SELF-AFFINE SETS IN ANALYTIC CURVES AND ALGEBRAIC SURFACES

DE-JUN FENG AND ANTTI KÄENMÄKI

Abstract. We characterize analytic curves that contain non-trivial self-affine sets. We also prove that compact algebraic surfaces cannot contain non-trivial self-affine sets.

1. Introduction

Self-similar and self-affine sets are among the most typical and important fractal objects; see e.g. [2]. They can be generated by the so-called iterated function systems; see Section 2. Although these sets can be very irregular as one expects, they often have very rigid geometric structure.

It is not surprising that typical non-flat smooth manifolds do not contain any non-trivial self-similar or self-affine set. For instance, circles are such examples. To see this, suppose to the contrary that a circle C contains a non-trivial self-affine set E. Let f be a contractive affine map in the defining iterated function system of E. Then $f(E) \subset E$ and thus $f(E)$ is contained in both C and $f(C)$. However, since $f(C)$ is an ellipse with diameter strictly smaller than that of C, the intersection of $f(C)$ and C contains at most two points. This is a contradiction since $f(E)$ is an infinite set.

The above general phenomena was first clarified by Mattila [6] in the self-similar case. He proved that a self-similar set E satisfying the open set condition either lies on an m-dimensional affine subspace or $\mathcal{H}^t(E \cap M) = 0$ for every m-dimensional C^1-submanifold of $\mathbb{R}^n$. Here t is the Hausdorff dimension of E and $\mathcal{H}^t$ is the t-dimensional Hausdorff measure. This result was later generalized to self-conformal sets in [4, 5, 7]. As a related work, Bandt and Kravchenko [1] showed that if E is a self-similar set which spans $\mathbb{R}^n$ and $x \in E$, then there does not exist a tangent hyperplane of E at x.

As an easy consequence of the result of Mattila or that of Bandt and Kravchenko, an analytic planar curve does not contain any non-trivial self-similar set unless it is a straight line segment. In a private communication, Mattila asked which kind of analytic planar curves

---

Date: December 21, 2016.

2000 *Mathematics Subject Classification.* Primary 28A80; Secondary 32C25, 51H30.

Key words and phrases. Self-affine set, analytic curve, algebraic surface.

can contain a non-trivial self-affine set. The main purpose of this article is to answer this question.

We first remark that any closed parabolic arc is a self-affine set. This interesting fact was first pointed out by Bandt and Kravchenko [1]. In that paper, they considered self-affine planar curves consisting of two pieces $E = f_1(E) \cup f_2(E)$. They showed that if a certain condition on the eigenvalues of f_1 and f_2 holds, then the curve E is differentiable at all except for countably many points. They also introduced a stronger condition on the eigenvalues which guarantees the curve E to be continuously differentiable. This result implies that there exist many continuously differentiable self-affine curves. However, Bandt and Kravchenko furthermore showed that self-affine curves cannot be very smooth: the only simple C^2 self-affine planar curves are parabolic arcs and straight lines.

In our main result, instead of curves that are itself self-affine, we consider general self-affine sets and examine when they can be contained in an analytic curve.

Theorem A. *An analytic curve in $\mathbb{R}^n$, $n \geq 2$, which cannot be embedded in a hyperplane contains a non-trivial self-affine set if and only if it is an affine image of $\eta\colon [c,d] \to \mathbb{R}^n$, $\eta(t) = (t, t^2, \ldots, t^n)$, for some $c < d$.*

The above result gives a complete answer to the question of Mattila: the only analytic planar curves that contain non-trivial self-affine sets are parabolic arcs and straight line segments. As explained by Mattila, the question is related to the study of singular integrals and self-similar sets in Heisenberg groups. In such groups, self-similar sets are self-affine in the Euclidean metric. From the singular integral theory point of view, it is thus important to understand when a self-affine set is contained in an analytic manifold.

Concerning manifolds, we study an analogue of Mattila's question. We examine which kind of algebraic surfaces can contain self-affine sets. Our result shows that this cannot happen on compact surfaces.

Theorem B. *A compact algebraic surface does not contain non-trivial self-affine sets.*

It is easy to see that non-compact surfaces, such as paraboloids, can contain non-trivial self-affine sets; see Example 4.1. To finish the article, we introduce in Proposition 4.3 a sufficient condition for the inclusion of a self-affine set in an algebraic surface.

2. Preliminaries

In this section, we introduce the basic concepts to be used throughout in the article. A mapping $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is *affine* if $f(x) = Tx + c$ for all $x \in \mathbb{R}^n$, where T is a $n \times n$ matrix and $c \in \mathbb{R}^n$. The matrix T is called a *linear part* of f. It is easy to see that an affine map is invertible if and only if its linear part is non-singular. A mapping $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is *strictly contractive* if $|f(x) - f(y)| < |x - y|$ for all $x, y \in \mathbb{R}^n$. Note that an affine mapping f is strictly contractive if and only if its linear part T has operator norm $\|T\|$ strictly less than 1. A non-empty compact set $E \subset \mathbb{R}^n$ is called *self-affine* if $E = \bigcup_{i=1}^{\ell} f_i(E)$, where $\{f_i\}_{i=1}^{\ell}$ is an *affine iterated function system (IFS)*, i.e. a finite collection of strictly contractive invertible affine maps $f_i\colon \mathbb{R}^n \to \mathbb{R}^n$; see [3]. Moreover, E is called *self-similar* if all the f_i's are similitudes. We say that a self-affine set is *non-trivial* if it is not a singleton.

If $a < b$, then a non-constant continuous function $\gamma\colon [a,b] \to \mathbb{R}^n$ is called a *curve*. We denote the set $\gamma([a,b]) \subset \mathbb{R}^n$ by $\mathrm{Img}(\gamma)$ and refer to it also as a *curve*. By saying that a curve γ contains a set A we obviously mean that $A \subset \mathrm{Img}(\gamma)$. A curve γ is *simple* if $\gamma(s) \neq \gamma(t)$ for $a \leq s < t < b$. We say that a curve $\gamma\colon [a,b] \to \mathbb{R}^n$, $\gamma(t) = (x_1(t), \ldots, x_n(t))$, is *analytic* if $x_i\colon [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ and real analytic on (a,b) for all $i \in \{1, \ldots, n\}$. Recall that a function is real analytic on an open set $U \subset \mathbb{R}$ if, at any point $t \in U$, it can be represented by a convergent power series on some interval of positive radius centered at t. Similarly, if x_i's are C^k functions for some $k \in \mathbb{N}$, then the curve γ is called C^k *curve*. The k-th derivative of a C^k curve γ is $\gamma^{(k)}(t) = (x_1^{(k)}(t), \ldots, x_n^{(k)}(t))$. If $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is an invertible affine mapping and $\gamma\colon [a,b] \to \mathbb{R}^n$ is a curve, then $f \circ \gamma$ is the *affine image* of the curve.

Let $P\colon \mathbb{R}^n \to \mathbb{R}$ be a non-constant polynomial with real coefficients. The set

$$S(P) = \{x \in \mathbb{R}^n : P(x) = 0\}$$

is called an *algebraic surface*. The *degree* of P, denoted by $\deg(P)$, is the highest degree of its terms, when P is expressed in canonical form. The degree of a term is the sum of the exponents of the variables that appear in it.

3. Self-affine sets and analytic curves

In this section, we prove Theorem A. Our arguments are inspired by the proof of [1, Theorem 3(i)]. We will first show that an affine image of $\eta\colon [c,d] \to \mathbb{R}^n$, $\eta(t) = (t, t^2, \ldots, t^n)$, contains a non-trivial self-affine set. This follows immediately from the following lemma.

Lemma 3.1. *If $\eta\colon [c,d] \to \mathbb{R}^n$, $\eta(t) = (t, t^2, \ldots, t^n)$, then $\mathrm{Img}(\eta)$ is a non-trivial self-affine set for all $c < d$.*

Proof. Let

$$0 < \lambda < (2^n \sqrt{n} \max\{(2|c|+1)^n, (|c|+|d|+1)^n\})^{-1} < 1$$

and choose $t_1, \ldots, t_\ell \in [c,d]$ with $\ell \in \mathbb{N}$ such that the self-similar set of $\{x \mapsto \lambda(x-c)+t_i\}_{i=1}^{\ell}$ is $[c,d]$. Write $c_{i,k,j} = \binom{k}{j}(\frac{t_i}{\lambda} - c)^{k-j}$ and observe that

$$\Bigl(t - \Bigl(c - \frac{t_i}{\lambda}\Bigr)\Bigr)^k = \sum_{j=1}^{k} c_{i,k,j}\Bigl(t^j - \Bigl(c - \frac{t_i}{\lambda}\Bigr)^j\Bigr)$$

for all $k \in \{1,\ldots,n\}$, $i \in \{1,\ldots,\ell\}$, and $t \in \mathbb{R}$.

Defining for each $i \in \{1,\ldots,\ell\}$ a lower-triangular matrix by

$$T_i = \begin{pmatrix} \lambda c_{i,1,1} & 0 & 0 & \cdots & 0 \\ \lambda^2 c_{i,2,1} & \lambda^2 c_{i,2,2} & 0 & \cdots & 0 \\ \lambda^3 c_{i,3,1} & \lambda^3 c_{i,3,2} & \lambda^3 c_{i,3,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda^n c_{i,n,1} & \lambda^n c_{i,n,2} & \lambda^n c_{i,n,3} & \cdots & \lambda^n c_{i,n,n} \end{pmatrix},$$

we see, by the choice of λ and the fact that $t_i \in [c,d]$, that

$$\begin{aligned} \|T_i\| &\le \sqrt{n} \max_{k\in\{1,\ldots,n\}} \sum_{j=1}^{k} |\lambda^k c_{i,k,j}| = \sqrt{n} \max_{k\in\{1,\ldots,n\}} \sum_{j=1}^{k} \lambda^k \binom{k}{j} \Bigl|\frac{t_i}{\lambda} - c\Bigr|^{k-j} \\ &\le \sqrt{n} \max_{k\in\{1,\ldots,n\}} \sum_{j=1}^{k} \lambda^j \binom{k}{j} (|t_i| + |c| + 1)^{k-j} \le \lambda\sqrt{n} \max_{k\in\{1,\ldots,n\}} (|t_i| + |c| + 1)^k 2^k < 1. \end{aligned}$$

Therefore, the affine map $f_i \colon \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$f_i(x_1,\ldots,x_n) = T_i(x_1,\ldots,x_n) - T_i\Bigl(c - \frac{t_i}{\lambda}, \Bigl(c - \frac{t_i}{\lambda}\Bigr)^2, \ldots, \Bigl(c - \frac{t_i}{\lambda}\Bigr)^n\Bigr)$$

is contractive and satisfies

$$\begin{aligned} f_i(t,t^2,\ldots,t^n) &= T_i\Bigl(t - \Bigl(c - \frac{t_i}{\lambda}\Bigr), t^2 - \Bigl(c - \frac{t_i}{\lambda}\Bigr)^2, \ldots, t^n - \Bigl(c - \frac{t_i}{\lambda}\Bigr)^n\Bigr) \\ &= \Bigl(\lambda\Bigl(t - \Bigl(c - \frac{t_i}{\lambda}\Bigr)\Bigr), \lambda^2\Bigl(t - \Bigl(c - \frac{t_i}{\lambda}\Bigr)\Bigr)^2, \ldots, \lambda^n\Bigl(t - \Bigl(c - \frac{t_i}{\lambda}\Bigr)\Bigr)^n\Bigr) \\ &= (\lambda(t-c)+t_i, (\lambda(t-c)+t_i)^2, \ldots, (\lambda(t-c)+t_i)^n) \end{aligned}$$

for all $t \in [c,d]$. Hence the self-affine set of $\{f_i\}_{i=1}^{\ell}$ is the curve $\mathrm{Img}(\eta)$. □

Let us next focus on the opposite claim.

Theorem 3.2. *If an analytic curve which cannot be embedded in a hyperplane contains a non-trivial self-affine set, then it is an affine image of* $\eta \colon [c,d] \to \mathbb{R}^n$, $\eta(t) = (t,t^2,\ldots,t^n)$, *for some* $c < d$.

Proof. Let $\gamma \colon [a,b] \to \mathbb{R}^n$ be an analytic curve such that $\mathrm{Img}(\gamma)$ is not contained in a hyperplane. Suppose that E is a non-trivial self-affine set of an affine IFS $\{f_i\}_{i=1}^{\ell}$ such that $E \subset \mathrm{Img}(\gamma)$. Let $\mathcal{S}$ be the semigroup generated by $f_1,\ldots,f_\ell$ under composition.

By analyticity and the assumption that $\mathrm{Img}(\gamma)$ is not contained in a hyperplane, without loss of generality, we may assume that $E \subset \gamma((a,b))$ and $\gamma'(t) \neq 0$ for all $t \in (a,b)$. Since (a,b) has a countable cover of open intervals I_i such that $\gamma(I_i)$ has no intersection points, we have $E \subset \bigcup_i E \cap \gamma(I_i)$ and therefore, by the Baire Category Theorem, there exists i and an open set U such that $\emptyset \neq E \cap U \subset E \cap \gamma(I_i)$. Since $E \cap U$ contains a non-trivial self-affine set, we see that no generality is lost if we assume the curve γ to be simple.

Fix $\varphi \in \mathcal{S}$ and write

$$\varphi(x) = M(x - x_0) + x_0 \tag{3.1}$$

for all $x \in \mathbb{R}^n$, where $x_0 \in \mathbb{R}^n$ is the fixed point of φ and M is an $n \times n$ invertible matrix. Note that $x_0 \in E$. Since $E \subset \gamma((a,b))$ there exists $t_0 \in (a,b)$ such that $x_0 = \gamma(t_0)$. Hence we may rewrite (3.1) as

$$\varphi(x) = M(x - \gamma(t_0)) + \gamma(t_0). \tag{3.2}$$

Since E is non-trivial, there exists a sequence $(t_i)_{i\in\mathbb{N}}$ of distinct numbers in (a,b) such that $t_i \to t_0$ as $i \to \infty$ and $\gamma(t_i) \in E$ for all $i \in \mathbb{N}$. Furthermore, since $\varphi(E) \subset E \subset \gamma((a,b))$, we see that $\varphi(\gamma(t_i)) \in \mathrm{Img}(\gamma)$ and therefore, for each $i \in \mathbb{N}$ there exists $t_i' \in (a,b)$ such that

$$\varphi(\gamma(t_i)) = \gamma(t_i'). \tag{3.3}$$

Recalling that γ is simple and $\varphi(\gamma(t_0)) = \gamma(t_0)$, we see that $t_i' \to t_0$ as $i \to \infty$. By (3.1) and (3.3), we have

$$M(\gamma(t_i) - \gamma(t_0)) = \varphi(\gamma(t_i)) - \gamma(t_0) = \gamma(t_i') - \gamma(t_0) \tag{3.4}$$

and therefore,

$$M\left(\frac{\gamma(t_i) - \gamma(t_0)}{t_i - t_0}\right) = \frac{\gamma(t_i') - \gamma(t_0)}{t_i' - t_0} \cdot \frac{t_i' - t_0}{t_i - t_0}.$$

Letting $i \to \infty$, we have

$$M\gamma'(t_0) = \lambda\gamma'(t_0), \tag{3.5}$$

where $\lambda = \lim_{i\to\infty}(t_i' - t_0)/(t_i - t_0) \neq 0$ by the invertibility of M.

Let J be an invertible matrix such that

$$J^{-1}\gamma'(t_0) = (1, 0, \ldots, 0)$$

and

$$J^{-1}MJ = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m \end{pmatrix}$$

is a real canonical Jordan form of M. Write $A = J^{-1}MJ$ and recall that if λ_i is a real eigenvalue of M, then

$$A_i = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix},$$

and if λ_i is a non-real eigenvalue of M with real part a_i and imaginary part b_i, then

$$A_i = \begin{pmatrix} C_i & I & 0 & \cdots & 0 & 0 \\ 0 & C_i & I & \cdots & 0 & 0 \\ 0 & 0 & C_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & C_i & I \\ 0 & 0 & 0 & \cdots & 0 & C_i \end{pmatrix},$$

where

$$C_i = \begin{pmatrix} a_i & b_i \\ -b_i & a_i \end{pmatrix} \quad \text{and} \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Note that by (3.5), we have $\lambda_1 = \lambda \in \mathbb{R}$. Observe also that, by (3.4), it holds that

$$AJ^{-1}(\gamma(t_i) - \gamma(t_0)) = J^{-1}(\gamma(t_i') - \gamma(t_0)) \tag{3.6}$$

for all $i \in \mathbb{N}$.

Defining $\tilde{\gamma} \colon [a,b] \to \mathbb{R}^n$ by

$$\tilde{\gamma}(t) = J^{-1}(\gamma(t) - \gamma(t_0)),$$

we clearly have $\tilde{\gamma}(t_0) = 0$ and $\tilde{\gamma}'(t_0) = (1, 0, \ldots, 0)$. Write $\tilde{\gamma}(t) = (\tilde{x}_1(t), \ldots, \tilde{x}_n(t))$. Since $\tilde{x}_1'(t_0) = 1 \neq 0$, the inverse $\tilde{x}_1^{-1}$ exists and is analytic on $(-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. To simplify notation, let us denote $\tilde{x}_1^{-1}$ by t and its parameters by $\tilde{x}_1$. Therefore, $\tilde{x}_k$ can be considered to be an analytic function of $\tilde{x}_1$ on $(-\varepsilon, \varepsilon)$ for all $k \in \{2, \ldots, n\}$. Note that

$$\tilde{x}_k(0) = 0 = \tilde{x}_k'(0)$$

for all $k \in \{2, \ldots, n\}$ and $\tilde{x}_2, \ldots, \tilde{x}_n$ are not constant functions. Indeed, if $\tilde{x}_k$ was a constant for some k, then, by the fact that each $\tilde{x}_k$ is a linear combination of $x_1, \ldots, x_n$, the curve γ would be contained in a hyperplane in $\mathbb{R}^n$. Let $\eta \colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ be defined by

$$\eta(\tilde{x}_1) = (\tilde{x}_1, \tilde{x}_2(\tilde{x}_1), \ldots, \tilde{x}_n(\tilde{x}_1)). \tag{3.7}$$

The goal of the proof is to show that the curve η is of the claimed form.

Let us next collect three facts related to the above defined setting.

Fact 1. Write $A = (a_{ij})_{1\le i,j\le n}$ and let $Y = \sum_{j=1}^{n} a_{1j}\tilde{x}_j$. Then

$$A(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) = (Y, \tilde{x}_2(Y), \dots, \tilde{x}_n(Y)) \tag{3.8}$$

for all $\tilde{x}_1 \in (-\varepsilon, \varepsilon)$.

Proof. By (3.6), the equality (3.8) holds for infinitely many different values of $\tilde{x}_1$. By analyticity, (3.8) holds on the whole interval $(-\varepsilon, \varepsilon)$. ∎

The next fact concerns the shape of the matrix A.

Fact 2. The matrix A is diagonal. In other words, all the block matrices A_i have dimension 1.

Proof. Let us first show that A_1 has dimension 1. Suppose to the contrary that $d_1 = \dim(A_1) > 1$. Since the eigenvalue associated to A_1 is $\lambda \in \mathbb{R}$, we have

$$A_1 = \begin{pmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}.$$

By Fact 1, we see that

$$\lambda\tilde{x}_{d_1}(\tilde{x}_1) = \tilde{x}_{d_1}(\lambda\tilde{x}_1 + \tilde{x}_2). \tag{3.9}$$

Notice that there exist integers $p_2, \dots, p_n \ge 2$ and reals $c_2, \dots, c_n \ne 0$ such that for each $k \in \{2, \dots, n\}$

$$\tilde{x}_k(\tilde{x}_1) = c_k(\tilde{x}_1)^{p_k} + o(\tilde{x}_1^{p_1}) \tag{3.10}$$

as $\tilde{x}_1 \to 0$. Plugging (3.10) into (3.9), and comparing the coefficients of Taylor series in $\tilde{x}_1$ on both sides, we get

$$\lambda c_{d_1} = c_{d_1}\lambda^{p_{d_1}}$$

which implies that $p_{d_1} = 1$, a contradiction. Hence we have $\dim(A_1) = 1$ and therefore $Y = \lambda\tilde{x}_1$.

Let us next assume inductively that for some $k \in \{1, \dots, n-1\}$ the matrices $A_1, \dots, A_k$ are of dimension 1 and show that $\dim(A_{k+1}) = 1$. Suppose to the contrary that $d = \dim(A_{k+1}) > 1$. Now there are two cases: either λ_{k+1} is real or not. If λ_{k+1} is real, then the same argument as that for A_1 gives a contradiction. We may thus assume that $\lambda_{k+1} = a + ib$ with $b \neq 0$. The matrix A_{k+1} is therefore of the form

$$A_i = \begin{pmatrix} a & b & 1 & 0 & \cdots & 0 & 0 \\ -b & a & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & a & b & \cdots & 0 & 0 \\ 0 & 0 & -b & a & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & a & b \\ 0 & 0 & 0 & 0 & \cdots & -b & a \end{pmatrix}.$$

Let $\ell = k + d$. Applying (3.8), we see that

$$\begin{aligned} a\tilde{x}_{\ell-1} + b\tilde{x}_\ell &= \tilde{x}_{\ell-1}(\lambda\tilde{x}_1), \\ -b\tilde{x}_{\ell-1} + a\tilde{x}_\ell &= \tilde{x}_\ell(\lambda\tilde{x}_1). \end{aligned}$$

Using the above identities and comparing the coefficients of $\tilde{x}_1^{p_\ell}$ and $\tilde{x}_1^{p_{\ell-1}}$ in the Taylor expansions of $\tilde{x}_\ell$ and $\tilde{x}_{\ell-1}$, we see that $p_\ell = p_{\ell-1}$; and moreover,

$$\begin{aligned} ac_{\ell-1} + bc_\ell &= c_{\ell-1}\lambda^{p_\ell}, \\ -bc_{\ell-1} + ac_\ell &= c_\ell\lambda^{p_\ell}, \end{aligned}$$

or, equivalently,

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} c_{\ell-1} \\ c_\ell \end{pmatrix} = \lambda^{p_\ell} \begin{pmatrix} c_{\ell-1} \\ c_\ell \end{pmatrix}.$$

This means that the real number λ^{p_ℓ} is an eigenvalue of the above matrix, a contradiction. ∎

By Fact 2, we may now write

$$A = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n), \tag{3.11}$$

where $\lambda_1 = \lambda \in (-1, 1) \setminus \{0\}$. With this observation, we can examine how the curve η defined in (3.7) looks like.

Fact 3. There exist integers $p_2 < p_3 < \cdots < p_n$ such that a piece of the curve $\operatorname{Img}(\gamma)$, namely $\gamma\colon (t_0 - \delta, t_0 + \delta) \to \mathbb{R}^n$ for some $\delta > 0$, is an affine image of the curve $\eta\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ defined by

$$\eta(t) = (t, t^{p_2}, \ldots, t^{p_n}).$$

Proof. By (3.11) and (3.8), we have

$$\tilde{x}_k(\lambda\tilde{x}_1) = \lambda_k\tilde{x}_k(\tilde{x}_1) \tag{3.12}$$

and hence, by (3.10), there exist integers $p_2, \dots, p_n \geq 2$ and reals $c_2, \dots, c_n \neq 0$ such that

$$c_k(\lambda \tilde{x}_1)^{p_k} = \lambda_k c_k \tilde{x}_1^{p_k} + o(\tilde{x}_1^{p_k}).$$

This implies that $\lambda_k = \lambda^{p_k}$ and $\tilde{x}_k(\lambda \tilde{x}_1) = \lambda^{p_k} \tilde{x}_k(\tilde{x}_1)$. Taking p_k-th derivative on both sides gives $\tilde{x}_k^{(p_k)}(\lambda \tilde{x}_1) = \tilde{x}_k^{(p_k)}(\tilde{x}_1)$. Hence $\tilde{x}_k^{(p_k)}(\lambda^j \tilde{x}_1) = \tilde{x}_k^{(p_k)}(\tilde{x}_1)$ for all $j \in \mathbb{N}$. Letting $j \to \infty$, we get $\tilde{x}_k^{(p_k)}(\tilde{x}_1) \equiv \tilde{x}_k^{(p_k)}(0) = c_k p_k!$ and therefore,

$$\tilde{x}_k(\tilde{x}_1) = c_k \tilde{x}_1^{p_k}.$$

Since the curve $\tilde{\gamma}$ is not contained in a hyperplane, we see that, for any non-zero vector $(b_1, \dots, b_n)$, the sum $\sum_{k=1}^n b_k \tilde{x}_k$ is not identically zero. Thus the integers $p_2, \dots, p_n$ are mutually distinct.

We have now proved that, possibly after a permutation on coordinate axis, the curve $\gamma\colon (t_0 - \delta, t_0 + \delta) \to \mathbb{R}^n$ for some $\delta > 0$, is an affine image under the affine transformation $u \mapsto J^{-1}(u - \gamma(t_0))$ of the curve

$$t \mapsto (t, c_2 t^{p_2}, \dots, c_n t^{p_n})$$

defined on $(-\varepsilon, \varepsilon)$ for some integers $2 \leq p_2 < p_3 < \cdots p_n$ and reals $c_2, \dots, c_n \neq 0$. Applying a further affine transformation $(u_1, u_2, \dots, u_n) \mapsto (u_1, u_2/c_2, \dots, u_n/c_n)$ we have finished the proof of Fact 3. ■

By Fact 3, it suffices to show that $p_k = k$ for all $k \in \{2, \dots, n\}$. Observe that $\eta\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ given by Fact 3 is an analytic simple curve which cannot be embedded in a hyperplane and it contains a non-trivial self-affine set. Therefore, applying the previous argument once more, we find integers $2 \leq q_2 < q_3 < \cdots < q_n$ and $t_1 \in (-\varepsilon, \varepsilon) \setminus \{0\}$ such that, under a suitable linear transformation J', the curve

$$t \mapsto J'(\eta(t) - \eta(t_1))$$

defined on $(t_1 - \xi, t_1 + \xi) \subset (-\varepsilon, \varepsilon)$ for some $\xi > 0$ can be parametrized by

$$t \mapsto (t, t^{q_2}, \dots, t^{q_n}).$$

This means that, writing $J' = (b_{kj})_{1 \leq k,j \leq n}$, we have

$$\sum_{j=1}^n b_{kj}(t^{p_j} - t_1^{p_j}) = \left(\sum_{j=1}^n b_{1j}(t^{p_j} - t_1^{p_j}) \right)^{q_k} \tag{3.13}$$

for all $t \in (t_1 - \xi, t_1 + \xi)$ and $k \in \{2, \dots, n\}$. By analyticity, (3.13) holds for all $t \in \mathbb{R}$.

We will next compare the degrees of polynomials on both sides of (3.13) for all $k \in \{2, \dots, n\}$. Let $d = \deg(\sum_{j=1}^n b_{1j}(t^{p_j} - t_1^{p_j})) \in \{1, p_2, \dots, p_n\}$. When k runs over $\{2, \dots, n\}$, the degrees of the right-hand side of (3.13) are $dq_2, dq_3, \dots, dq_n$, whereas the left-hand side has degree in $\{1, p_2, \dots, p_n\}$. Therefore,

$$\{dq_2, dq_3, \dots, dq_n\} \subset \{1, p_2, \dots, p_n\}$$

which implies that

$$p_k = dq_k \tag{3.14}$$

for all $k \in \{2, \dots, n\}$. Since $d \in \{1, p_2, \dots, p_n\}$, we must have $d = 1$ – otherwise, by (3.14), $q_k = 1$ for some $k \in \{2, \dots, n\}$ which is a contradiction. But since $d = 1$, we may write (3.13) as

$$\sum_{j=1}^{n} b_{kj}(t^{p_j} - t_1^{p_j}) = (c(t - t_1))^{p_k}$$

for all $k \in \{2, \dots, n\}$. In particular, this shows that $(t - t_1)^{p_n}$ is a linear combination of $(t - t_1), (t^{p_2} - t_1^{p_2}), \dots, (t^{p_n} - t_1^{p_n})$. Since $t_1 \neq 0$, all powers t^j, $j \in \{1, \dots, p_n\}$, appear in $(t - t_1)^{p_n}$ with non-degenerate coefficients, and it follows that $p_k = k$ for all $k \in \{2, \dots, n\}$. □

Remark 3.3. (1) Bandt and Kravchenko showed that there are plenty of C^1 planar self-affine curves (i.e. self-affine sets that are C^1 planar curves); see [1, Theorem 2]. Furthermore, in [1, Theorem 3(ii)], they showed that parabolic arcs and straight line segments are the only simple C^2 planar self-affine curves. This result also follows from Theorem A by a simple modification. It would be interesting to know that if a self-affine set E is contained in a C^2 planar curve, then does there exists an analytic curve containing E?

(2) The analyticity assumption in Theorem A is well motivated since for each $k \in \mathbb{N}$ it is easy to construct a non-quadratic C^k planar curve containing a self-affine set. It would also be interesting to know if there exists a self-affine set E which is a subset of a strictly convex C^2 planar curve, but is not a subset of any quadratic curve. Also, when can a self-affine set intersect an analytic curve in a set of positive measure for some relevant measure such as the self-affine measure? In the self-conformal case, this property implies that the whole set is contained in an analytic curve; see [4, Theorem 2.1].

4. Self-affine sets and algebraic surfaces

In this section, we prove Theorem B and introduce self-affine polynomials.

Proof of Theorem B. Let $P \colon \mathbb{R}^d \to \mathbb{R}$ be a non-constant polynomial with real coefficients such that $S(P)$ is compact. Suppose to the contrary that there exists a non-trivial self-affine set E contained in $S(P)$. Let f be one of the mappings of the affine IFS defining E and set $P_n = P \circ f^{-n}$ for all $n \in \mathbb{N}$. Observe that the degree of P_n is at most $\deg(P)$. It is easy to see that $S(P_n) = f^n(S(P))$ for all $n \in \mathbb{N}$ and therefore $\operatorname{diam}(S(P_n)) \to 0$ as $n \to \infty$. By the assumption, we have $f^n(E) \subset f^n(S(P)) = S(P_n)$ for all $n \in \mathbb{N}$, and by the invariance, we have $f^n(E) \subset f^{n-1}(E) \subset \cdots \subset E$ for all $n \in \mathbb{N}$.

Since the ring of polynomials having degree at most $\deg(P)$ is finite dimensional there exist $P_{k_1}, \ldots, P_{k_m}$ such that each P_n is a linear combination of these polynomials. Choose n so large that

$$\operatorname{diam}(S(P_n)) < \min_{i \in \{1,\ldots,m\}} \operatorname{diam}(f^{k_i}(E)) = \operatorname{diam}\left(\bigcap_{i=1}^m f^{k_i}(E)\right).$$

But since $P_n = \sum_{i=1}^m c_i P_{k_i}$ for some c_i, we have

$$\bigcap_{i=1}^m f^{k_i}(E) \subset \bigcap_{i=1}^m S(P_{k_i}) \subset S(P_n).$$

This contradiction finishes the proof. □

Example 4.1. It is clear that a hyperplane can contain a non-trivial self-affine set. In this example, we show that also other kinds of non-compact algebraic surfaces can contain non-trivial self-affine sets. Let $P \colon \mathbb{R}^d \to \mathbb{R}$, $P(x_1, \ldots, x_d) = x_1^2 + \cdots + x_{d-1}^2 - x_d$ and fix an interval $[a,b] \subset \mathbb{R}$. Define a mapping $\eta \colon [a,b]^{d-1} \to \mathbb{R}^d$ by setting $\eta(x_1, \ldots, x_{d-1}) = (x_1, \ldots, x_{d-1}, x_1^2 + \cdots + x_{d-1}^2)$. Let $\{c_i(x_1, \ldots, x_{d-1}) + (d_i, \ldots, d_i)\}_{i=1}^{\ell}$ be an affine IFS on $\mathbb{R}^{d-1}$ so that $[a,b]^{d-1}$ is the self-affine set generated by it. Define $f_i \colon \mathbb{R}^d \to \mathbb{R}^d$ by setting

$$f_i(x_1, \ldots, x_d) = \begin{pmatrix} c_i & 0 & \cdots & 0 & 0 \\ 0 & c_i & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c_i & 0 \\ 2c_i d_i & 2c_i d_i & \cdots & 2c_i d_i & c_i^2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{d-1} \\ x_d \end{pmatrix} + \begin{pmatrix} d_i \\ d_i \\ \vdots \\ d_i \\ (d-1)d_i^2 \end{pmatrix}$$

for all $(x_1, \ldots, x_d) \in \mathbb{R}^d$ and $i \in \{1, \ldots, \ell\}$. Since $f_i(\eta(x_1, \ldots, x_{d-1})) = \eta(c_i x_1 + d_i, \ldots, c_i x_{d-1} + d_i)$ the image $\eta([a,b]^{d-1}) \subset S(P)$ is invariant under the affine IFS $\{f_i\}_{i=1}^{\ell}$.

The previous example does not characterize the polynomials for which the associated algebraic surface contains non-trivial self-affine sets. Suppose that $P \colon \mathbb{R}^d \to \mathbb{R}$ is a non-constant polynomial with real coefficients. We say that a contractive invertible affine map f is a *scaling factor* for P if there exists a constant $C \in \mathbb{R}$ such that

$$P \circ f = CP. \tag{4.1}$$

A polynomial P is called *self-affine* if it has two scaling factors with distinct fixed points.

Example 4.2. Let $P \colon \mathbb{R}^2 \to \mathbb{R}$, $P(x_1, x_2) = x_2 - x_1$. It is easy to see that $f \colon \mathbb{R}^2 \to \mathbb{R}^2$, $f(x_1, x_2) = \frac{1}{2}(x_1, x_2)$, and $g \colon \mathbb{R}^2 \to \mathbb{R}^2$, $g(x_1, x_2) = \frac{1}{2}(x_1 + 1, x_2 + 1)$, are scaling factors for P and have distinct fixed points.

The following proposition shows that a polynomial P being self-affine is sufficient for the inclusion of self-affine sets.

Proposition 4.3. *If $P\colon \mathbb{R}^d \to \mathbb{R}$ is a self-affine polynomial, then $S(P)$ contains a non-trivial self-affine set.*

Proof. Let f be a scaling factor for P with a constant C. Note that there exists a non-singular $d \times d$ matrix M with $\|M\| < 1$ and $a \in \mathbb{R}^d$ so that $f(x) = Mx + a$ for all $x \in \mathbb{R}^d$. Observe that

$$f^n(x) = M^n x + \sum_{i=0}^{n-1} M^i a \to \sum_{i=0}^{\infty} M^i a =: x_0$$

as $n \to \infty$, where $x_0 \in \mathbb{R}^d$ is the fixed point of f. Choose $x \in \mathbb{R}^d$ such that

$$|P(x_0)| + 1 < |P(x)|.$$

Such a point x exists since P is not bounded. Since

$$C^n P(x) = P \circ f^n(x) \to P(x_0)$$

as $n \to \infty$ we may choose n large enough so that $|C^n P(x)| < |P(x_0)| + 1$. Thus $|C| < 1$.

Let h and g be scaling factors for P with distinct fixed points. If f is any finite composition of the mappings h and g, then f is a scaling factor for P. If C is the constant associated to the scaling factor f, then the above reasoning implies that $|C| < 1$. Furthermore, if x_0 is the fixed point of f, then $P(x_0) = P \circ f(x_0) = CP(x_0)$. Since $|C| < 1$, this implies $P(x_0) = 0$ and $x_0 \in S(P)$. Recalling that $S(P)$ is closed it thus contains the self-affine set generated by the affine IFS $\{h, g\}$. $\square$

Remark 4.4. It would be interesting to characterize all the algebraic surfaces associated to self-affine polynomials. For example, in the two-dimensional case, is the surface always contained in a line through the origin? Of course, the ultimate open question here is to characterize all the algebraic surfaces containing self-affine sets.

Acknowledgement. Feng was partially supported by the HKRGC GRF grants (projects CUHK401013, CUHK14302415).

References

[1] C. Bandt and A. Kravchenko. Differentiability of fractal curves. *Nonlinearity*, 24(10):2717–2728, 2011.

[2] K. Falconer. *Fractal geometry*. John Wiley & Sons Ltd., Chichester, 1990. Mathematical foundations and applications.

[3] J. E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30(5):713–747, 1981.

[4] A. Käenmäki. On the geometric structure of the limit set of conformal iterated function systems. *Publ. Mat.*, 47(1):133–141, 2003.

[5] A. Käenmäki. Geometric rigidity of a class of fractal sets. *Math. Nachr.*, 279(1):179–187, 2006.

[6] P. Mattila. On the structure of self-similar fractals. *Ann. Acad. Sci. Fenn. Ser. A I Math.*, 7(2):189–195, 1982.

[7] V. Mayer and M. Urbański. Finer geometric rigidity of limit sets of conformal IFS. *Proc. Amer. Math. Soc.*, 131(12):3695–3702 (electronic), 2003.

Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong

E-mail address: djfeng@math.cuhk.edu.hk

Department of Mathematics and Statistics, P.O. Box 35 (MaD), FI-40014 University of Jyväskylä, Finland

E-mail address: antti.kaenmaki@jyu.fi