

ELEMENTARY APPROACHES TO RIEMANN INTEGRABILITY FOR DISCONTINUOUS FUNCTIONS

SEOUNGWAN SONG

ABSTRACT. In this paper, we investigate elementary approaches to Riemann integrability through the structure of the set of discontinuities: functions whose sets of discontinuities are finite, countably infinite, and certain uncountable sets. In each case, we establish Riemann integrability directly through Darboux sums, without appealing to the Lebesgue criterion. In the uncountable case, we focus on the Cantor set and extend the resulting argument to a broader class of sets having Cantor behaviour. These results provide an elementary illustration of the classical Lebesgue criterion for Riemann integrability.

1. INTRODUCTION

In our previous work [3], we considered a special case of Riemann integrability, particularly functions whose derivatives are bounded on a closed interval. We obtained the following main result:

Theorem. *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous on $[a, b]$. If f is differentiable on (a, b) and f' is bounded on (a, b) , then f is Riemann integrable on $[a, b]$.*

This theorem indeed provides a sufficient condition for Riemann integrability. However, it does not characterize all Riemann integrable functions. More precisely, differentiability is not necessary, and moreover the presence of discontinuities does not preclude Riemann integrability. One classical result in integration theory, namely Lebesgue's criterion for Riemann integrability [1] gives a complete characterization of bounded Riemann integrable functions on a closed interval: a bounded function on a closed interval is Riemann integrable if and only if its set of discontinuities has measure zero. The main purpose of this paper is to study this idea through several cases. We first consider functions with finitely many discontinuities, followed by functions with countably infinite set of discontinuities. In both cases, we aim to derive Riemann integrability directly using elementary arguments based on Darboux sums, without appealing to Lebesgue's criterion. Then our two main theorems are as follows:

Theorem 3.4. *Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded. If f has finitely many discontinuities, then f is Riemann integrable on $[a, b]$.*

Theorem 4.4. *Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded. If f has countably infinitely many discontinuities, then f is Riemann integrable on $[a, b]$.*

Although the preceding theorems concern finite and countably infinite sets of discontinuities, they do not cover all possible cases. In particular, a set may be uncountable while still having measure zero. This leads to the final part of this

paper, where we examine the Cantor set as a representative example. The resulting argument is then extended to a broader class of sets having a similar behaviour.

2. SETUP

The proofs of our main theorems require several common definitions and notation. This section is devoted to establishing the necessary background for the arguments that follow.

Definition 2.1. Let $a < b$ be real numbers. A finite set of points $P = \{x_0, x_1, \dots, x_n\}$ that satisfy $a = x_0 < \dots < x_n = b$ is called a *partition* of $[a, b]$.

Definition 2.2. Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded, and let P be partition of $[a, b]$. For each subinterval $[x_{i-1}, x_i]$, define $M_i = \sup\{f(x) : x \in [x_{i-1}, x_i]\}$ and $m_i = \inf\{f(x) : x \in [x_{i-1}, x_i]\}$. Then the *upper Darboux sum* and the *lower Darboux sum* of f with respect to P are defined by

$$U(f, P) = \sum_{i=1}^n M_i \Delta x_i, \quad L(f, P) = \sum_{i=1}^n m_i \Delta x_i$$

respectively, where $\Delta x_i = x_i - x_{i-1}$.

Definition 2.3. Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded. The function f is called *Riemann integrable* on $[a, b]$ if

$$\inf\{U(f, P) : P \text{ is partition of } [a, b]\} = \sup\{L(f, P) : P \text{ is partition of } [a, b]\}.$$

Equivalently, f is *Riemann integrable* on $[a, b]$ if, for every $\epsilon > 0$, there exist partition P of $[a, b]$ such that

$$U(f, P) - L(f, P) < \epsilon.$$

Note that, by substituting the expressions for the upper and lower Darboux sums from Definition 2.2, this condition can also be written as

$$\sum_{i=1}^n (M_i - m_i) \Delta x_i < \epsilon.$$

Definition 2.4. Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded. The set of discontinuities of f on $[a, b]$ is denoted by $D_f = \{x \in [a, b] : f \text{ is discontinuous at } x\}$.

Then the case where f is continuous, has finitely many discontinuities, or has countably infinitely many discontinuities on $[a, b]$ can be expressed as $D_f = \emptyset$, $|D_f| < \infty$, or $|D_f| = \aleph_0$, respectively.

3. FUNCTIONS WITH FINITELY MANY DISCONTINUITIES

In this section, we consider bounded functions having only finitely many points of discontinuity. We begin with the continuous case, which provides the first step of our main theorem. It is important to note that in order to apply the Darboux criterion, we must control the oscillation of the function on every subinterval of a partition. Hence, uniform continuity plays an essential role here. Appealing Bolzano-Weierstrass theorem, we initially prove that every continuous function on $[a, b]$ is uniformly continuous. We then use this result to show that the difference between the upper and lower Darboux sums can be made arbitrarily small. We begin by recalling the following classical theorem.

Theorem 3.1 (Bolzano–Weierstrass Theorem). *Every bounded sequence contains a convergent subsequence.*

Lemma 3.2. *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is uniformly continuous on $[a, b]$.*

Proof. Let us assume the contrary, that is, f is not uniformly continuous. Then there exist $\epsilon > 0$ such that for every $\delta > 0$, there exist points $x, y \in [a, b]$ satisfying $|x - y| < \delta$ but $|f(x) - f(y)| \geq \epsilon$. We now define sequence $(x_n), (y_n) \subseteq [a, b]$ such that for each $n \in \mathbb{N}$, $|x_n - y_n| < 1/n$ and $|f(x_n) - f(y_n)| \geq \epsilon$. Notice that (x_n) lies inside $[a, b]$, that means, (x_n) is bounded. Hence by Bolzano–Weierstrass Theorem, there exist convergent subsequence (x_{n_k}) converging to some $c \in [a, b]$. We then have $|x_{n_k} - y_{n_k}| < 1/n_k$, where $1/n_k \rightarrow 0$. It follows from triangle inequality that $|y_{n_k} - c| \leq |y_{n_k} - x_{n_k}| + |x_{n_k} - c| \rightarrow 0$, which gives $y_{n_k} \rightarrow c$. Considering that f is continuous at c , we then have $f(x_{n_k}) \rightarrow f(c)$ and $f(y_{n_k}) \rightarrow f(c)$. And consequently, $|f(x_{n_k}) - f(y_{n_k})| \rightarrow 0$. However, our assumption tells for every k , $|f(x_{n_k}) - f(y_{n_k})| \geq \epsilon$, a contradiction. Therefore, we conclude that f is uniformly continuous on $[a, b]$. $\square$

The uniform continuity obtained in the preceding result enables us to bound the oscillation of the continuous function on arbitrarily small subintervals. This immediately yields the following theorem.

Theorem 3.3. *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is Riemann integrable on $[a, b]$.*

Proof. We start by fixing $\epsilon > 0$. Lemma 3.2. tells us that f is uniformly continuous on $[a, b]$. Hence, for every $\epsilon > 0$, there exist $\delta > 0$ such that, for all $x, y \in [a, b]$, if $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon/(b - a)$. Denote $P = \{x_0, x_1, \dots, x_n\}$ as a partition of $[a, b]$ whose mesh is less than δ . We then take any subintervals $[x_{i-1}, x_i]$, and define M_i and m_i as supremum and infimum of f on $[x_{i-1}, x_i]$, respectively. Considering the length of interval is less than δ , uniform continuity induces that $M_i - m_i < \epsilon/(b - a)$. Thus, we now have:

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{i=1}^n (M_i - m_i)(x_i - x_{i-1}) \\ &< \frac{\epsilon}{b - a} \sum_{i=1}^n (x_i - x_{i-1}) \\ &= \frac{\epsilon}{b - a} \cdot (b - a) = \epsilon. \end{aligned}$$

Showing that $U(f, P) - L(f, P) < \epsilon$. Therefore, we conclude that f is Riemann integrable on $[a, b]$ $\square$

We now turn to the main result of this section. We generalize the preceding theorem to bounded functions with finitely many discontinuities by induction on the number of discontinuities. The central idea is that the finiteness of D_f enables us to isolate one discontinuity inside an interval of arbitrarily small length, allowing us to control the difference between the upper and lower Darboux sums in that interval. On the other hand, the remaining intervals contain fewer discontinuities, and hence the induction hypothesis can be applied. The following proof explains this.

Theorem 3.4. *Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded. If f has finitely many discontinuities, then f is Riemann integrable on $[a, b]$.*

Proof. Let $P(n)$ be the statement that every bounded function on $[a, b]$ with at most n discontinuities is Riemann integrable. Notice that a function with no discontinuities is continuous. Hence, our base case $P(0)$ follows from Theorem 3.3. We now suppose $P(n)$ holds for arbitrary $n \geq 0$, and let $f : [a, b] \rightarrow \mathbb{R}$ be a bounded function with at most $n + 1$ discontinuities. Observe that if f has at most n discontinuities, then the result follows immediately from our induction hypothesis. Therefore, it is enough to consider the case in which f has exactly $n + 1$ discontinuities.

Let $\epsilon > 0$ be given. Since f is bounded, there exist $M > 0$ such that $|f(x)| \leq M$ for every $x \in [a, b]$. Then it follows that the oscillation of f on any subinterval of $[a, b]$ is at most $2M$. We next choose one discontinuity c . Considering that c either lies in the interior of $[a, b]$ or is one of its endpoints, the proof can be now separate into following two cases.

Case 1. $c \in (a, b)$. Given that c lies in the interior of $[a, b]$, we can define interval $[a', b'] \subset [a, b]$ that contains no discontinuity of f other than c , and have length less than $\epsilon/(4M)$. Then $[a, b]$ now can be expressed as union of following three intervals: $[a, a']$, $[a', b']$, and $[b', b]$. Notice that the restrictions of f to $[a, a']$ and $[b', b]$ each have at most n discontinuities. Hence our induction hypothesis implies that both restrictions are Riemann integrable, and therefore there exist partitions P_a of $[a, a']$ and P_b of $[b', b]$ satisfying $U(f, P_a) - L(f, P_a) < \epsilon/4$ and $U(f, P_b) - L(f, P_b) < \epsilon/4$.

We now move our attention to the interval $[a', b']$. Let M_0 and m_0 denote the supremum and infimum of f on $[a', b']$. We know that the oscillation of f on any subinterval of $[a, b]$ is at most $2M$. Hence, it induces that $M_0 - m_0 \leq 2M$. Therefore, the difference between the lower and upper sum is bounded above by $(M_0 - m_0)(b' - a') < 2M \cdot \epsilon/(4M) = \epsilon/2$.

Let P be the partition of $[a, b]$ obtained by combining P_a and P_b , and a' and b' . Then by using the result found above, we now have:

$$\begin{aligned} U(f, P) - L(f, P) &= U(f, P_a) - L(f, P_a) + (M_0 - m_0)(b' - a') \\ &\quad + U(f, P_b) - L(f, P_b) \\ &< \frac{\epsilon}{4} + \frac{\epsilon}{2} + \frac{\epsilon}{4} = \epsilon. \end{aligned}$$

Showing that f is Riemann integrable on $[a, b]$ in this case.

Case 2. $c \in \{a, b\}$. Since the proof remains the same for two endpoints, we only consider the case where $c = a$. Let $b' \in (a, b)$ such that $[a, b']$ contains no discontinuity of f other than a , and $b' - a < \epsilon/(4M)$. We know that on the remaining interval $[b', b]$, f has n discontinuities. Hence by our induction hypothesis, it gives us partition P_b such that $U(f, P_b) - L(f, P_b) < \epsilon/2$. Also, by using the same estimate as in Case 1, we know that error contributed by the interval $[a, b']$ is less than $\epsilon/2$. Together, these give us a partition P of $[a, b]$ obtained by adjoining a to P_b , such that $U(f, P) - L(f, P) < \epsilon/2 + \epsilon/2 = \epsilon$. Thus, f is Riemann integrable on $[a, b]$ when $c = a$, and the case $c = b$ follows in the same manner.

Considering both two cases, we now have $P(n + 1)$ holds. This proves our main theorem of this section. $\square$

We have now established the result for bounded functions with finitely many points of discontinuity. This naturally raises the question of whether our finiteness assumption can be weakened to countability. In the next section, we answer this question in the affirmative.

4. FUNCTIONS WITH COUNTABLY MANY DISCONTINUITIES

Herein, we turn our attention to bounded functions whose sets of discontinuities are countably infinite. Unlike in our previous section, we cannot directly apply the induction argument. Instead, following the approach based on tagged partitions presented by Gordon [2], we define a gauge differently at continuity and discontinuity points. With this approach in mind, we first show that a countable set can be covered by open intervals whose total length is arbitrarily small, and recall the notions of tagged partitions and gauges, following with Cousin's theorem.

Lemma 4.1. *Let $G = \{a_1, a_2, \dots\} \subseteq [a, b]$ be countable. Then for every $\epsilon > 0$, there exists a set of open intervals $\{I_1, I_2, \dots\}$ such that $G \subseteq \bigcup_{i=1}^{\infty} I_i$ and $\sum_{i=1}^{\infty} \ell(I_i) < \epsilon$, where $\ell(I_i)$ denotes the length of I_i .*

Proof. Let $\epsilon > 0$ be arbitrary. For each $i \in \mathbb{N}$, we cover $a_i \in G$ with a open interval I_i with length less than $\epsilon/2^i$. Then by sum of geometric series, it follows that the total length of the open intervals is less than $\sum_{i=1}^{\infty} \ell(I_i) < \sum_{i=1}^{\infty} \epsilon/2^i = \epsilon$. This completes our lemma. $\square$

Remark. Lemma 4.1 shows that every countable subset of $[a, b]$ has arbitrarily small total length, in other words, measure zero. Therefore, Lebesgue's criterion for Riemann integrability immediately implies that every bounded function whose set of discontinuities is countable is Riemann integrable. Nevertheless, we give a direct proof based on the following well-known result of Cousin

Definition 4.2. Let $P = \{x_0, x_1, \dots, x_n\}$ be a partition of $[a, b]$. For each subinterval $[x_{i-1}, x_i]$, choose one point $t_i \in [x_{i-1}, x_i]$. Then the point t_i is called the *tag* of the subinterval $[x_{i-1}, x_i]$, and the collection of pairs $\{(t_i, [x_{i-1}, x_i]) : i = 1, \dots, n\}$ is called a *tagged partition* of $[a, b]$. Moreover, a tagged partition is said to be δ -fine if $[x_{i-1}, x_i] \subseteq (t_i - \delta(t_i), t_i + \delta(t_i))$ for all $i = 1, \dots, n$, where $\delta : [a, b] \rightarrow \mathbb{R}^+$ is a positive function called *gauge*.

Theorem 4.3 (Cousin's theorem). *Let $\delta : [a, b] \rightarrow \mathbb{R}^+$ be gauge on $[a, b]$. Then there exists a δ -fine tagged partition of $[a, b]$.*

The preceding results now allows us to address the main conclusion of this section. It is important to note that the difference between the upper and lower Darboux sums depends on both the oscillation of the function and the length of the corresponding subintervals. That means, it is unnecessary to make both quantities small simultaneously. Hence, we separate the tagged subintervals based on whether their tags are continuity or discontinuity points. For the tags which are continuity points, the gauge is chosen so that the oscillation of the function is small. For the remaining subintervals where the function can fluctuate, the gauge is chosen so that the corresponding subintervals lie inside a set of open intervals of arbitrarily small total length that can be obtain by Lemma 4.1. As a result, we can control the two types of subintervals in different ways, which yields the following theorem.

Theorem 4.4. *Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded. If f has countably infinitely many discontinuities, then f is Riemann integrable on $[a, b]$.*

Proof. Let $\epsilon > 0$ be given. Since $|D_f| = \aleph_0$, we can write $D_f = \{d_1, d_2, \dots\}$. Also, considering that f is bounded, there exist $M > 0$ such that $|f(x)| \leq M$ for every $x \in [a, b]$. Consequently, the oscillation of f on any subinterval of $[a, b]$ is at most $2M$. We now use Lemma 4.1 to cover the discontinuities by open intervals whose total length is arbitrarily small. More precisely, for each $i \in \mathbb{N}$, choose an open interval I_i have length less than $\epsilon/(2^{i+2}M)$. Then $\{I_1, I_2, \dots\}$ covers D_f , and by the sum of geometric series, the total length is less than $\sum_{i=1}^{\infty} \ell(I_i) < \sum_{i=1}^{\infty} \epsilon/(2^{i+2}M) = \epsilon/(4M)$.

We next define a gauge $\delta : [a, b] \rightarrow \mathbb{R}^+$ according to whether a given point is a continuity or a discontinuity point of f as following:

$$\delta(x) = \begin{cases} \delta_j(x), & \text{if } x = d_j \in D_f \\ \delta_c(x), & \text{if } x \notin D_f, \end{cases}$$

where, if $x = d_j \in D_f$, we set $\delta_j(x) > 0$ sufficiently small so that satisfies:

$$(x - \delta_j(x), x + \delta_j(x)) \subseteq I_j.$$

On the other hand, if $x \notin D_f$, then f is continuous at x . Hence, we can choose $\delta_c(x) > 0$ such that satisfies:

$$|x - y| < \delta_c(x) \implies |f(x) - f(y)| < \frac{\epsilon}{4(b-a)} \quad \text{for every } y \in [a, b].$$

With gauge δ over all $[a, b]$, we are now ready to apply Cousin's theorem. Then, we can obtain δ -fine tagged partition $P = \{x_0, x_1, \dots, x_n\}$, with tags $t_i \in [x_{i-1}, x_i]$ such that $[x_{i-1}, x_i] \subseteq (t_i - \delta(t_i), t_i + \delta(t_i))$ for every i . Let M_i and m_i denote the supremum and infimum of f on $[x_{i-1}, x_i]$, respectively. Considering that t_i is either a point of continuity or a point of discontinuity of f , the purpose of this choice of gauge now becomes clear. We can therefore separate the proof into following two cases.

Case 1. $t \notin D_f$. We know that f is continuous at t_i , and hence $\delta(t_i) = \delta_c(t_i)$. Let $\alpha, \beta \in [x_{i-1}, x_i]$ be arbitrary. As the partition is δ -fine, $[x_{i-1}, x_i] \subseteq (t_i - \delta(t_i), t_i + \delta(t_i))$, and it follows that $|\alpha - t_i| < \delta_c(t_i)$ and $|\beta - t_i| < \delta_c(t_i)$. Therefore by definition δ_c , we obtain:

$$|\alpha - t_i| < \delta_c(t_i) \implies |f(\alpha) - f(t_i)| < \frac{\epsilon}{4(b-a)}, \text{ and}$$

$$|\beta - t_i| < \delta_c(t_i) \implies |f(\beta) - f(t_i)| < \frac{\epsilon}{4(b-a)}.$$

Then it follows from triangle inequality that, $|f(\alpha) - f(\beta)| \leq |f(\alpha) - f(t_i)| + |f(t_i) - f(\beta)| < \epsilon/(4(b-a)) + \epsilon/(4(b-a)) = \epsilon/(2(b-a))$. Since α and β was arbitrary, it means that the oscillation of f on the interval is less than $\epsilon/(2(b-a))$, in other words, $M_i - m_i < \epsilon/(2(b-a))$. Finally, we now can find the contribution from all subintervals whose tags are continuity points. Considering that the total length is at most $b - a$, we have:

$$\begin{aligned} \sum_{t_i \notin D_f} (M_i - m_i)(x_i - x_{i-1}) &< \frac{\epsilon}{2(b-a)} \sum_{t_i \notin D_f} (x_i - x_{i-1}) \\ &\leq \frac{\epsilon}{2(b-a)} \cdot (b-a) = \frac{\epsilon}{2}. \end{aligned}$$

Case 2. $t \in D_f$. We know that f is discontinuous at t_i . Hence, $t_i = d_j$ for some $j \in \mathbb{N}$. Also, by the choice of δ_j , it induces that $(d_j - \delta_j(t_i), d_j + \delta_j(t_i)) \subseteq I_j$. Bearing in mind that P is a δ -fine tagged partition, we know that:

$$\begin{aligned} [x_{i-1}, x_i] &\subseteq (t_i - \delta(t_i), t_i + \delta(t_i)) \\ &= (d_j - \delta_j(t_i), d_j + \delta_j(t_i)) \subseteq I_j. \end{aligned}$$

Showing that $[x_{i-1}, x_i] \subseteq I_j$. That means, every subinterval whose tag lies in D_f is contained in one of the intervals I_j . Moreover, considering the partition subintervals do not overlap except at endpoints, their total length satisfies $\sum_{t_i \in D_f} (x_i - x_{i-1}) \leq \sum_{i=1}^{\infty} \ell(I_i) < \epsilon/(4M)$. On the other hand, recall that the oscillation of f on any subintervals is at most $2M$. Therefore by substituting the result that we obtained, we have:

$$\begin{aligned} \sum_{t_i \in D_f} (M_i - m_i)(x_i - x_{i-1}) &\leq 2M \sum_{t_i \in D_f} (x_i - x_{i-1}) \\ &< 2M \cdot \frac{\epsilon}{4M} = \frac{\epsilon}{2}. \end{aligned}$$

Finally we can now combine the contribution obtained in the two cases. Then it follows that:

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{t_i \notin D_f} (M_i - m_i)(x_i - x_{i-1}) + \sum_{t_i \in D_f} (M_i - m_i)(x_i - x_{i-1}) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Therefore, we conclude that f is Riemann integrable on $[a, b]$. This completes our proof. $\square$

5. FUNCTIONS WITH UNCOUNTABLY MANY DISCONTINUITIES

We have now reached the final section of this paper. In this section, we turn to bounded functions whose sets of discontinuities are uncountable. In contrast with the prior cases, having uncountably many discontinuities does to the same result. For instance, consider Dirichlet's nowhere-continuous function:

$$f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \\ 0, & \text{if } x \notin \mathbb{Q}. \end{cases}$$

It is easy to see that every subinterval of $[0, 1]$ contains both rationals and irrationals, and hence $M_i - m_i = 1$ for every i . Accordingly, $U(f, P) - L(f, P) = 1$ for any choice of partition, and so f is not Riemann integrable. However on the other hand, as we will show in this section, the characteristic function of the Cantor set is Riemann integrable although its set of discontinuities is uncountable. This tells us one thing: uncountable set of discontinuities does not necessarily preclude Riemann integrability. Therefore, the aim of this section is to understand what distinguishes these two situations. More particularly, we consider why a function with discontinuities contained in the Cantor set can still be Riemann integrable by proving directly. Finally, we then generalize this argument to a broader class of sets with similar behaviour. As usual, we begin by establishing the preliminaries that will be needed for the main result.

Definition 5.1. Let C_0 be the interval $[0, 1]$, and define C_1 to be the set that obtained by removing the middle open third of C_0 . That is, $C_1 = C_0 \setminus (1/3, 2/3) = [0, 1/3] \cup [2/3, 1]$. Continuing this process, let C_n be obtained from C_{n-1} by removing the open middle third of each subinterval. Then *Cantor set* C is the intersection $C = \bigcap_{n=0}^{\infty} C_n$.

Remark 5.2. Considering that Cantor set is defined by $C = \bigcap_{n=0}^{\infty} C_n$, it follows that $C \subseteq C_n$ for every $n \geq 0$. That means, the Cantor set lies inside any finite stage C_n .

Before moving on to our next result, it is good to first note that the Cantor set is uncountable [1]. Considering our aim of this section is to focus on uncountable sets of discontinuities, it is clear why the Cantor set deserves further discussion in this section.

Lemma 5.3. *Let C_n denote the n -th stage in forming the Cantor set. Then C_n consists of 2^n closed intervals, where each has length $(1/3)^n$. Moreover, the total length of the interval in C_n is $(2/3)^n$.*

Proof. The proof is straightforward from the definition of C_n . Notice, at each stage the number of closed interval doubles and the length of each subinterval is divided by 3. Hence after n stages, it is formed of 2^n subintervals, where each length is $(1/3)^n$. Then it follows that their total length is $2^n(1/3)^n = (2/3)^n$. $\square$

We now consider bounded functions whose sets of discontinuities are contained in the Cantor set. The following result answers to the question raised at the beginning of this section by showing directly how Riemann integrability can still hold in this uncountable case. The main idea of this proof is by using sufficiently large finite stage C_n to separate the interval $[0, 1]$ into different types of subintervals and control the contribution separately. Details are given in the following theorem.

Theorem 5.4. *Let $f : [0, 1] \rightarrow \mathbb{R}$ be bounded, if the set of discontinuities is contained in the Cantor set C , then f is Riemann integrable on $[0, 1]$.*

Proof. As usual, let $\epsilon > 0$ be given. Since f is bounded, there exists $M > 0$ such that $|f(x)| \leq M$ for all $x \in [0, 1]$. Then it follows that the oscillation of f is at most $2M$ on any subinterval of $[0, 1]$.

We first divide the interval $[0, 1]$ into several parts. Recall that $D_f \subseteq C$, and by definition of Cantor set, $C \subseteq C_n$ for every $n \in \mathbb{N}$. Hence it implies that every discontinuity of f is contained in C_n as well. By Lemma 5.3, we know that the total length of the subintervals forming C_n is $(2/3)^n$. Notice that as $n \rightarrow \infty$, the total length of C_n converges to 0. Therefore, we can choose n large enough so that satisfy $(2/3)^n < \epsilon/(6M)$. Now bearing in mind that C_n is finite union of closed interval, it follows that the complement $[0, 1] \setminus C_n$ is finite union of open intervals. We denote them as $(a_1, b_1), \dots, (a_m, b_m)$. For each interval (a_j, b_j) , we then choose points α_j and β_j sufficiently close to the end points so that the total length of all the intervals on the edge is less than $\epsilon/(6M)$, in other words, $\sum_{j=1}^m ((\alpha_j - a_j) + (b_j - \beta_j)) < \epsilon/(6M)$. Thus, $[0, 1]$ now can be written as union of following three types of subintervals: closed intervals forming C_n , small intervals near endpoints of each (a_j, b_j) , and the closed middle intervals $[\alpha_j, \beta_j]$.

We now estimate the contribution from each type. First, we consider the subintervals forming C_n . Since their total length is less than $\epsilon/(6M)$, and the oscillation

of f on each interval is at most $2M$, it induces that the difference between the upper and lower Darboux sum is less than $\epsilon/(6M) \cdot 2M = \epsilon/3$. We next consider the small intervals near end point of each (a_j, b_j) . As before, the oscillation of f on these intervals is at most $2M$. Also as their total length was chosen to be less than $\epsilon/(6M)$, the contribution for this type is bounded by $2M \cdot \epsilon/(6M) = \epsilon/3$. Now, it remains to consider the closed middle interval $[\alpha_j, \beta_j]$. Given that $D_f \subset C_n$, and each $[\alpha_j, \beta_j]$ lies on $[0, 1] \setminus C_n$, we know that none of these intervals contains a discontinuity of f . That means, f is continuous on every $[\alpha_j, \beta_j]$. Therefore by Theorem 3.3, f is Riemann integrable on each of these intervals, and consequently for each $j = 1, 2, \dots, m$, there exist partition P_j of $[\alpha_j, \beta_j]$ that satisfy $\sum_{j=1}^m (U(f, P_j) - L(f, P_j)) < \epsilon/3$.

Finally, let P be partition of $[0, 1]$ obtained by combining the endpoints of subintervals forming C_n , the points α_j, β_j , and all the partition points P_j . By combining the result obtained above, we now have: $U(f, P) - L(f, P) < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon$. Therefore we conclude that f is Riemann integrable. $\square$

The preceding argument relies on an important property of the Cantor set. Namely, the discontinuity set can be contained in finitely many closed intervals whose total length can be made arbitrarily small. This suggests that the same argument may be extended to sets with the same property. We now formalize this idea.

Definition 5.5. Let $E \subseteq [a, b]$. We say E has *Cantor behaviour* if there exist sequence of finite unions of closed intervals $\{E_n\}_{n \geq 1}$ such that $E \subseteq E_n$ for every $n \in \mathbb{N}$, and the total length of the intervals forming E_n converges to 0 as $n \rightarrow \infty$.

Remark. The Cantor set has Cantor behaviour. Indeed, taking $E_n = C_n$ we have $C \subseteq C_n$ for all n , and the total length of the intervals forming C_n is $(2/3)^n$, which converges to 0.

By following the argument of Theorem 5.4 and replacing the finite stages C_n with the sets E_n , we obtain a stronger results for sets with Cantor behaviour. This gives the following theorem.

Theorem 5.6. Let $f : [a, b] \rightarrow \mathbb{R}$ be bounded. If the set of discontinuities D_f has Cantor behaviour, then f is Riemann integrable on $[a, b]$.

Proof. Let $\epsilon > 0$ be arbitrary. Given that f is bounded, there exists $M > 0$ such that $|f(x)| \leq M$ for all $x \in [a, b]$, and it follows that the oscillation of f is bounded above by $2M$ on any subinterval.

In the same manner as in Theorem 5.4, we separate the interval $[a, b]$ into three types of subintervals. Since D_f has Cantor behaviour, there exists a sequence of finite unions of closed intervals such that $\{E_n\}$ that satisfy $D_f \subseteq E_n$ for all n , where the total length of the intervals forming E_n converges to 0. Thus, we choose n sufficiently large so that the total length is less than $\epsilon/(6M)$. We next consider the complement $[a, b] \setminus E_n$. Because E_n is a finite union of closed intervals $[a, b] \setminus E_n$ consists of finitely many open intervals. As in the Theorem 5.4, we divide each open intervals into two small intervals near endpoints and one closed interval in the middle, where the total length of the intervals at the edge is less than $\epsilon/(6M)$.

We now consider the three resulting parts. The interval forming E_n have total length less than $\epsilon/(6M)$, together with the fact that oscillation is at most $2M$, it implies that the contribution is less than $\epsilon/3$, and the same estimate applies to

the intervals near the endpoints. Considering that $D_f \subseteq E_n$, we know that the remaining middle intervals contains no discontinuities, in other words, continuous. Therefore Theorem 3.3 tells us that f is Riemann integrable on each of the intervals. Equivalently, we can choose partitions so that their total contribution is less than $\epsilon/3$.

Finally, by combining all the partition point into a partition P of $[a, b]$, we obtain: $U(f, P) - L(f, P) < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon$. Therefore we conclude that f is Riemann integrable. This completes our proof. $\square$

In conclusion, the result obtained in each section tells us that the cardinality of D_f alone does not determine Riemann integrability. Moreover, the different approaches used in the preceding sections are aligned with one common principle: the discontinuities are controlled by covering them with intervals of arbitrarily small total length. This brings the discussion back to Lebesgue's criterion, where the measure zero condition captures this principle in its general form. As a result, although the results obtained in this paper do not provide a complete characterization of Riemann integrability, I believe that these direct arguments show how parts of the Lebesgue criterion can be understood through elementary Darboux sum arguments.

REFERENCES

- [1] Stephen Abbott, *Understanding Analysis*, 2nd ed., Springer, 2015.
- [2] Russell A. Gordon, *The Use of Tagged Partitions in Elementary Real Analysis*, The American Mathematical Monthly, Vol. 105, No. 2, pp. 107–117, 1998.
- [3] Seoungwan Song, *A Criterion for Riemann Integrability with Bounded Derivatives*, preprint, 2026.