Three Distance Theorem: Liang's Proof

Nitin Verma
mathsanew.com

October 7, 2021

Let α be any positive real number with $\alpha < 1$, and N any positive integer. For any real number x, let $\{x\}$ denote the fractional part of x, i.e. $\{x\} = x - \lfloor x \rfloor$. Consider the points $\{0\alpha\}(= 0), \{1\alpha\}, \{2\alpha\}, \{3\alpha\}, \ldots, \{N\alpha\}$ in interval $[0, 1]$. For simplicity, we will assume in this discussion that all these points are distinct; similar reasoning will apply if they start repeating after $\{i\alpha\}$ for some i. These points will divide the interval $[0, 1]$ in $N + 1$ non-empty intervals, which we will call "Gaps".

The basic statement of *Three Distance Theorem* (also known as *Three Gap Theorem* and *Steinhaus Conjecture*) is as follows. Please refer the article titled *Three Distance Theorem* [1] for more insights into this theorem.

Theorem 1 (Three Distance Theorem)**.** *The gaps created (as described above) have only upto three distinct lengths.*

F. M. Liang gave a simple and elegant proof for a generic version of this theorem [2]. In this article, we will see an elaboration of that proof for the theorem presented above.

Any region (not necessarily a gap) with endpoints $\{i\alpha\}$ and $\{j\alpha\}$, $\{i\alpha\} < \{j\alpha\}$, will be denoted as (i, j) (i can be 0). A region from some endpoint $\{i\alpha\}$ to endpoint 1 will be denoted by a special notation $(i, 0)$, because a region ending at 1 can be seen as wrapping around and ending at 0.

If a region contains some point $\{k\alpha\}$, excluding its own endpoints, we will say that the region "encloses" that point.

Note that a gap is also one of the regions just described, which does not enclose any point. So we will use the same notation to denote a gap.

Reducing a Gap

Consider some gap (i, j) with $i > 0$ and $j > 0$. We already know $\{i\alpha\} < \{j\alpha\}$. The two points $p_1 = \{(i-1)\alpha\}$ and $p_2 = \{(j-1)\alpha\}$ are simply a left-shift of $\{i\alpha\}$ and $\{j\alpha\}$ by amount α, wrapping around 0 if necessary. If both or none of the two points wrap around 0, we see that they will form a region from p_1 to p_2 having the same length as gap (i, j). This region is $(i-1, j-1)$.

We can prove that it is impossible to have only $\{i\alpha\}$ wrap around 0 with $p_2 > 0$. Because then, 0 will be a point within this "wrapped region" from p_1 to p_2. That means, right-shift of p_1, 0 and p_2 by α will give us gap (i, j) but with point $\{\alpha\}$ enclosed by the gap. That is impossible since a gap cannot enclose a point.

Though, we still can have only $\{i\alpha\}$ wrap around 0 with $p_2 = 0$. This is possible only with $j = 1$. In this case, we will define the region formed by p_1 and p_2 to be from endpoint p_1 to 1, which is denoted by $(i-1, 0)$. Since $j = 1$, it is $(i-1, j-1)$.

Thus, we can always associate a well-defined region with points p_1 and p_2, which is precisely the region $(i-1, j-1)$ and is of the same length as gap (i, j).

So, given any gap (i, j) with $i > 0$ and $j > 0$, we can obtain the region $(i-1, j-1)$ from it. We will say, the gap (i, j) was "reduced" to the region $(i-1, j-1)$.

The obtained region $(i-1, j-1)$ may enclose a point and so need not be a gap. Suppose it encloses a point $\{k\alpha\}$ with $k < N$. But that means, $\{(k+1)\alpha\}$, which is a right-shift of $\{k\alpha\}$ by α, must be enclosed by gap (i, j). That is impossible since a gap cannot enclose a point. So we must have $k = N$. We conclude that, the region $(i-1, j-1)$ encloses either no point (so, is a gap) or exactly one point $\{N\alpha\}$.

Further, since there is only one final point $\{N\alpha\}$, there must be exactly one region (f_1, f_2) which encloses this point and no other point.

Thus, the region $(i-1, j-1)$ must be either a gap or the unique region (f_1, f_2).

Proof

Given any gap (a, b) (a or b may be 0), we can repetitively reduce it to obtain regions (a, b), $(a-1, b-1)$, $(a-2, b-2)$ etc. until either of these two happen: (1) the obtained region (a', b') encloses a point, (2) the obtained region (a', b') has a' or b' equaling 0.

Upon termination, we will map the initial gap (a, b) to the final obtained region (a', b').

If (1) holds upon termination, region (a', b') must be the region (f_1, f_2).

Otherwise ((2) holds, (1) doesn't hold), the region (a', b') must be a gap, either $(0, b - a)$ (if $b > a$), or $(a - b, 0)$ (if $a > b$). Note that gap $(a - b, 0)$ represents the gap ending at 1. We know that there is exactly one gap $(0, s)$ starting at 0, and exactly one gap $(e, 0)$ ending at 1. So, in this case, the region (a', b') must be either $(0, s)$ or $(e, 0)$.

Since reducing a gap doesn't change its length, the above process will end up mapping any gap to a gap/region of the same length. But we found that, any gap can map to only one of these three: the region (f_1, f_2), the gap $(0, s)$, the gap $(e, 0)$. Thus, there can be only upto three distinct gap lengths.

Relation Between a Gap's Endpoints

Suppose a gap (a, b) maps to $(0, s)$. Then, we must have $b > a$ and $b - a = s$. Similarly, if the gap maps to $(e, 0)$, we must have $a > b$ and $a - b = e$.

Notice that the region (f_1, f_2) encloses a single point $\{N\alpha\}$. So, it consists of two adjacent gaps (f_1, N) and (N, f_2). But each of these two gaps must map to either $(0, s)$ or $(e, 0)$ (they can't map to region (f_1, f_2)). Since gap (f_1, N) has $f_1 < N$, it must map to $(0, s)$. Similarly, gap (N, f_2) must map to $(e, 0)$. Thus we must have: $N - f_1 = s$ and $N - f_2 = e$. Hence, $f_2 - f_1 = s - e$.

If a gap (a, b) maps to region (f_1, f_2), we must have $b - a = f_2 - f_1$ (both a and b were decremented by same amount to give f_1 and f_2). So due to above, $b - a = s - e$.

In summary, each gap (a, b) must have the difference $b - a$ one among: s, $-e$, $s - e$.

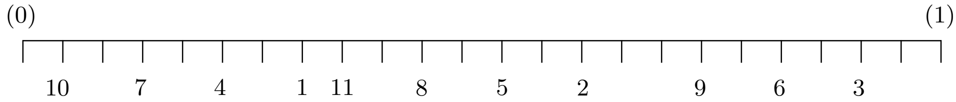
Also note that the gap-length corresponding to region (f_1, f_2) is simply the sum of lengths of the gaps $(0, s)$ and $(e, 0)$.

Theorem 1 and 2 in article [1] include the same results we just proved above; s and e are referred as u_1 and u_N there, and the three gap lengths as L_1, L_2 and $L_1 + L_2$.

Example

The below figure shows the example of $\alpha = 7/23$ and $N = 11$. All multiples of $1/23$ in interval $[0, 1]$ have been marked. Numbers $1, 2, 3, \ldots$ below the interval line $[0, 1]$ indicate points $\{\alpha\}, \{2\alpha\}, \{3\alpha\}, \ldots$.

Figure 1: $\alpha = 7/23$ and $N = 11$



Below is shown the repetitive reducing of all gaps, where each arrow indicates a reduce operation. Note that all reduce sequences are terminating at gaps $(0, 10)$ or $(3, 0)$, or region $(1, 8)$.

$(1, 11) \rightarrow (0, 10)$

$(11, 8) \rightarrow (10, 7) \rightarrow (9, 6) \rightarrow (8, 5) \rightarrow (7, 4) \rightarrow (6, 3) \rightarrow (5, 2) \rightarrow (4, 1) \rightarrow (3, 0)$

$(2, 9) \rightarrow (1, 8)$

■

References

[1] Nitin Verma. *Three Distance Theorem.*
`https://mathsanew.com/articles/three_distance_theorem.pdf`.

[2] F. M. Liang. *A Short Proof of the 3d Distance Theorem.* Discrete Math., Vol 28 (3) (1979), 325–326.
`https://doi.org/10.1016/0012-365X(79)90140-7`.