CANTOR SET ARITHMETIC

JAYADEV S. ATHREYA, BRUCE REZNICK, AND JEREMY T. TYSON

Abstract. Every element u of $[0,1]$ can be written in the form $u = x^2 y$, where x, y are elements of the Cantor set C. In particular, every real number between zero and one is the product of three elements of the Cantor set. On the other hand the set of real numbers v that can be written in the form $v = xy$ with x and y in C is a closed subset of $[0,1]$ with Lebesgue measure strictly between $\frac{17}{21}$ and $\frac{8}{9}$. We also describe the structure of the quotient of C by itself, that is, the image of $C \times (C \setminus \{0\})$ under the function $f(x,y) = x/y$.

1. Introduction.

One of the first exotic mathematical objects encountered by the post-calculus student is the Cantor set

$$C = \left\{ \sum_{k=1}^{\infty} \alpha_k 3^{-k}, \alpha_k \in \{0,2\} \right\}. \tag{1.1}$$

(See Section 2 for several equivalent definitions of C.) One of its most beautiful properties is that

$$C + C := \{x + y : x, y \in C\} \tag{1.2}$$

is equal to $[0,2]$. (The whole interval is produced by adding dust to itself!)

The first published proof of (1.2) was by Hugo Steinhaus [9] in 1917. The result was later rediscovered by John Randolph in 1940 [7].

We remind the reader of the beautiful constructive proof of (1.2). It is enough to prove the containment $C + C \supset [0,2]$. Given $u \in [0,2]$, consider the ternary representation for $u/2$:

$$\frac{u}{2} = \sum_{k=1}^{\infty} \frac{\epsilon_k}{3^k}, \quad \epsilon_k \in \{0,1,2\}. \tag{1.3}$$

Date: November 23, 2017.

2000 *Mathematics Subject Classification.* Primary: 28A80, Secondary: 11K55 .

J.S.A. was supported by NSF CAREER grant DMS-1559860 and NSF grants DMS-1069153, DMS-1107452, DMS-1107263 and DMS-1107367. B.R. was supported by Simons Collaboration Grant 280987. J.T.T. was supported by NSF grants DMS-1201875 and DMS-1600650 and Simons Collaboration Grant 353627.

Define pairs (α_k, β_k) to be $(0,0), (2,0), (2,2)$ according to whether $\epsilon_k = 0, 1, 2$, respectively, and define elements $x, y \in C$ by

$$x = \sum_{k=1}^{\infty} \frac{\alpha_k}{3^k}, \qquad y = \sum_{k=1}^{\infty} \frac{\beta_k}{3^k}. \tag{1.4}$$

Since $\alpha_k + \beta_k = 2\epsilon_k$, $x + y = 2 \cdot \frac{u}{2} = u$.

While presenting this proof in a class, one of the authors (BR) wondered what would happen if addition were replaced by the other arithmetic operations. Another author (JT) immediately pointed out that subtraction is easy, because of a symmetry of C:

$$x = \sum_{k=1}^{\infty} \frac{\epsilon_k}{3^k} \in C \iff 1 - x = \sum_{k=1}^{\infty} \frac{2 - \epsilon_k}{3^k} \in C.$$

Thus

$$C - C := \{x - y : x, y \in C\} = \{x - (1 - z) : x, z \in C\} = C + C - 1 = [-1, 1].$$

More generally, to understand the structure of linear combinations $aC + bC$, $a, b \in \mathbb{R}$, it suffices to consider the case $a = 1$ and $0 \le b \le 1$. (If $a > b > 0$, then $aC + bC = a(C + (b/a)C)$; the remaining cases are left to the reader.) The precise structure of the linear combination set

$$aC + bC$$

was obtained by Pawłowicz [6], who extended an earlier result by Utz [10], published in 1951. Utz's result states that

$$C + bC = [0, 1 + b] \qquad \text{for every } \tfrac{1}{3} \le b \le 3. \tag{1.5}$$

Multiplication is trickier. The inclusion $C \subset [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ implies that any element of the interval $(\frac{1}{3}, \frac{4}{9})$ cannot be written as a product of two elements from C. Thus the measure of the product of C with itself is at most $\frac{8}{9}$. This paper grew out of a study of multiplication on C.

In this paper we will prove the following results.

Theorem 1.1.

(1) *Every $u \in [0, 1]$ can be written as $u = x^2 y$ for some $x, y \in C$.*
(2) *The set of quotients from C can be described as follows:*

$$\left\{ \frac{x}{y} : x, y \in C, y \neq 0 \right\} = \bigcup_{m=-\infty}^{\infty} \left[\frac{2}{3} \cdot 3^m, \frac{3}{2} \cdot 3^m \right]. \tag{1.6}$$

(3) *The set $\{xy : x, y \in C\}$ is a closed set with Lebesgue measure strictly greater than $\frac{17}{21}$.*

In particular, part (1) of Theorem 1.1 implies that every real number in the interval $[0, 1]$ is the product of three elements of C.

In words, (1.6) says that each positive real number is a quotient of two elements of the Cantor set if and only if either the left-most nonzero digit in the ternary representation of u is "2," or the left-most nonzero digit is "1," but the first subsequent non-"1" digit is "0," not "2."

This paper is organized as follows. We begin in section 2 with several different descriptions of the Cantor set, and then the key tools, all of which are accessible to students in a good undergraduate analysis class. As a warmup, we use these tools to give a short proof of Utz's result (1.5) in Section 3.1. Sections 3.2, 3.3 and 3.4 are devoted to the proofs of parts (1), (2) and (3) of Theorem 1.1, respectively. Sprinkled throughout are relevant open questions. This article began as a standard research paper, but we realized that many of our main results might be of interest to a wider audience. In Section 4, we discuss some of our other results, which will be published elsewhere, with different combinations of co-authors.

2. Tools.

We begin by recalling several equivalent and well-known definitions of the Cantor set. See Fleron [3] and the references within for an excellent overview of the history of the Cantor set and the context in which several of these definitions first arose.

The standard ternary representation of a real number x in $[0,1]$ is

$$x = \sum_{k=1}^{\infty} \frac{\alpha_k(x)}{3^k}, \qquad \alpha_k(x) \in \{0,1,2\}. \tag{2.1}$$

This representation is unique, except for the *ternary rationals*, $\{\frac{m}{3^n}, m, n \in \mathbb{N}\}$, which have two ternary representations. Supposing $\alpha_n > 0$ and $m \neq 0$ mod 3, so that $\alpha_n \in \{1,2\}$ below, we have

$$\begin{aligned} \frac{m}{3^n} &= \sum_{k=1}^{n-1} \frac{\alpha_k}{3^k} + \frac{\alpha_n}{3^n} + \sum_{k=n+1}^{\infty} \frac{0}{3^k} \\ &= \sum_{k=1}^{n-1} \frac{\alpha_k}{3^k} + \frac{\alpha_n - 1}{3^n} + \sum_{k=n+1}^{\infty} \frac{2}{3^k}. \end{aligned} \tag{2.2}$$

The *Cantor set* C consists those $x \in [0,1]$ admitting a ternary representation as in (2.1) with $\alpha_k(x) \in \{0,2\}$ for all k. Note that C also contains those ternary rationals as in (2.2) whose final digit is "1." These may be transformed as above into a representation in which $\alpha_n(x) = 0$ and $\alpha_k(x) = 2$ for $k > n$. As noted earlier, $x \in C$ if and only if $1 - x \in C$. Further, if $k \in \mathbb{N}$, then

$$x \in C \implies 3^{-k} x \in C. \tag{2.3}$$

This definition arises in dynamical systems, as the Cantor set C can be viewed as an invariant set for the map $x \mapsto 3x \bmod 1$, or equivalently, as the image of an invariant set C' for the one-sided shift map σ acting on $\Omega = \{0,1,2\}^{\mathbb{N}}$. Given a sequence $\omega = (\omega_n)_{n=1}^{\infty} = (\omega_1, \omega_2, \ldots)$ in Ω,

$$\sigma\omega = (\omega_2, \omega_3, \ldots).$$

Letting $C' = \{0,2\}^{\mathbb{N}} \subset \Omega$, we realize the Cantor set C as the image of C' under the coding map $T : \Omega \to [0,1]$ given by

$$T(\omega) = \sum_{n=1}^{\infty} \omega_n 3^{-n}.$$

We now present the usual "middle-third" definition of the Cantor set: Define $C_n = \{x : \alpha_k(x) \in \{0,2\}, 1 \le k \le n\}$, which is a union of 2^n closed intervals of length 3^{-n}, written as

$$C_n = \bigcup_{i=1}^{2^n} I_{n,i}. \tag{2.4}$$

The left-hand endpoints of the $I_{n,i}$'s comprise the set

$$\left\{ \sum_{k=1}^{n} \frac{\epsilon_k}{3^k} : \epsilon_k \in \{0,2\} \right\}. \tag{2.5}$$

The right-hand endpoints have "1" as their final nonzero ternary digit when written as a finite ternary expansion.

The more direct definition of C is as a nested intersection of closed sets:

$$C = \bigcap_{n=1}^{\infty} C_n; \qquad C_1 \supset C_2 \supset C_3 \supset \cdots. \tag{2.6}$$

This definition is standard in fractal geometry, where the Cantor set C is seen as the invariant set for the pair of contractive linear mappings $f_1(x) = \frac{1}{3}x$ and $f_2(x) = \frac{1}{3}x + \frac{2}{3}$ acting on the real line. That is, C is the unique nonempty compact set that is fully invariant under f_1 and f_2:

$$C = f_1(C) \cup f_2(C).$$

Observe that each "parent" interval $I_{n,i} = [a, a + \frac{1}{3^n}]$ in C_n has two "child" intervals

$$I_{n+1,2i-1} = \left[a, a + \tfrac{1}{3^{n+1}}\right], \qquad I_{n+1,2i} = \left[a + \tfrac{2}{3^{n+1}}, a + \tfrac{3}{3^{n+1}}\right] \tag{2.7}$$

in C_{n+1}, and C_{n+1} is the union of all children intervals whose parents are in C_n.

It is useful to introduce the following notation to represent the omission of the middle third:

$$I = [a, a+3t] \implies \ddot{I} = [a, a+t] \cup [a+2t, a+3t]. \tag{2.8}$$

Using this notation,

$$C_{n+1} = \bigcup_{i=1}^{2^{n+1}} I_{n+1,i} = \bigcup_{i=1}^{2^n} \ddot{I}_{n,i}. \tag{2.9}$$

It will also be useful, for studying products and quotients, to give a third definition. Let $\widetilde{C} = C \cap [1/2, 1] = C \cap [2/3, 1] = \frac{2}{3} + \frac{1}{3} \cdot C$. Then by examining the smallest k for which $\epsilon_k = 2$, we see that

$$C = \{0\} \cup \bigcup_{k=0}^{\infty} 3^{-k}\widetilde{C}. \tag{2.10}$$

Similarly, let $\widetilde{C}_n = C_n \cap [1/2, 1]$, consisting of 2^{n-1} closed intervals of length 3^{-n}:

$$\widetilde{C}_n = \bigcup_{i=2^{n-1}+1}^{2^n} I_{n,i}. \tag{2.11}$$

Then

$$\widetilde{C} = \bigcap_{n=1}^{\infty} \widetilde{C}_n. \tag{2.12}$$

And, analogously,

$$\widetilde{C}_{n+1} = \bigcup_{i=2^n+1}^{2^{n+1}} I_{n+1,i} = \bigcup_{i=2^{n-1}+1}^{2^n} \ddot{I}_{n,i}. \tag{2.13}$$

By looking at the left-hand endpoints, we see that each interval $I_{n,i} \neq [0, \frac{1}{3^n}]$ can be written as $\frac{1}{3^{n-k}} I_{k,j}$ for some $I_{k,j} \in \widetilde{C}_k$; hence

$$C_n = \left[0, \frac{1}{3^n}\right] \cup \bigcup_{k=1}^{n} \frac{1}{3^{n-k}} \widetilde{C}_k. \tag{2.14}$$

The keys to our method lie in two lemmas which might appear on a first serious analysis exam. (Please do not send such exams to the authors!)

Lemma 2.1. *Suppose* $\{K_i\} \subset \mathbb{R}$ *are nonempty compact sets,* $K_1 \supseteq K_2 \supseteq K_3 \supseteq \cdots$, *and* $K = \cap K_i$.
(i) If $(x_j) \to x$, $x_j \in K_j$, *then* $x \in K$.
(ii) If $F : \mathbb{R}^m \to \mathbb{R}$ *is continuous, then* $F(K^m) = \cap F(K_i^m)$.

Proof. (i). If $x \notin K$, then $x \notin K_r$ for some r. Since K_r^c is open, there exists $\epsilon > 0$ so that $(x - \epsilon, x + \epsilon) \subseteq K_r^c \subseteq K_{r+1}^c \cdots$ and hence $|x_j - x| \geq \epsilon$ for $j \geq r$, a contradiction to $x_j \to x$.

(ii). Since $K \subseteq K_i$, we have $F(K^m) \subseteq \cap F(K_i^m)$. Conversely, suppose $u \in \cap F(K_i^m)$. We need to find $x \in K^m$ such that $F(x) = u$. For each i, choose $x_i = (x_{i,1}, \ldots, x_{i,m}) \in K_i^m$ so that $F(x_i) = u$. Since K_1^m is compact, the Bolzano–Weierstrass theorem implies that the sequence (x_i) has a convergent subsequence $x_{r_j} = (x_{r_j,1}, \ldots x_{r_j,m}) \to y = (y_1, \ldots, y_m)$. Applying (i) to the subsequence $K_{r_1} \supseteq K_{r_2} \supseteq K_{r_3} \supseteq \cdots$, we see that each $y_k \in K$ and since F is continuous, $F(y) = u$, as desired. □

If we perform the middle-third construction with an initial interval of $[a, b]$, it is easy to see that the limiting object is a translate of the Cantor set, specifically $C_{a,b} := a + (b - a)C$.

Lemma 2.2. *Suppose $F : \mathbb{R}^m \to \mathbb{R}$ is continuous, and suppose that for every choice of disjoint or identical subintervals $I_k \subset [a, b]$ of common length,*

$$F(I_1, \dots, I_m) = F(\ddot{I}_1, \dots, \ddot{I}_m). \tag{2.15}$$

Then $F(C^m_{a,b}) = F([a, b]^m)$.

Proof. We prove the result for $[a, b] = [0, 1]$; the result follows generally by composing F with an appropriate linear function. Let

$$C_k = \bigcup_{j=1}^{2^k} I_{k,j}, \tag{2.16}$$

where each interval $I_{k,j}$ has length 3^{-k}. It follows that

$$F(C^m_k) = \bigcup_{1 \le j_1, \dots, j_m \le 2^k} F(I_{k,j_1}, \dots, I_{k,j_m}), \tag{2.17}$$

where for each pair (ℓ, ℓ'), I_{k,j_ℓ} and I_{k,j'_ℓ} are either identical or disjoint. Since

$$C_{k+1} = \bigcup_{j=1}^{2^k} \ddot{I}_{k,j}, \tag{2.18}$$

the hypothesis implies that $F(C^m_k) = F(C^m_{k+1})$, and the result then follows from Lemma 2.1(ii). □

We only apply Lemma 2.2 in the cases that $m = 2$ and $[a, b] = [0, 1]$ and $[\frac{2}{3}, 1]$. (In the latter case, when $F(x, y) = xy$ or x/y, it is helpful to have control of the ratio x/y.)

3. Arithmetic on the Cantor set.

3.1. **Addition and subtraction.** Sums and differences of Cantor sets have been widely studied in connection with dynamical systems. In this section we give a brief proof of the following result of Utz [10]. Further information about sums of Cantor sets can be found in [1] and [2].

Theorem 3.1 (Utz)**.** *If $\lambda \in [\frac{1}{3}, 3]$, then every element u in $[0, 1 + \lambda]$ can be written in the form $u = x + \lambda y$ for $x, y \in C$.*

We include this proof in order to introduce the main ideas in the proof of Theorem 1.1 in a simpler context. The key tool is Lemma 2.2.

Let $f_\lambda(x, y) = x + \lambda y$; we wish to show that $f_\lambda(C^2) = [0, 1 + \lambda]$. Observe that $C + \lambda C = \lambda(C + \lambda^{-1}C)$ for $\lambda \ne 0$, so it suffices to consider $\frac{1}{3} \le \lambda \le 1$.

Proof. We apply Lemma 2.2 and show that for any two closed intervals I_1, I_2 of the same length in $[a, b]$, $f_\lambda(I_1, I_2) = f_\lambda(\ddot{I}_1, \ddot{I}_2)$. For clarity, we write $I_1 = [r, r+3t]$, $I_2 = [s, s+3t]$, and $w = r+\lambda s$, so that $f_\lambda(I_1, I_2) = [w, w+3(1+\lambda)t]$.

Observe that $\ddot{I}_1 = [r, r+t] \cup [r+2t, r+3t]$ and $\ddot{I}_2 = [s, s+t] \cup [s+2t, s+3t]$, so

(3.1)
$$f_\lambda(\ddot{I}_1, \ddot{I}_2) = (\ [w, w+(1+\lambda)t] \cup [w+2\lambda t, w+(1+3\lambda)t]\ ) \\ \cup (\ [w+2t, w+(3+\lambda)t] \cup [w+(2+2\lambda t, w+(3+3\lambda)t]\ )\ .$$

Since $\lambda \leq 1$, $1+\lambda \geq 2\lambda$ and $3+\lambda \geq 2+2\lambda$, the pairs of intervals coalesce into

$$f_\lambda(\ddot{I}_1, \ddot{I}_2) = [w, w+(1+3\lambda)t]) \cup [w+2t, w+(3+3\lambda)t]. \tag{3.2}$$

Since $\lambda \geq \frac{1}{3}$, we have $2 \leq 1+3\lambda$. Hence $f_\lambda(\ddot{I}_1, \ddot{I}_2) = f_\lambda(I_1, I_2)$, completing the proof. □

Remark 1. Unlike the folklore proof for $\lambda = 1$, there seems to be no obvious algorithmic proof, save for $\lambda = \frac{1}{3}$. In this case, suppose $u \in [0, \frac{4}{3}]$. If $u = \frac{4}{3}$, then $u = 1 + \frac{1}{3} \cdot 1 \in C + \frac{1}{3}C$. If $u < \frac{4}{3}$, then we can write $u = x+y$, $x, y \in C$, and assume $x \geq y$. Since $y \leq \frac{u}{2} < \frac{2}{3}$ is in C, $y \leq \frac{1}{3}$, hence $y = \frac{1}{3}z$, $z \in C$. This produces the desired construction.

The case of subtraction, that is, the case of f_λ when $\lambda < 0$, is easily handled.

Theorem 3.2. *If $\beta = -\lambda < 0$, then*

$$f_\beta(C^2) = -\lambda + f_\lambda(C^2). \tag{3.3}$$

Proof. If $\beta < 0$, $x, y \in C$, we have

$$x + \beta y = x + \beta(1-z) = -\lambda + x + \lambda z \tag{3.4}$$

for x and z in C. □

Remark 2. Arithmetic sums of Cantor sets and more general compact sets have been studied intensively. For instance, Mendes and Oliveira [5] discuss the topological structure of sums of Cantor sets, while Schmeling and Shmerkin [8] characterize those nondecreasing sequences $0 \leq d_1 \leq d_2 \leq d_3 \leq \cdots \leq 1$ that can arise as the sequence of Hausdorff dimensions of iterated sumsets $A, A+A, A+A+A, \ldots$ for a compact subset A of $\mathbb{R}$. Recent work of Gorodetski and Northrup [4] involves the Lebesgue measure of sumsets of Cantor sets and other compact subsets of the real line. We refer the interested reader to these papers and the references therein for more information.

3.2. **Multiplication.** We let $f(x,y) = x^2 y$ and shall show that $f(C^2) = [0,1]$. We begin by showing that it suffices to consider $f(\widetilde{C}^2)$.

Lemma 3.3. *If $f(\widetilde{C}^2) = [\frac{8}{27}, 1]$, then $f(C^2) = [0,1]$.*

Proof. Suppose $u \in [0,1]$. If $u = 0$, then $u = 0^2 \cdot 0$. If $u > 0$, then there exists a unique integer $r \geq 0$ so that $u = 3^{-r}v$, where $v \in (\frac{1}{3}, 1]$. Since $\frac{8}{27} < \frac{1}{3}$, $v = x^2 y$ for $x, y \in \widetilde{C} \subset C$, and since $x, 3^{-r}y \in C$, $u = x^2(3^{-r}y)$ is the desired representation. □

Accordingly, we confine our attention to $\widetilde{C}$.

Lemma 3.4. *If* $I = [a, a+3t]$ *and* $J = [b, b+3t]$ *are in* $[\frac{2}{3}, 1]$*, then* $f(I,J) = f(\ddot{I}, \ddot{J})$.

Proof. We first define

$$\begin{aligned} [a^2 b, (a+t)^2(b+t)] &=: [u_1, v_1]; \\ [a^2(b+2t), (a+t)^2(b+3t)] &=: [u_2, v_2]; \\ [(a+2t)^2 b, (a+3t)^2(b+t)] &=: [u_3, v_3]; \\ [(a+2t)^2(b+2t), (a+3t)^2(b+3t)] &=: [u_4, v_4]. \end{aligned} \tag{3.5}$$

Evidently, $f(I,J) = [u_1, v_4]$, and also, $u_1 < u_2, v_1 < v_2$, and $u_3 < u_4, v_3 < v_4$. If we can first show that $v_1 > u_2$ and $v_3 > u_4$, then $[u_1, v_1] \cup [u_2, v_2] = [u_1, v_2]$ and $[u_3, v_3] \cup [u_4, v_4] = [u_3, v_4]$. Second, since $u_1 < u_3$ and $v_2 < v_4$, if we can show that $v_2 > u_3$, then $[u_1, v_2] \cup [u_3, v_4] = [u_1, v_4]$, and the proof will be complete.

We compute

$$\begin{aligned} v_1 - u_2 &= a(2b-a)t + (2a+b)t^2 + t^3; \\ v_2 - u_3 &= a(3a-2b)t + (6a-3b)t^2 + 3t^3; \\ v_3 - u_4 &= a(2b-a)t + (5b-2a)t^2 + t^3. \end{aligned} \tag{3.6}$$

Since $2b - a \geq 2 \cdot \frac{2}{3} - 1 > 0$, $3a - 2b \geq 3 \cdot \frac{2}{3} - 2 \geq 0$, $6a - 3b \geq 6 \cdot \frac{2}{3} - 3 > 0$, and $5b - 2a \geq 5 \cdot \frac{2}{3} - 2 > 0$, each of the quantities in (3.6) is positive and the proof is complete. □

Theorem 3.5.

$$f(\widetilde{C}^2) = [\tfrac{8}{27}, 1]. \tag{3.7}$$

Proof. Apply Lemma 2.2, noting that $f([\frac{2}{3}, 1]^2) = [\frac{8}{27}, 1]$. □

Remark 3. Observe that $u \in [0,1]$, written as $x^2 y$, where $x, y \in C$, is also $x \cdot x \cdot y$, so this implies that every element in $[0,1]$ is a product of three elements from the Cantor set. (This can also be proved in a more ungainly way, by taking $m = 3$ in Lemma 2.2 with $f(x_1, x_2, x_3) = x_1 x_2 x_3$.)

Remark 4. We do not know an algorithm for expressing $u \in [0,1]$ in the form of $x^2 y$ or $x_1 x_2 x_3$ as a product of elements of the Cantor set.

Remark 5. More generally, one can look at $f_{a,b}(C^2)$ where $f_{a,b}(x,y) := x^a y^b$. Since $u = x^a y^b$ if and only if $u^{1/a} = x y^{b/a}$, for $u \in (0,1)$, it suffices to consider $a = 1$. By looking at $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, it is not hard to see that if $f(x,y) = xy^t$, $t \geq 1$, and $(\frac{2}{3})^{1+t} > \frac{1}{3}$, then $f(C_1^2)$ is already missing an interval from $[0,1]$. This condition occurs when $t < \frac{\log 2}{\log 3/2} \approx 1.7095$.

Remark 6. By taking logarithms, we can convert the question about products of elements of C into a question about sums. (We can omit the point 0 since its multiplicative behavior is trivial.) Of course, the underlying set is no longer the standard self-similar Cantor set but is a more general ("nonlinear") closed subset of $\mathbb{R}$. Some conclusions about the number of factors needed to recover all of $[0, 1]$ can be obtained from the general results in the papers of Cabrelli–Hare–Molter [1], [2], but it does not appear that one obtains the precise conclusion of part 2 of Theorem 1.1 in this fashion.

3.3. **Division.** In this section, we complete our arithmetic discussion by considering quotients.

Theorem 3.6.

$$\left\{\frac{u}{v} : u, v \in C\right\} = \bigcup_{m=-\infty}^{\infty} \left[\frac{2}{3} \cdot 3^m, \frac{3}{2} \cdot 3^m\right]. \tag{3.8}$$

Proof. As with multiplication, it suffices to consider $\widetilde{C}$.

Lemma 3.7. *Theorem 3.6 is implied by the identity*

$$\left\{\frac{u}{v} : u, v \in \widetilde{C}\right\} = \left[\frac{2}{3}, \frac{3}{2}\right]. \tag{3.9}$$

Proof. Write $u, v \in C$ as $u = 3^{-s}\tilde{u}$, $v = 3^{-t}\tilde{v}$ for integers $c, d \geq 0$ and $\tilde{u}, \tilde{v} \in \widetilde{C}$. Then $u/v = 3^{d-c}\tilde{u}/\tilde{v}$, where $m = d - c$ can attain any integer value. □

We now prove (3.9). Consider $\widetilde{C}_1 = [\frac{2}{3}, 1]$ and apply Lemma 2.2. Clearly, $\{\frac{u}{v} : u, v \in \widetilde{C}_1\} = [\frac{2}{3}, \frac{3}{2}]$. Consider two intervals in $\widetilde{C}_n$, $I_1 = [a, a + 3t]$ and $I_2 = [b, b + 3t]$. These intervals are either identical or disjoint. Since $x = \frac{u}{v}$ implies $\frac{1}{x} = \frac{v}{u}$, there is no harm in assuming $a \leq b$, and either $I_1 = I_2$ and $a = b$, or the intervals are disjoint and $a + 3t \leq b$. The quotients from these intervals will lie in

$$J_0 := \left[\frac{a}{b + 3t}, \frac{a + 3t}{b}\right] := [r_0, s_0].$$

Since $\ddot{I}_1 = [a, a + t] \cup [a + 2t, a + 3t]$ and $\ddot{I}_2 = [b, b + t] \cup [b + 2t, b + 3t]$, we obtain four subintervals

$$J_1 = \left[\frac{a}{b + 3t}, \frac{a + t}{b + 2t}\right] = [r_1, s_1],$$

$$J_2 = \left[\frac{a}{b + t}, \frac{a + t}{b}\right] = [r_2, s_2],$$

$$J_3 = \left[\frac{a + 2t}{b + 3t}, \frac{a + 3t}{b + 2t}\right] = [r_3, s_3],$$

$$J_4 = \left[\frac{a + 2t}{b + t}, \frac{a + 3t}{b}\right] = [r_4, s_4].$$

We need to see how $J_0 = J_1 \cup J_2 \cup J_3 \cup J_4$. There are two cases, depending on whether $a = b$ or $a < b$.

We first record some algebraic relations. We have $r_1 = r_0$ and $s_4 = s_0$, and, evidently, $r_1 < r_2$, $s_1 < s_2$, $r_3 < r_4$, $s_3 < s_4$. Further,

$$\begin{aligned}
r_3 - r_2 &= \frac{a+2t}{b+3t} - \frac{a}{b+t} = \frac{2t(b-a+t)}{(b+t)(b+3t)}, \\
s_3 - s_2 &= \frac{a+3t}{b+2t} - \frac{a+t}{b} = \frac{2t(b-a-t)}{b(b+2t)}, \\
s_1 - r_2 &= \frac{a+t}{b+2t} - \frac{a}{b+t} = \frac{t(b-a+t)}{(b+t)(b+2t)}, \\
s_2 - r_3 &= \frac{a+t}{b} - \frac{a+2t}{b+3t} = \frac{t(3a+3t-b)}{b(b+3t)} \geq \frac{t(3\cdot\frac{2}{3}+0-1)}{b(b+3t)} > 0, \\
s_3 - r_4 &= \frac{a+3t}{b+2t} - \frac{a+2t}{b+t} = \frac{t(b-a-t)}{(b+t)(b+2t)}.
\end{aligned}$$

Suppose first that $a < b$, so $a+3t < b$. Then each of the differences above is positive, so $r_1 < r_2 < r_3 < r_4$ and $s_1 < s_2 < s_3 < s_4$; further, the intervals overlap: $s_1 > r_2$, $s_2 > r_3$ and $s_3 > r_4$. Thus $J_0 = J_1 \cup J_2 \cup J_3 \cup J_4$.

If $a = b$, then $J_3 = \left[\frac{a+2t}{a+3t}, \frac{a+3t}{a+2t}\right] \subset \left[\frac{a}{a+t}, \frac{a+t}{a}\right] = J_2$, so we may drop J_3 from consideration. We have $r_1 < r_2 < r_4$ and $s_1 < s_2 < s_4$ and need only show that $s_1 > r_2$ and $s_2 > r_4$. The first is clear, and for the second,

$$s_2 - r_4 = \frac{a+t}{a} - \frac{a+2t}{a+t} = \frac{t^2}{a(a+t)} > 0, \tag{3.10}$$

so $J_0 = J_1 \cup J_2 \cup J_4$, and we are done. □

Remark 7. We do not know an algorithm for expressing a feasible u as a quotient of elements in C.

3.4. **Multiplication, revisited.** Let $g(x, y) = xy$. As noted earlier, $g(C^2)$ is not the full interval $[0, 1]$, though $g(C^2) = \cap g(C_i^2)$ is the intersection of a descending chain of closed sets and so is closed. In order to gain some information about $g(C^2)$, we look carefully at how Lemma 2.2 fails.

Lemma 3.8. *Let $I = [a, a+3t]$ and $J = [b, b+3t]$, with $\frac{2}{3} \leq a \leq b \leq 1$, be either identical or disjoint intervals. Then*

$$\begin{aligned}
a < b &\implies g(\ddot{I}, \ddot{J}) = g(I, J)\,; \\
a = b &\implies g(\ddot{I}, \ddot{I}) = g(I, I) \setminus ((a+2t)^2 - t^2, (a+2t)^2).
\end{aligned}$$

Proof. We have

$$g([a, a+3t], [b, b+3t]) = [ab, ab + 3(a+b)t + 9t^2]$$

and

$$
\begin{aligned}
&g([a, a+t] \cup [a+2t, a+3t], [b, b+t] \cup [b+2t, b+3t]) = \\
&[ab, ab+(a+b)t+t^2] \cup [ab+2at, ab+(3a+b)t+3t^2] \\
&\cup [ab+2bt, ab+(a+3b)t+3t^2] \\
&\cup [ab+(2a+2b)t+4t^2, ab+3(a+b)t+9t^2].
\end{aligned} \tag{3.11}
$$

Since $a \le b$, it follows that $ab + 2at \le ab + (a+b)t + t^2$, and the first two intervals coalesce into $[ab, ab + (3a+b)t + 3t^2]$.

Suppose that $a < b$, and recall that we have assumed $\frac{2}{3} \le a < b \le 1$. Since $a + t \le b$, it follows that $ab + (2a+2b)t + 4t^2 \le ab + (a+3b)t + 3t^2$ and the last two intervals coalesce into $[ab + 2bt, ab + 3(a+b)t + 9t^2]$. Thus the right-hand side of (3.11) reduces to

$$
[ab, ab+(3a+b)t+3t^2] \cup [ab+2bt, ab+3(a+b)t+9t^2] \tag{3.12}
$$

Moreover,

$$
ab+(3a+b)t+3t^2-(ab+2bt) = t(3a+3t-b) \ge t(3 \cdot \tfrac{2}{3}+0-1) = t \ge 0,
$$

which shows that the pair of intervals in (3.12) coalesces to a single interval. This proves the first statement.

If $a = b$, then the middle two intervals in (3.11) are the same, and

$$
\begin{aligned}
&g(([a, a+t] \cup [a+2t, a+3t])^2) \\
&\quad = [a^2, a^2+4at+3t^2] \cup [a^2+4at+4t^2, a^2+6at+9t^2] \\
&\quad = [a^2, (a+3t)^2] \setminus ((a+2t)^2 - t^2, (a+2t)^2)).
\end{aligned}
$$

□

This leads to the following estimate for the Lebesgue measure of $g(C^2)$.

Theorem 3.9.

$$
\mu(g(C^2)) \ge \frac{17}{21}.
$$

Proof. First note that $g(\widetilde{C}^2) \subset g(\widetilde{C}_1^2) = [\frac{4}{9}, 1]$. It follows as before

$$
g(C^2) = \{0\} \cup \bigcup_{k=0}^{\infty} 3^{-k} g(\widetilde{C}^2), \tag{3.13}
$$

and since $\frac{1}{3} < \frac{4}{9}$, the sets $3^{-k} g(\widetilde{C}^2)$ are disjoint. Therefore,

$$
\mu(g(C^2)) = \sum_{k=0}^{\infty} 3^{-k} \mu(g(\widetilde{C}^2)) = \tfrac{3}{2} \mu(g(\widetilde{C}^2)). \tag{3.14}
$$

Since $\widetilde{C}_n$ consists of 2^{n-1} intervals of length 3^{-n}, it follows from Lemma 3.8 that

$$\begin{aligned} &\mu(g(\widetilde{C}^2_{n+1})) \geq \mu(g(\widetilde{C}^2_n)) - \frac{2^{n-1}}{3^{2n+2}} \\ \text{(3.15)} \qquad &\implies \mu(g(\widetilde{C}^2)) \geq \left(1 - \frac{4}{9}\right) - \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^{2n+2}} = \frac{34}{63} \\ &\implies \mu(g(C^2)) \geq \frac{3}{2} \cdot \frac{34}{63} = \frac{17}{21}. \end{aligned}$$

□

Remark 8. This argument shows that for all m,

$$\mu(g(\widetilde{C}^2_m)) \geq \mu(g(\widetilde{C}^2)) \geq \mu(g(\widetilde{C}^2_m)) - \sum_{n=m+1}^{\infty} \frac{2^{n-1}}{3^{2n+2}}. \tag{3.16}$$

The reason that Theorem 3.9 is only an estimate is that there is no guarantee that intervals missing from $f(\ddot{I}^2)$ cannot be covered elsewhere. The first instance in which this occurs is for $n = 4$: one of the intervals in $\widetilde{C}_4$ is $I_0 = [\frac{62}{81}, \frac{63}{81}] = [.2022_3, .21_3]$. By Lemma 3.8,

$$(\tfrac{188^2-1}{243^2}, \tfrac{188^2}{243^2}) = (\tfrac{35343}{59049}, \tfrac{35344}{59049}) \approx (.5985368, .5985537) \notin f(\ddot{I}^2_0). \tag{3.17}$$

However, $\widetilde{C}_5$ contains the intervals

$$J_1 = [\tfrac{162}{243}, \tfrac{163}{243}] = [.2_3, .20001_3], \quad J_2 = [\tfrac{216}{243}, \tfrac{217}{243}] = [.22_3, .22001_3] \tag{3.18}$$

and

$$J_1 J_2 = [\tfrac{34992}{59049}, \tfrac{35371}{59049}] \approx (.5925926, .59901099)$$

covers the otherwise-missing interval. A more detailed *Mathematica* computation, using $m = 11$, gives the first eight decimal digits for $\mu(g(C^2))$:

$$\mu(g(C^2)) = .80955358\cdots \approx \frac{17}{21} + 2.97 \times 10^{-5}. \tag{3.19}$$

4. Final remarks.

As mentioned in the introduction, this paper is part of a larger project. We discuss a few results from this project whose proofs will appear elsewhere, written by various combinations of the authors and their students.

4.1. **Self-similar Cantor sets.** The Cantor set easily generalizes to sets defined with different "middle-fractions" removed. Consider the self-similar Cantor set $D^{(t)}$ obtained as the invariant set for the pair of contractive mappings

$$f_1(x) = tx, \qquad f_2(x) = tx + (1-t)$$

acting on the real line. Thus

$$D^{(t)} = \bigcap_{n \geq 0} D^{(t)}_n,$$

where, for each $n \geq 0$, $D_n^{(t)}$ is a union of 2^n intervals, each of length t^n, contained in $[0,1]$. For instance,

$$D_1^{(t)} = [0,t] \cup [1-t,1],$$
$$D_2^{(t)} = [0,t^2] \cup [t(1-t),t] \cup [1-t,1-t+t^2] \cup [1-t^2,1]\,.$$

For an integer $m \geq 2$, let t_m be the unique solution to $(1-t)^m = t$ in $[0,1]$. Then

$$\cdots < t_4 < t_3 < t_2 < \frac{1}{2}$$

and $t_m \to 0$ as $m \to \infty$. Numerical values are

$$t_2 = \tfrac{3-\sqrt{5}}{2} \approx 0.381966\ldots$$
$$t_3 \approx 0.317672\ldots$$
$$t_4 \approx 0.275508\ldots$$
$$t_5 \approx 0.245122\ldots$$
$$t_6 \approx 0.22191\ldots$$
$$t_7 \approx 0.203456\ldots$$

If we choose t such that $t_m \leq t < t_{m-1}$ (that is, $(1-t)^m \leq t < (1-t)^{m-1}$) and let $g_n(x_1,\ldots,x_n) = x_1 x_2 \cdots x_n$, then

$$g_m((D^{(t)})^m) = [0,1],$$

but $g_{m-1}((D^{(t)})^{m-1})$ has Lebesgue measure strictly less than one. In particular, if a Cantor set D in which a middle fraction λ is taken with $\lambda \leq 1-2t_2 = \sqrt{5}-2 \approx .23607$, then every element in $[0,1]$ can be written as a product of two elements of D.

4.2. **Number Theory.** There is much more to be said about the representation of specific numbers in C/C. For example, if $v = 2u$ for $u,v \in C$, then

$$u = \frac{1}{3^n},\ v = \frac{2}{3^n}, \quad n \geq 1;$$

if $v = 11u$ for $u,v \in C$, then

$$u = \frac{1}{4\cdot 3^n},\ v = \frac{11}{4\cdot 3^n}, \quad n \geq 1.$$

By contrast, if $v = 4u$ for $u,v \in C$, then there exists a finite or infinite sequence of integers (n_k) with $n_1 \geq 2$ and $n_{k+1} - n_k \geq 2$, so that

$$u = \sum_k \frac{2}{3^{n_k}}, \qquad v = \sum_k \left(\frac{2}{3^{n_k-1}} + \frac{2}{3^{n_k}}\right).$$

The proof of the second result is trickier than the other two.

We also mention a conjecture for which there is strong numerical evidence.

Conjecture 4.1. *Every $u \in [0,1]$ can be written as $x_1^2+x_2^2+x_3^2+x_4^2$, $x_i \in C$.*

We need a minimum of four squares, since $(\frac{1}{3})^2 + (\frac{1}{3})^2 + (\frac{1}{3})^2 < (\frac{2}{3})^2$, so the open interval $(\frac{1}{3}, \frac{4}{9})$ will be missing from the sum of three squares.

5. Acknowledgments

The authors wish to thank the referees and editors for their rapid, sympathetic and extremely useful suggestions for improving the manuscript. BR wants to thank Prof. W. A. J. Luxemburg's Math 108 at Caltech in 1970–1971 for introducing him to the beauties of analysis.

References

[1] Cabrelli, C. A., Hare, K. E., Molter, U. M. (1997). Sums of Cantor sets. *Ergodic Theory Dynam. Systems.* 17(6):1299–1313.
[2] Cabrelli, C. A., Hare, K. E., Molter, U. M. (2002). Sums of Cantor sets yielding an interval. *J. Aust. Math. Soc.* 73(3):405–418.
[3] Fleron, J. F. (1994). A note on the history of the Cantor set and Cantor function. *Math. Mag.* 67(2): 136–140.
[4] Gorodetski, A., Northrup, S. (2015). On sums of nearly affine Cantor sets. Preprint, arxiv.org/pdf/1510.07008.pdf.
[5] Mendes, P., Oliveira, F. (1994). On the topological structure of the arithmetic sum of two Cantor sets. *Nonlinearity.* 7(2):329–343.
[6] Pawłowicz, M. (2013). Linear combinations of the classic Cantor set. *Tatra Mt. Math. Publ.* 56:47–60.
[7] Randolph, J. F. (1940). Distances between points of the Cantor set. *Amer. Math. Monthly.* 47(8):549–551.
[8] Schmeling, J., Shmerkin, P. (2010). On the dimension of iterated sumsets. In Barral, J., Seuret, S., eds. *Recent Developments in Fractals and Related Fields*, Appl. Numer. Harmon. Anal., Boston, MA: Birkhäuser Boston, Inc., pp. 55–72.
[9] Steinhaus, H. (1917). Mowa Własność Mnogości Cantora. *Wector*, 1–3. English translation in: STEINHAUS, H. D. (1985) *Selected Papers.*
[10] Utz, W. R. (1951). The distance set for the Cantor discontinuum. *Amer. Math. Monthly.* 58(6):407–408.

Department of Mathematics, University of Washington, Seattle, WA 98195
E-mail address: jathreya@math.uiuc.edu

Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801
E-mail address: reznick@math.uiuc.edu

Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801
E-mail address: tyson@math.uiuc.edu