Diophantine approximation and Cantor sets

Yann Bugeaud

Abstract. *We provide an explicit construction of elements of the middle third Cantor set with any prescribed irrationality exponent. This answers a question posed by Kurt Mahler.*

1. Introduction

In Section 2 of his paper *Some suggestions for further research*, Mahler [12] posed the following question:

How close can irrational elements of Cantor's set be approximated by rational numbers
(i) in Cantor's set, and
(ii) by rational numbers not in Cantor's set ?

Here, and throughout the present note, *Cantor's set* is the middle third Cantor set, that is, the set of all real numbers $c_1 3^{-1} + c_2 3^{-2} + \cdots + c_i 3^{-i} + \cdots$ with $a_i = 0$ or 2 for every $i \geq 1$. We denote it by K.

In other words, Mahler asked (see also Problem 35 in [5]) whether there are elements of Cantor's set with any prescribed *irrationality exponent*, where the irrationality exponent $\mu(\xi)$ of an irrational number ξ is the supremum of the real numbers μ such that

$$\left| \xi - \frac{p}{q} \right| < \frac{1}{q^{\mu}}$$

has infinitely many solutions in rational numbers p/q.

The irrationality exponent of every irrational number is greater than or equal to 2 and it is precisely equal to 2 for almost all real numbers (with respect to the Lebesgue measure) ; see Section 1 of [15]. As a first step towards Mahler's question, Weiss [17] established that the irrationality exponent of almost all elements of K (with respect to the standard measure supported on K) is equal to 2, as well. Furthermore, metric number theory was used in [7, 8, 10] to show that K contains badly approximable numbers, that is, numbers ξ for which $|\xi - p/q| > c/q^2$ holds for every rational p/q and some positive constant c.

Recently, Levesley, Salp and Velani [11] used metric number theory to bring some new light on Mahler's problem. They established the existence of elements in K that can be approximated at any prescribed order by rational numbers whose denominators

2000 *Mathematics Subject Classification :* 11J82.

are powers of 3. This gives a very satisfactory answer to the 3-adic analogue of Mahler's question. Furthermore, they complemented this result by constructing explicit elements of K having any prescribed irrationality exponent greater than or equal to $(3+\sqrt{5})/2$.

This leaves open the problem whether there are elements in K with any prescribed irrationality exponent between 2 and $(3+\sqrt{5})/2$. The purpose of the present note is to give a positive answer to this question and to establish a much stronger assertion, namely that there are elements in K with, roughly speaking, any prescribed approximation order. In particular, we construct uncountably many explicit elements of K with any prescribed irrationality exponent at least equal to 2.

2. Results

Let $\Psi : \mathbf{R}_{\geq 1} \to \mathbf{R}_{>0}$ be a non-increasing function and set

$$\mathcal{K}(\Psi) := \left\{ \xi \in \mathbf{R} : \left| \xi - \frac{p}{q} \right| < \Psi(q) \ \text{ for infinitely many rational numbers } \frac{p}{q} \right\}.$$

In 1931, Jarník [6] used the theory of continued fractions to construct explicit examples of real numbers in $\mathcal{K}(\Psi)$ which do not belong to any set $\mathcal{K}(c\Psi)$ with $0 < c < 1$, provided that the function Ψ satisfies $\Psi(x) = o(x^{-2})$. This means that the set

$$\mathrm{Exact}(\Psi) := \mathcal{K}(\Psi) \setminus \bigcup_{m \geq 2} \mathcal{K}\big((1-1/m)\Psi\big)$$

of real numbers approximable to order Ψ and to no better order is non-empty. A strong form of Mahler's question can be formulated as:

Is the intersection $\mathrm{Exact}(\Psi) \cap K$ *always non-empty ?*

Under a slight additional assumption on the approximation function Ψ, we give a positive answer to a slightly weaker question. To state our main result, we consider general 'missing digits' sets as in Section 7 of [11]. Let $b \geq 3$ be an integer and let $J(b)$ be a proper subset of $\{0, 1, \ldots, b-1\}$ with at least two elements. We denote by $K_{J(b)}$ the set of real numbers in the unit interval $[0, 1]$ whose b-ary expansion consists exclusively of digits in $J(b)$.

Theorem 1. *Let* $\Psi : \mathbf{R}_{\geq 1} \to \mathbf{R}_{>0}$ *be a function such that* $x \mapsto x^2\Psi(x)$ *is non-increasing and tends to* 0. *For any positive real number* c *less than* $1/b$, *the set*

$$\big(\mathcal{K}(\Psi) \setminus \mathcal{K}(c\Psi)\big) \cap K_{J(b)}$$

is uncountable.

Choosing $b = 3$, $J(b) = \{0, 2\}$ and taking for Ψ the function $x \mapsto x^{-\tau}(\log 2x)^{-2}$, with $\tau \geq 2$, we derive at once from Theorem 1 a satisfactory answer to Mahler's question. In all what follows, $\lfloor x \rfloor$ denotes the greatest integer less than or equal to the real number x.

Theorem 2. *Let τ be a real number with $\tau \geq 2$. Cantor's set contains uncountably many elements whose irrationality exponent is equal to τ. Let $\lambda > 0$ and set $n_0(\tau, \lambda) = 1 + \max\{0, \lfloor (\log(2/\lambda))/\log \tau \rfloor\}$. Then, the real number*

$$\xi_{\tau,\lambda} := 2 \sum_{n \geq n_0(\tau,\lambda)} 3^{-\lfloor \lambda \tau^n \rfloor} \tag{2.1}$$

is an element of K with $\mu(\xi_{\tau,\lambda}) = \tau$. Furthermore, $\xi_{2,\lambda}$ is a badly approximable real number.

Remark 1. For a real number $\tau > 2$, Levesley, Salp and Velani [11] studied the rational approximation of the irrational number $\xi_{\tau,1}$, that from now on we simply denote by ξ_τ. Estimating the distance between ξ_τ and its rational approximants $\sum_{n=1}^{N} 3^{-\lfloor \tau^n \rfloor}$, $N \geq 1$, they deduced straightforwardly that $\mu(\xi_\tau) \geq \tau$. To bound $\mu(\xi_\tau)$ from above, they used triangular inequalities and obtained the upper estimate $\mu(\xi_\tau) \leq \max\{\tau, (2\tau - 1)/(\tau - 1)\}$. This gives $\mu(\xi_\tau) = \tau$ for $\tau \geq (3 + \sqrt{5})/2$, but, unfortunately, not the exact value of $\mu(\xi_\tau)$ when $2 < \tau < (3 + \sqrt{5})/2$. Statements of this type are reminiscent of numerous results in transcendence theory ; see, for example, Theorems 7.7 and 8.8 in [5]. To prove our theorems, we develop an alternative approach ; namely, we construct explicitly the continued fraction expansion and the expansion in base 3 of suitable real numbers ξ. The former gives us the irrationality exponent of ξ, and the latter ensures us that ξ lies in Cantor's set.

Remark 2. A more general question than Mahler's one consists in studying how the elements of K are approximable by algebraic numbers of bounded degree. Even the extension of Weiss' result [17] is not known ; see Kristensen [9] for partial results.

Remark 3. It follows from Ridout's Theorem [14] that all the real numbers constructed in the proof of Theorem 1 are transcendental. Furthermore, the main result of [1] yields a transcendence measure for these numbers, implying that they are either S- or T-numbers in Mahler's classification.

Remark 4. Under the additional assumptions that $x \mapsto x^2 \Psi(x)$ is non-increasing and that the sum $\sum_{x \geq 1} x \Psi(x)$ converges, the Hausdorff dimension of the set $\mathrm{Exact}(\Psi)$ has been determined in [4]: it is equal to the Hausdorff dimension of $\mathcal{K}(\Psi)$. A challenging and difficult open problem would be, first, to show that the set $\mathrm{Exact}(\Psi) \cap K_{J(b)}$ is non-empty (that may even be not true), and, second, to compute, under these assumptions, its Hausdorff dimension. It is reasonable to expect that the Hausdorff dimension of $\mathcal{K}(\Psi) \cap K_{J(b)}$ is equal to the product of the Hausdorff dimensions of the sets $\mathcal{K}(\Psi)$ and $K_{J(b)}$.

Remark 5. Let $\mathcal{A} = \{1, 3, 3^2, \ldots\}$ be the set of powers of 3. For a non-increasing function $\Psi : \mathbf{R}_{\geq 1} \to \mathbf{R}_{>0}$ set

$$\mathcal{K}_{\mathcal{A}}(\Psi) := \left\{ \xi \in \mathbf{R} : \left| \xi - \frac{p}{q} \right| < \Psi(q) \text{ for infinitely many rational numbers } \frac{p}{q} \text{ with } q \in \mathcal{A} \right\}.$$

Theorem 1 asserts that, if $x \mapsto x^2 \Psi(x)$ is non-increasing and tends to 0, then, for any positive c less than 1/3, the set $\mathcal{K}_{\mathcal{A}}(\Psi) \setminus \mathcal{K}_{\mathcal{A}}(c\Psi)$ is non-empty. This result is new and cannot be obtained via the metric theory as used in [11], see [3] and the Remark at the end of Section 3 of [11].

Remark 6. We refer the reader to [2] for the definitions of the notions discussed in this remark. Let p be an odd integer and ℓ be a positive integer such that $p/2^\ell > 2$, or let $(p, \ell) = (2, 0)$. Define the sequence $\mathbf{x} = (x_n)_{n\geq 0}$ by setting $x_n = 1$ if n is of the form $2^h p^a$ for some integers $a \geq 1$ and $h = 0, \ldots, \ell$, and setting $x_n = 0$ otherwise. Then, $\mathbf{x}$ is an *automatic sequence* and the irrationality exponent of the automatic real number $\sum_{n\geq 0} 3^{-x_n}$ is equal to $p/2^\ell$. A slightly modified construction shows that there exist automatic real numbers with any prescribed *rational* irrationality exponent. This is justified at the end of Section 3. Furthermore, since the subword complexity of any real number $\xi_{\tau,\lambda}$ defined in (2.1) is sub-linear, Theorem 2 implies that there exist real numbers of sub-linear subword complexity with any prescribed irrationality exponent.

3. Proofs

The key tool for our constructive proofs is the Folding Lemma, which we deduce at once from Section 2 of [13] ; see also [16].

Folding Lemma. *Let t be a positive integer and r/s be a rational number written in lowest terms. If $r/s = [0; 1, 1, a_3, \ldots, a_{h-1}, a_h]$ with $h \geq 4$ and $a_h \geq 2$, then*

$$\frac{r}{s} + \frac{(-1)^h}{ts^2} = \frac{rst + (-1)^h}{ts^2} = [0; 1, 1, a_3, \ldots, a_{h-1}, a_h, t-1, 1, a_h - 1, a_{h-1}, \ldots, a_3, 2],$$

for $t \geq 2$, and

$$\frac{r}{s} + \frac{(-1)^h}{s^2} = \frac{rs + (-1)^h}{s^2} = [0; 1, 1, a_3, \ldots, a_{h-1}, a_h + 1, a_h - 1, a_{h-1}, \ldots, a_3, 2].$$

We now explain how to use repeatedly the Folding Lemma in order to construct real numbers satisfying the conclusion of Theorem 1. Without any loss of generality, we assume that $J(b) = \{0, 1\}$. Let $v \geq 4$ be an integer such that $x^2\Psi(x) \leq b^{-2}$ for $x \geq b^v$. Observe that there is a positive integer r, coprime with b, such that the rational r/b^v lies in the open real interval with endpoints $[0; 1, 1, 1, 2]$ and $[0; 1, 1, 1, 3]$. Consequently, the continued fraction of r/b^v reads

$$\frac{r}{b^v} = [0; 1, 1, a_3, \ldots, a_{h-1}, a_h],$$

with $h \geq 4$ and $a_h \geq 2$.

Let $\mathbf{u} = (u_k)_{k\geq 1}$ be a sequence of positive integers. Applying the Folding Lemma first with $t = b^{u_1}$, then with $t = b^{u_2}$, and so on, we get a sequence $(d_k)_{k\geq 1}$ of integers such that the real number

$$\xi_{\mathbf{u}} = \frac{r}{b^v} + \frac{(-1)^h}{b^{u_1+2v}} - \frac{1}{b^{u_2+2(u_1+2v)}} - \cdots - \frac{1}{b^{u_k+2(u_{k-1}+\cdots+2^{k-2}u_1+2^{k-1}v)}} - \cdots$$

satisfies

$$\frac{1}{b^{u_{k+1}+2v_k}} \le \left|\xi_{\mathbf{u}} - \frac{d_k}{b^{v_k}}\right| \le \frac{1+b^{-v_k}}{b^{u_{k+1}+2v_k}}, \quad \text{for } k \ge 1, \tag{3.1}$$

where we have set

$$v_k = u_k + 2u_{k-1} + \cdots + 2^{k-1}u_1 + 2^k v, \quad \text{for } k \ge 1.$$

Here, d_k is the nearest integer to $\xi_{\mathbf{u}} b^{v_k}$, for $k \ge 1$.

We consider the sequence $\mathbf{u}$ defined inductively by $u_1 = 1$ and

$$1 + b^{-v_k} < b^{u_{k+1}} b^{2v_k} \Psi(b^{v_k}) \le b + b^{1-v_k}, \quad \text{for } k \ge 1. \tag{3.2}$$

Since $x \mapsto x^2\Psi(x)$ is non-increasing and tends to 0, the sequence $\mathbf{u}$ is unbounded and non-decreasing.

It follows from (3.1) and (3.2) that

$$\frac{\Psi(b^{v_k})}{b + b^{1-v_k}} \le \left|\xi_{\mathbf{u}} - \frac{d_k}{b^{v_k}}\right| < \Psi(b^{v_k}), \quad \text{for } k \ge 1. \tag{3.3}$$

We have now to show that the sequence $(d_k/b^{v_k})_{k\ge1}$ comprises all the best rational approximations to $\xi_{\mathbf{u}}$. Write

$$\xi_{\mathbf{u}} = [0; a_1, a_2, \ldots], \qquad \frac{p_j}{q_j} = [0; a_1, a_2, \ldots, a_j], \quad j \ge 1,$$

for the continued fraction expansion of $\xi_{\mathbf{u}}$ and its convergents, and recall the classical inequalities (see e.g., Section 1.2 from [5])

$$\frac{1}{(a_j+2)q_{j-1}^2} < \left|\xi_{\mathbf{u}} - \frac{p_{j-1}}{q_{j-1}}\right| < \frac{1}{a_j q_{j-1}^2}, \quad \text{for } j \ge 2. \tag{3.4}$$

Let $m \ge 2$ be an integer. It is easily checked that

$$a_{2^m(h+1)} = b^{u_{m+1}} - 1 \quad \text{and} \quad \frac{p_{2^m(h+1)-1}}{q_{2^m(h+1)-1}} = \frac{d_m}{b^{v_m}}. \tag{3.5}$$

By construction, all the partial quotients a_j with $2^{m-1}(h+1) < j < 2^m(h+1)$ are less than or equal to $b^{u_{m-1}} - 1$. Then, (3.4) yields

$$\left|\xi_{\mathbf{u}} - \frac{p_{j-1}}{q_{j-1}}\right| > \frac{1}{(a_j+2)q_{j-1}^2} \ge \frac{1}{(b^{u_{m-1}}+1)q_{j-1}^2}, \tag{3.6}$$

for $2^{m-1}(h+1) < j < 2^m(h+1)$. Let ε be a positive real number such that $c < 1/(b+2\varepsilon)$. Since $x \mapsto x^2\Psi(x)$ is non-increasing and $\mathbf{u}$ is non-decreasing and unbounded, we infer from (3.2) that

$$\begin{aligned} (b^{u_{m-1}}+1)q_{j-1}^2\Psi(q_{j-1}) &\le (b^{u_{m-1}}+1)b^{2v_{m-1}}\Psi(b^{v_{m-1}}) \\ &\le (b+\varepsilon)(b^{u_{m-1}}+1)b^{-u_m} \le b + 2\varepsilon, \end{aligned}$$

hence, by (3.6),

$$\left|\xi_{\mathbf{u}} - \frac{p_{j-1}}{q_{j-1}}\right| > \frac{\Psi(q_j)}{b+2\varepsilon}$$

for $2^{m-1}(h+1) < j < 2^m(h+1)$, provided that m is sufficiently large. Combined with (3.3) and (3.5), this shows that

$$\left|\xi_{\mathbf{u}} - \frac{p_j}{q_j}\right| \geq \frac{\Psi(q_j)}{b+2\varepsilon} > c\Psi(q_j) \tag{3.7}$$

holds for every sufficiently large integer j.

It now follows at once from (3.3) and (3.7) that $\xi_{\mathbf{u}}$ lies in $\mathcal{K}(\Psi) \setminus \mathcal{K}(c\Psi)$. The real number

$$\frac{r}{b^v} + \frac{(-1)^h}{b^{u_1+2v}} - \xi_{\mathbf{u}} = \sum_{k\geq 2} \frac{1}{b^{v_k}}$$

belongs to $\mathcal{K}(\Psi) \setminus \mathcal{K}(c\Psi)$ and is written only with the digits 0 and 1 in base b. Let $\mathbf{u}' = (u'_k)_{k\geq 1}$ be a sequence defined from the above sequence $\mathbf{u}$ by setting $u'_1 = 1$, $u'_{2k} = u_k$ and $u'_{2k+1} \in \{1,2\}$ for $k \geq 1$. Then, the same proof yields that the real number $rb^{-v} + (-1)^h b^{-u'_1-2v} - \xi_{\mathbf{u}'}$ lies in $\mathcal{K}(\Psi) \setminus \mathcal{K}(c\Psi)$ and is written only with the digits 0 and 1 in base b. Thus, we have constructed uncountably many real numbers with the required property. This completes the proof of Theorem 1.

About the second assertion of Theorem 2, we content ourselves to say that it is sufficient to perform the above construction with $b = 3$, $v = \lfloor \lambda\tau^{3+n_0(\tau,\lambda)} \rfloor$ and the sequence $\mathbf{u}$ defined by $u_k = \lfloor \lambda\tau^{k+3+n_0(\tau,\lambda)} \rfloor - 2\lfloor \lambda\tau^{k+2+n_0(\tau,\lambda)} \rfloor$ if $\tau > 2$ and $k \geq 1$. The resulting real number is equal to $\xi_{\tau,\lambda}$ plus some rational number, and its irrationality exponent is equal τ, hence, $\mu(\xi_{\tau,\lambda}) = \tau$. For $\tau = 2$, since $\lfloor \lambda 2^{k+1} \rfloor = 2\lfloor \lambda 2^k \rfloor + \varepsilon_k$, with $\varepsilon_k \in \{0,1\}$, we have to apply repeatedly the Folding Lemma with $t = 3^{\varepsilon_n}$, and the partial quotients in the continued fraction expansion of the resulting number remain bounded ; thus, $\xi_{2,\lambda}$ is a badly approximable number.

To conclude, we briefly justify the claim asserted in Remark 6. Let p and q be coprime integers with $p > q \geq 1$ and $p/q > 2$. Let $a \geq 4$ and ℓ be integers such that

$$\frac{2}{q} < \frac{p^{a-1}}{2^\ell} < \frac{p}{q^2}. \tag{3.8}$$

We start the construction with a rational number $r/3^{p^a}$ in the open real interval with endpoints $[0; 1, 1, 1, 2]$ and $[0; 1, 1, 1, 3]$. Applying the Folding Lemma ℓ times with $t = 1$, we reach a rational with denominator $3^{2^\ell p^a}$. Then, we apply the Folding Lemma with $t = 3^{(qp^{a-1}-2^{\ell+1})p^a}$ and with $t = 3^{(p^a-2qp^{a-1})p^a}$. This gives rational numbers with denominators $3^{qp^{2a-1}}$ and $3^{p^{2a}}$. We continue with ℓ further applications of the Folding Lemma with $t = 1$, one with $t = 3^{(qp^{a-1}-2^{\ell+1})p^{2a}}$ and one with $t = 3^{(p^a-2qp^{a-1})p^{2a}}$, and so on. By (3.8), the irrationality exponent of the resulting irrational number ξ is equal to p/q. Define the sequence $\mathbf{y} = (y_n)_{n\geq 0}$ by setting $y_n = 1$ if n is of the form $2^j p^{ha}$ or

$qp^{(h+1)a-1}$ for some integers $h \geq 1$ and $j = 0, \ldots, \ell$, and setting $y_n = 0$ otherwise. Then, $\mathbf{y}$ is an *automatic sequence* and the automatic real number $\sum_{n \geq 0} 3^{-y_n}$ is equal to ξ plus some rational number. Consequently, its irrationality exponent is precisely p/q.

Acknowledgements. I am pleased to thank the referee for his very careful reading.

References

[1] B. Adamczewski and Y. Bugeaud, *Mesures de transcendance et aspects quantitatifs de la méthode de Thue–Siegel–Roth–Schmidt*. Preprint.

[2] J.-P. Allouche and J. O. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.

[3] V. Beresnevich, H. Dickinson and S. L. Velani, *Sets of exact "logarithmic order" in the theory of Diophantine approximation*, Math. Ann. 321 (2001), 253–273.

[4] Y. Bugeaud, *Sets of exact approximation order by rational numbers*, Math. Ann. 327 (2003), 171–190.

[5] Y. Bugeaud, Approximation by algebraic numbers. Cambridge Tracts in Mathematics 160, Cambridge, 2004.

[6] V. Jarník, *Über die simultanen Diophantische Approximationen*, Math. Z. 33 (1931), 505–543.

[7] D. Kleinbock, E. Lindenstrauss, and B. Weiss, *On fractal measures and Diophantine approximation*, Selecta Math. 10 (2004), 479–523.

[8] D. Kleinbock and B. Weiss, *Badly approximable vectors on fractals*, Israel J. Math. 149 (2005), 137–170.

[9] S. Kristensen, *Approximating numbers with missing digits by algebraic numbers*, Proc. Edinb. Math. Soc. 49 (2006), 657–666.

[10] S. Kristensen, R. Thorn and S. Velani, *Diophantine approximation and badly approximable sets*, Adv. Math. 203 (2006), 132–169.

[11] J. Levesley, C. Salp and S. L. Velani, *On a problem of K. Mahler: Diophantine approximation and Cantor sets*, Math. Ann. 338 (2007), 97–118.

[12] K. Mahler, *Some suggestions for further research*, Bull. Austral. Math. Soc. 29 (1984), 101–108.

[13] A. J. van der Poorten and J. Shallit, *Folded continued fractions*, J. Number Theory 40 (1992), 237–250.

[14] D. Ridout, *Rational approximations to algebraic numbers*, Mathematika 4 (1957), 125–131.

[15] W. M. Schmidt, Approximation to algebraic numbers. Monographie de l'Enseignement Mathématique 19, Genève, 1971.

[16] J. Shallit, *Simple continued fractions for some irrational numbers*, J. Number Theory 11 (1979), 209–217.

[17] B. Weiss, *Almost no points on a Cantor set are very well approximable*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. 457 (2001), 949–952.

Yann Bugeaud
Université Louis Pasteur
U. F. R. de mathématiques
7, rue René Descartes
67084 STRASBOURG (FRANCE)

`bugeaud@math.u-strasbg.fr`