real-number values between 1 and $2 - 2^{-52}$; the computer number system is not a continuum! Rather, $(1.b_1b_2\cdots b_{52})_2$ can take the values $1 + k2^{-52}$ for $k = 0, 1, \ldots, 2^{52} - 1$. For **(c)** the sign bit is 0; this string defines a positive x. The exponent F is determined by the string 11111111110. But

$$(11111111110)_2 = (11111111111)_2 - (00000000001)_2 = 2047 - 1 = 2046.$$

The true exponent is then $2046 - 1023 = 1023$, and using the result from **(b)** we see that

$$x = (2 - 2^{-52})2^{1023} = (1 - 2^{-53})2^{1024}.$$

The last expression is a fine final form, but we might also report the following.

$$\boxed{x = (2^{53} - 1)2^{971}}$$

Remark 1. The described representation for a *normalized double precision number* is not unique to MATLAB, but also used in `C`, `Fortran`, `C++`, `Python`, etc. The exceptional values $F = 0, 2047$ for the exponent are reserved for special purposes. $F = 2047$ corresponds to either $\pm$`Inf`, provided the mantissa is 52 zeros, or `NaN` (Not a Number), provided the mantissa has a 1 as any bit. $F = 0$ is reserved for *unnormalized numbers* and corresponds to a change in the representation (notice the absence of the lead 1):

$$x_{\text{unnorm}} = \pm(0.b_1b_2\cdots b_{52})_2 \times 2^{-1022}.$$

Change of the representation for unnormalized numbers allows for more numbers crammed near 0. The number in **(c)** is actually the largest normalized double precision number, and so the largest double precision number represented on the computer (unless, that is, you would like to view `Inf` or `NaN` as larger still). In fact, $x \simeq 1.7977 \times 10^{308}$ which is quite large. The smallest nonzero double (an unnormalized number) is $x_{\text{unnorm}} = 2^{-1074} \simeq 4.9407 \times 10^{-324}$.

Remark 2. For each exponent value in the range $1 \leq F \leq 2046 = 2^{11} - 2$ and choice of sign bit, there are 2^{52} possible mantissa values. Therefore, $2 \cdot (2^{11} - 2) \cdot 2^{52} = 2^{64} - 2^{54}$ is the total number of normalized double precision numbers. To this count, we add 2 for $\pm$`Inf`, realized for $F = 2047$, $s = 0$ or 1, and a mantissa which is 52 zeros. We add $2 \cdot (2^{52} - 1)$ for the possible ways to get a `NaN`, corresponding to $F = 2047$, $s = 0$ or 1, and any mantissa, except the one with 52 zeros. At this point we have $2^{64} - 2^{54} + 2^{53}$ computer numbers. But let us not forget the unnormalized numbers; these are $2 \cdot 2^{52}$ (the possible sign times the possible mantissa values) in number. So that total count for "double precision numbers" is

$$2^{64} - 2^{54} + 2^{53} + 2^{53} = 2^{64} - 2^{54} + 2 \cdot 2^{53} = 2^{64},$$

exactly what we should expect, since the storage format is a 64-bit binary string.