

# Intrinsic Approximation on Cantor-like Sets, a Problem of Mahler

Ryan Broderick, Lior Fishman, Asaf Reich and Barak Weiss

July 2010

## Abstract

In 1984, Kurt Mahler posed the following fundamental question: How well can irrationals in the Cantor set be approximated by rationals in the Cantor set? Towards such a theory, we prove a Dirichlet-type theorem for this intrinsic diophantine approximation on Cantor-like sets. The resulting approximation function is analogous to that for  $\mathbb{R}^d$ , but with  $d$  being the Hausdorff dimension of the set, and logarithmic dependence on the denominator instead.

**Keywords:** Cantor sets, Diophantine Approximation, Dirichlet-type Theorem

## 1 Introduction

The diophantine approximation theory of the real line is classical, extensive, and essentially complete as far as characterizing how well real numbers can be approximated by rationals (Schmidt 1980 is a standard reference). The basic result on approximability of all reals is

**Theorem 1** (Dirichlet's Approximation Theorem). *For each  $x \in \mathbb{R}$ , and for any  $Q \in \mathbb{N}$ , there exist  $p, q \in \mathbb{N}$ ,  $q \leq Q$ , such that*

$$\left| x - \frac{p}{q} \right| < \frac{1}{qQ}.$$

**Corollary 1.** *For each  $x \in \mathbb{R}$ , there exist infinitely many  $p, q \in \mathbb{N}$  satisfying*

$$\left| x - \frac{p}{q} \right| < \frac{1}{q^2}.$$

It is a classical result of Hurwitz that Corollary 1 is false if 1 is replaced by any constant less than  $\frac{1}{\sqrt{5}}$ . Furthermore the rate cannot be improved for most irrationals, in the sense that if the exponent 2 is replaced by any real number greater than 2, the resulting set for which the inequality holds (VWA) is null. In addition, the subject of approximating points on fractals by rationals

has been extensively studied in recent years; see for example Kleinbock-Weiss (2005), Kristensen *et al.* (2006), and Fishman (2009) for the BA case, and Weiss (2001) and Kleinbock *et al.* (2004) for VWA. In 1984, Mahler posed a natural problem concerning not whether the metric properties of these diophantine sets are preserved under intersection with fractals, but studying approximation of irrationals in the fractal by rationals also in the fractal. As far as the authors are aware, nothing is known about such intrinsic diophantine approximation on fractals (though there is such research for algebraic varieties, for example Ghosh *et al.* 2010). We present a first step towards an intrinsic theory determining the analogous rate of approximation for the Cantor set  $C$  (and those constructed similarly). Namely, our result implies

**Corollary 2.** *For all  $x \in C$ , there exist infinitely many solutions  $p \in \mathbb{N}$ ,  $q \in \mathbb{N}$ ,  $\frac{p}{q} \in C$  to*

$$\left|x - \frac{p}{q}\right| < \frac{1}{q(\log_3 q)^{1/d}},$$

where  $d = \dim C$ .

We conjecture that the set of numbers  $x$  satisfying, for some  $\epsilon > 0$ ,

$$\left|x - \frac{p}{q}\right| < \frac{1}{q^{1+\epsilon}(\log_3 q)^{1/d}}$$

for infinitely many  $p$  and  $q$  is null with respect to the standard measure on  $C$ . This would be analogous to the classical setting, where VWA has zero Lebesgue measure. (See section 3.)

## 2 Theorem

Let  $C$  denote the Cantor-like set consisting of numbers in  $I = [0, 1]$  which can be written in base  $b > 2$  using only the digits in  $S \subset \{0, 1 \dots b-1\}$ , where  $|S| = a > 1$ . Obviously this is equivalent to partitioning  $I$  into  $b$  equal subintervals, only keeping those indexed by  $S$ , and successively continuing this on each remaining subinterval. (The usual middle-thirds set is given by  $b = 3$ ,  $S = \{0, 2\}$ .)

Throughout the paper we denote by  $\{x\}$  the fractional part of  $x$  and by  $[x]$  the integer part. To state the theorem we also associate to the set the number  $b_0$ , where  $b_0$  is the least integer such that  $C$  is invariant under  $x \mapsto \{b_0 x\}$  (as we shall see, this is simply equal to  $b$  except when the given definition of  $C$  is redundant in a certain sense).

**Theorem 2** (Dirichlet for Cantor sets). *For all  $x \in C$ , for every  $Q$  of the form  $b^n$ , there exists  $p/q \in \mathbb{Q} \cap C$ , such that*

$$\left|x - \frac{p}{q}\right| < \frac{1}{qQ},$$

where  $q \leq Q$  if  $C = I$  or  $q \leq b_0^{Q^d}$  otherwise, where  $d = \dim C$  and  $b = b_0^r$  for some  $r \in \mathbb{N}$ .

**Remark 1.** Notice it follows from the statement that for any  $Q \in \mathbb{N}$ , the same statement holds with the bound multiplied by  $b$  (by letting  $b^n \leq Q < b^{n+1}$ .)

*Proof.* First we need a characterization of the rationals in  $C$ :

**Lemma 1.** A rational number is in  $C$  if and only if it can be written either as a terminating base- $b$  expansion (left end points in the construction) or as

$$\frac{(\sum_{i=0}^{k+l-1} c_i b^{k+l-1-i} - \sum_{j=0}^{k-1} c_j b^{k-1-j})}{b^{k+l} - b^k}. \quad (1)$$

where  $l, k \in \mathbb{N}$ , and  $c_i \in S$ .

Equivalently (1) can also be expressed in terms of base- $b$  expansions, i.e.,

$$\frac{((c_0 c_1 \dots c_{k+l-1})_b - (c_0 c_1 \dots c_{k-1})_b)}{b^{k+l} - b^k}. \quad (2)$$

*Proof.* A rational in  $C$  has either a terminating  $b$ -ary expansion (consisting of digits from  $S$ ) or an eventually periodic one. If it is purely periodic of period  $l$ , it has the form

$$\sum_{n=1}^{n=\infty} \frac{b^{l-1}x_0 + b^{l-2}x_1 + \dots + x_{l-1}}{(b^l)^n} = \frac{b^{l-1}x_0 + b^{l-2}x_1 + \dots + x_{l-1}}{b^l - 1},$$

where the digits  $x_i$  are in  $S$ . If the rational is not purely periodic then one must insert some number  $k$  of initial zeros and then add the initial terminating expansion of length  $k$ , so we obtain the form

$$\frac{b^{l-1}x_0 + b^{l-2}x_1 + \dots + x_{l-1}}{(b^l - 1)b^k} + \frac{(b^l - 1)(b^{k-1}y_0 + b^{k-2}y_1 + \dots + y_{k-1})}{(b^l - 1)b^k}$$

for some digits  $y_i \in S$ . Rearranging this gives the result.  $\square$

Now let  $x \in C \setminus \mathbb{Q}$ . Given any  $n \in \mathbb{N}$ , let  $Q = b^n$ . Denote by  $M = (m_i)$  the semigroup of positive integer multiplication maps mod 1 leaving  $C$  invariant. We order the elements in increasing order from  $1 = m_0$ .

There are  $a^n$  possibilities for the first  $n$  digits in the  $b$ -ary expansion of  $\{qx\}$ . Consider the elements of  $C$  given by  $0, \{m_0x\}, \dots, \{m_{a^n-1}x\}$ . By the pigeonhole principle either there exist  $0 \leq q, q' < a^n$  such that  $\{m_qx\}$  and  $\{m_{q'}x\}$  have the same first  $n$  digits, or the same holds for some  $0 \leq q < a^n$  and 0. That is, they are in the same interval of the  $n$ -th stage of  $C$ 's construction. Assuming the former, it follows that  $|\{m_{q'}x\} - \{m_qx\}| < \frac{1}{Q}$ . Rewriting gives

$$|m_{q'}x - [m_{q'}x] - m_qx + [m_qx]| < \frac{1}{Q}.$$

Setting  $p = [m_{q'}x] - [m_qx]$  we get

$$\left| x - \frac{p}{m'_{q'} - m_q} \right| < \frac{1}{(m'_{q'} - m_q)Q}.$$

If one of the above values is 0 rather than an  $m_i$  the calculation is trivial. If  $M$  is generated by more than one element, by Furstenberg's result (1967) that the only infinite closed subset of  $\mathbb{R}/\mathbb{Z}$  invariant under  $M$  is  $\mathbb{R}/\mathbb{Z}$  itself,  $C = I$ . Thus  $C$  is invariant under all integers, i.e.  $M = \mathbb{N}$ , so  $m_{q'} - m_q \leq a^n = Q$  and Dirichlet's original function is recovered.

If  $M$  has a single generator, denote it  $b_0$ , and then  $b = b_0^r$  for some  $r$ , and  $a = a_0^r$ . In this case  $C$  is also the set constructed in the corresponding way for  $b_0$ , and some  $S_0$  (which can be defined as the set satisfying that when written as base  $b_0$  digits,  $S = S_0^r$ , the  $r$ -fold concatenations of elements of  $S_0$ .) Then  $m_{Q-1} \leq b_0^{a^n-1}$ . Since  $m_{q'} - m_q = b_0^k(b_0^d - 1)$  for some integers  $k, d$ , following (2) it suffices to observe that  $p = \lfloor b_0^{k+d}x \rfloor - \lfloor b_0^kx \rfloor$ . Thus  $\frac{p}{m_{q'} - m_q} \in C$ , and  $m_{q'} - m_q < b_0^{a^n} = b_0^{Q^d}$ .  $\square$

**Corollary 2.** *For all  $x \in C$ , there exist infinitely many solutions  $p \in \mathbb{N}$ ,  $q \in \mathbb{N}$ ,  $\frac{p}{q} \in C$  to*

$$\left| x - \frac{p}{q} \right| < \frac{1}{q(\log_{b_0} q)^{1/d}}.$$

Notice that the approximation function's asymptotic behavior gets better as  $d \rightarrow 0$ , even though the first such  $q$  whose existence we prove can tend to infinity as  $b$  does.

### 3 Further investigation

The obvious next step would be to check whether this is the "right" rate of approximation for  $C$ . For this purpose, we make the following definitions: whereas the usual set BA consists of all reals  $x$  such that

$$\left| \frac{p}{q} - x \right| > \frac{c(x)}{q^2} \quad \forall p/q \in \mathbb{Q},$$

we define, relative to the approximation function proven above, the intrinsic BA numbers to be all  $x \in C$  satisfying

$$\left| \frac{p}{q} - x \right| > \frac{c(x)}{q(\log_{b_0} q)^{1/d}} \quad \forall p/q \in \mathbb{Q} \cap C,$$

and in comparison to the classical VWA definition

$$\left| \frac{p}{q} - x \right| < \frac{1}{q^{2+\epsilon(x)}}$$

for infinitely many  $p/q$ , the intrinsic VWA numbers are those in  $C$  with

$$\left| \frac{p}{q} - x \right| < \frac{1}{q^{1+\epsilon(x)}}$$

for infinitely many  $p/q \in C$  instead<sup>1</sup>. Then, one could say this function is correct if one proved that intrinsic BA is nonempty (and perhaps of full dimension), or that intrinsic VWA has zero Hausdorff measure. These results appear difficult to achieve, however, without having a deeper understanding of the rationals in  $C$ . For the real line, there exists a large arsenal of useful information regarding distribution properties of rationals, their quantity within bounds on  $q$ , etc., whereas for  $C$  nothing is even known about which denominators can appear - in reduced form, of course; the expression (2) is not useful here since it is not reduced. In fact Mahler points out basically the same fundamental difficulty. To obtain the desired results, we believe a major new piece of information such as knowing exactly which denominators appear in  $C$  within given bounds (i.e. their "density" in the same sense as the classical number-theoretical study of the density of the primes), or knowing how these rationals "repel" each other as a function of  $q$ , will be necessary. Until then, however, we make the following conjecture (which seems to be supported by our experimentation with some computer data):

**Conjecture 1.** *Let  $\Phi(i, j) = \#\{\frac{p}{q} \in C, \gcd(p, q) = 1, i \leq q \leq j\}$ . Then  $\Phi(b_0^n, b_0^{n+1}) = O(a_0^{(1+\epsilon)n})$ , for all  $\epsilon > 0$ .*

This would imply that Intrinsic VWA has zero  $d$ -Hausdorff measure.

---

<sup>1</sup>This set is equivalent to the one defined by a similar inequality with a  $(\log_{b_0} q)^{1/d}$  factor in the denominator, so the definition is analogous to the traditional one in that we simply divide our approximation function by  $q^\epsilon$ . We remove the logarithmic factor for simplicity.

## 4 References

- L. Fishman, Schmidt's game on fractals, *Israel J. Math.* **171** (2009), no. 1, 7792.
- H. Furstenberg, Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation, *Math. Systems Theory* **1** (1967), 1-49.
- A. Ghosh, A. Gorodnik, and A. Nevo, Diophantine approximation and automorphic spectrum, Preprint, arXiv:1007.0593v1.
- D. Kleinbock, E. Lindenstrauss and B. Weiss, On fractal measures and diophantine approximation, *Selecta Math.* **10** (2004), 479523.
- S. Kristensen, R. Thorn, and S.L. Velani, Diophantine approximation and badly approximable sets, *Advances in Math.* **203** (2006), 132–169.
- D. Kleinbock and B. Weiss, Badly approximable vectors on fractals, *Israel Journal of Mathematics*, **149** (2005), 137-170.
- K. Mahler, Some suggestions for further research, *Bulletin of the Australian Mathematical Society*, **29** (1984), 101-108.
- W.M. Schmidt, *Diophantine approximation*, Lecture Notes in Mathematics, vol. 785, Springer-Verlag, Berlin, 1980.
- B. Weiss, Almost no points on a Cantor set are very well approximable, *Proc. R. Soc. Lond. A* **457** (2001), 949–952.