*Japan J. Appl. Math.*, 2 (1985), 381–414

# On the Structure of Self-Similar Sets

Masayoshi HATA

*Department of Mathematics, Faculty of Science,*
*Kyoto University, Kyoto 606, Japan*

Received October 3, 1984
Revised May 14, 1985

We shall investigate topological properties of a uniquely determined compact set $K$ such that $K=\sum_{\lambda\in\Lambda}f_\lambda(K)$, where each $f_\lambda$ is a weak contraction of a complete metric space and $\Lambda=\{1, 2, \cdots, m\}$ or $\Lambda=N$. Such a set $K$ is said to be self-similar. Many classical peculiar sets can be represented in this form. We shall also discuss the interesting problem presented by R. F. Williams.

*Key words*: contractions, locally connected continua, functional equations, Hausdorff dimension

CONTENTS

1. Introduction .......... 381
2. Preliminaries .......... 382
3. Existence and Uniqueness .......... 384
4. Connectedness .......... 388
5. Cut Points, End Points, and Simple Links .......... 392
6. Parameterization .......... 396
7. Regularity of $G$ .......... 400
8. Reversed Equations .......... 403
9. Space-Filling Curves .......... 406
10. Hausdorff Dimension of Self-Similar Sets .......... 407
11. Williams' Problem .......... 410

## 1. Introduction

The notion of "fractals" was introduced by Mandelbrot [30] in the description of Nature. A set $S$ is said to be a *fractal* provided that the Hausdorff dimension of $S$ strictly exceeds the topological dimension of $S$. For example, Cantor's ternary set is a typical example of fractals. Of course, it is a classical problem to investigate such fractal sets in Mathematics. Indeed, measure theory is a fundamental and powerful tool to analyse fractals. See Rogers [45], Falconer [13] and the references given there.

On the other hand, as is pointed out by Mandelbrot, "self-similarity" is very important in the study of such sets. Actually, most classical fractal sets constructed by many mathematicians have the self-similarity in some sense.

The aim of this paper is to investigate various topological structures of self-similar sets, whose definition will be given later, and to analyse many classical pathological sets and curves through the notion of self-similarity. With this formulation, one can easily create and handle self-similar fractals.

Let $X$ be a separable complete metric space with a metric $d$. A mapping $f: X \to X$ is said to be a *contraction* provided that the Lipschitz constant

$$\mathrm{Lip}(f) = \sup_{x \neq y} \frac{d(f(x), f(y))}{d(x, y)} \tag{1.1}$$

satisfies $\mathrm{Lip}(f) < 1$. Every contraction $f$ has a unique fixed point $\mathrm{Fix}(f)$ in $X$. Recently Hutchinson [22] considered the non-empty subset $K \subset X$ satisfying

$$K = f_1(K) \cup f_2(K) \cup \cdots \cup f_m(K) \tag{1.2}$$

where $m \geq 2$ and $\{f_j\}_{1 \leq j \leq m}$ is a given finite family of contractions.

On the other hand, Williams [54] studied the following set

$$K = \mathrm{closure}\left( \bigcup_{\substack{1 \leq i_1, \cdots, i_n \leq m \\ n \geq 1}} \mathrm{Fix}(f_{i_1} \circ f_{i_2} \circ \cdots \circ f_{i_n}) \right) \tag{1.3}$$

toward a study of generic properties of the action of free (non-abelian) groups on manifolds. He proved essentially that there exists a unique compact solution of (1.2); it is therefore given by (1.3). This result was also proved by Hutchinson in a different way. Several properties of $K$ on geometric measure theory were proved in [22]. Mattila [32] strengthened some of them.

In this paper, the equation (1.2) will be generalized to weak contractions and the solution $K$ will be regarded as a fixed point of some set-dynamical system.

For another method to describe self-similar fractals using endomorphisms of words in free groups, see Dekking [9].

The author would like to thank Prof. M. Yamaguti for his useful comments.

## 2. Preliminaries

We begin with some definitions. Let $X$ be the same space as in the previous section.

DEFINITION 2.1. A mapping $f: X \to X$ is said to be a *weak contraction* provided that $\Omega_f(t) \equiv \lim_{s \to t+} \omega_f(s) < t$ for any $t > 0$, where $\omega_f$ is the modulus of continuity of $f$:

$$\omega_f(s) = \sup_{d(x, y) \leq s} d(f(x), f(y)). \tag{2.1}$$

Obviously $\Omega_f(t)$ is non-decreasing and right-continuous. Note that every weak contraction $f$ is uniformly continuous in $X$ and has a unique fixed point $\mathrm{Fix}(f)$ in $X$. The regularity of $\omega_f$ may depend on the space $X$. Indeed, if $X$ is compact, $\omega_f$ is right-continuous; that is, $\Omega_f = \omega_f$. If $X$ is a closed convex subset of a Banach space, then $\omega_f$ is concave, therefore $\Omega_f = \omega_f$ is continuous. For example, let $X$ be the unit interval with the Euclidean distance. Then the function $f(x) = x/(1+x)$ is a weak contraction with $\omega_f = f$, while $f$ is not a contraction since $\mathrm{Lip}(f) = 1$.

The power set $2^X$ of all subsets of $X$ forms a poset under set-inclusion in a natural way; $x \leq y$ means $x$ is a subset of $y$. Moreover, $2^X$ is a complete lattice with operations "+" (join, set-union) and "·" (meet, set-intersection). See Birkhoff [4] for lattice theory.

Let $\mathscr{C}(X) \subset 2^X$ be the set of all non-empty compact subsets of $X$. $\mathscr{C}(X)$ is not a lattice but a join-semilattice. It is known that $\mathscr{C}(X)$ is a complete metric space equipped with the Hausdorff metric:

$$d_H(x, y) = \max(\inf\{\varepsilon > 0;\ N_\varepsilon(x) \geq y\},\ \inf\{\varepsilon > 0;\ N_\varepsilon(y) \geq x\}), \tag{2.2}$$

where $N_\varepsilon(x)$ is an $\varepsilon$-neighborhood of the set $x$. Michael [33] proved that if $X$ is compact, then $\mathscr{C}(X)$ is also compact. Note that the mapping $i: X \to \mathscr{C}(X)$, which maps a point $p$ into the set consisting of the single point $p$, is an isometry.[†)]

We now give a remark. Let $\{x_n\}_{n \geq 1}$ be a Cauchy sequence in $\mathscr{C}(X)$. Then we will denote by $\lim_{n\to\infty} x_n$ the unique limit of $\{x_n\}$ in $\mathscr{C}(X)$; this means $\lim_{n\to\infty} x_n = \bigcap_{m \geq 1} \text{closure}\,(\bigcup_{n \geq m} x_n)$ in the usual notation. Therefore, an infinite sum $\sum_{n=1}^{\infty} y_n$, if it exists, means the set closure $(\bigcup_{n \geq 1} y_n)$.

For any continuous mapping $f$: $X \to X$, we can define the *induced mapping* $f^*$: $\mathscr{C}(X) \to \mathscr{C}(X)$ by $f^*(x) = f(x)$ in a natural way.

DEFINITION 2.2. A set $K \in \mathscr{C}(X)$ is said to be *self-similar* provided that the set $K$ can be expressed in the form

$$K = \sum_{\lambda \in \Lambda} f_\lambda^*(K), \tag{2.3}$$

where $\{f_\lambda\}_{\lambda \in \Lambda}$ is a set of weak contractions of $X$ and the index set $\Lambda$ is $\{1, 2, \cdots, m\}$, $m \geq 2$, or $N$.

(2.3) means that the set $K$ consists of a finite or an infinite number of miniatures of $K$ itself. Thus our terminology will be justified in a sense. Note that Hutchinson's definition of self-similarity differs from ours; he required some separation conditions in addition.

A mapping $F$: $\mathscr{C}(X) \to \mathscr{C}(X)$ is said to be *isotone* provided that $x \leq y$ implies $F(x) \leq F(y)$; a *join-endomorphism* provided that $F(x+y) = F(x) + F(y)$ for any $x, y \in \mathscr{C}(X)$. Let $\mathscr{F}(\mathscr{C}(X))$ be the set of all isotone join-endomorphisms (not necessarily continuous) defined on $\mathscr{C}(X)$. Obviously every induced mapping belongs to $\mathscr{F}(\mathscr{C}(X))$ and is further continuous. Again $\mathscr{F}(\mathscr{C}(X))$ becomes a join-semilattice; $F \leq G$ means $F(x) \leq G(x)$ and $F+G$ means $(F+G)(x) = F(x) + G(x)$ for any $x \in \mathscr{C}(X)$. The following properties on the induced mappings were proved by the author [19].

LEMMA 2.3. *If $f$ is a weak contraction of $X$, then $f^*$ is also a weak contraction of $\mathscr{C}(X)$ with $\Omega_{f^*} = \Omega_f$. Moreover, if $\{f_j\}_{1 \leq j \leq m}$ is a finite set of weak contractions of $X$, then $F = \sum_{j=1}^m f_j^*$ is also a weak contraction of $\mathscr{C}(X)$ with $\Omega_F(t) \leq \max_{1 \leq j \leq m} \Omega_{f_j}(t)$.*

†) Throughout this paper we shall make no distinction in notation between the point $p$ and the set consisting of the single point $p$.

## 3. Existence and Uniqueness

In this section, we shall discuss the equation (2.3) and generalize the results of Williams and Hutchinson mentioned in Section 1. In addition, we shall discuss different types of set-equations.

By Lemma 2.3 we get a generalization of Hutchinson's result immediately.

THEOREM 3.1. *Suppose that* $\{f_j\}_{1\le j\le m}$, $m\ge 2$, *is a finite set of weak contractions of* $X$. *Then there exists a unique compact subset* $K=K(f_1,\cdots,f_m)$ *satisfying the equation* (2.3) *with* $\Lambda=\{1,2,\cdots,m\}$. *Moreover, for any compact subset* $Q\in\mathscr{C}(X)$, *we have*

$$\lim_{n\to\infty} F^n(Q)=K(f_1,\cdots,f_m), \tag{3.1}$$

*where* $F=\sum_{j=1}^m f_j^* \in\mathscr{F}(\mathscr{C}(X))$.

To investigate the structure of the set $K(f_1,\cdots,f_m)$, it is convenient to introduce the *one-sided symbol space* $\Sigma=\{1,2,\cdots,m\}^N$ on $m$ symbols. Endowed with the metric

$$d_\Sigma(\alpha,\beta)=\sum_{n\ge 1}2^{-n}\tau(\alpha_n,\beta_n) \qquad \text{for} \quad \alpha=(\alpha_n), \quad \beta=(\beta_n)\in\Sigma, \tag{3.2}$$

where $\tau(i,j)=1$ if $i\ne j$ and $\tau(i,j)=0$ if $i=j$, $\Sigma$ becomes a compact metric space. Then we have

THEOREM 3.2. *Suppose that* $\{f_j\}_{1\le j\le m}$, $m\ge 2$, *is a finite set of weak contractions of* $X$. *Then there exists a continuous onto mapping* $\psi:\Sigma\to K(f_1,\cdots,f_m)$ *such that the following diagram is commutative*:

$$\begin{array}{ccc} \Sigma & \xrightarrow{\sigma_j} & \Sigma \\ \psi\downarrow & & \downarrow\psi \\ K & \xrightarrow[f_j]{} & K \end{array}$$

*where* $\sigma_j$ *is the right-shift operator*: $\sigma_j(\alpha_1\alpha_2\cdots)=(j\alpha_1\alpha_2\cdots)$ *for any* $1\le j\le m$. *In particular, Williams' formula* (1.3) *holds true.*

*Proof.* Let $\Omega(t)=\max_{1\le j\le m}\Omega_{f_j}(t)$ for brevity. First we will show that the sequence defined by $p_n(\alpha)=f_{\alpha_1}\circ f_{\alpha_2}\circ\cdots\circ f_{\alpha_n}(p_0)$, $n\ge 1$, is a Cauchy sequence in $X$. To show this, for any $\varepsilon>0$, define a sufficiently large integer $N=N(\varepsilon)$ such that

$$\Omega^N(M)\le\varepsilon-\Omega(\varepsilon) \qquad \text{where} \quad M=\max_{1\le j\le m} d(p_0,f_j(p_0)).$$

Then $d(p_N(\alpha),p_{N+1}(\alpha))\le\Omega^N(M)\le\varepsilon-\Omega(\varepsilon)<\varepsilon$ for any $\alpha\in\Sigma$. Suppose now that $d(p_N(\alpha),p_{N+j}(\alpha))\le\varepsilon$ for any $1\le j\le k$ and $\alpha\in\Sigma$. Then it follows that

$$\begin{aligned} d(p_N(\alpha),p_{N+k+1}(\alpha)) &\le d(p_N(\alpha),p_{N+1}(\alpha))+d(p_{N+1}(\alpha),p_{N+k+1}(\alpha)) \\ &\le\varepsilon-\Omega(\varepsilon)+\Omega(d(p_N(\alpha'),p_{N+k}(\alpha')))\le\varepsilon, \end{aligned}$$

where $\alpha'=(\alpha_2\alpha_3\cdots)$. Hence $d(p_N(\alpha), p_{N+j}(\alpha))\leq\varepsilon$ for any $j\geq 1$ by induction and therefore $\{p_n(\alpha)\}$ is a Cauchy sequence. It is easily seen that $p_\infty(\alpha)=\lim_{n\to\infty}p_n(\alpha)$ is independent of the choice of $p_0$.

Now define $\psi(\alpha)=p_\infty(\alpha)$ for $\alpha\in\Sigma$. Since $d(p_\infty(\alpha), p_N(\alpha))\leq\varepsilon$, the set $\psi(\Sigma)$ is bounded and therefore $\psi$ is continuous. Thus $\psi(\Sigma)$ is a compact subset satisfying the equation (1.2). Therefore we have $K(f_1,\cdots,f_m)=\psi(\Sigma)$ by Theorem 3.1. □

For a fixed weak contraction $f$ of $X$, let $\mathscr{W}_f(X)$ be the set of all weak contractions $g$ satisfying $\Omega_g(t)\leq\Omega_f(t)$ for any $t>0$. $\mathscr{W}_f(X)$ is endowed with topology of uniform convergence on compact sets. Then we have

THEOREM 3.3. *Suppose that $f$ is a weak contraction of $X$. Then the mapping*

$$K\colon \mathscr{W}_f(X)\times\cdots\times\mathscr{W}_f(X)\longrightarrow\mathscr{C}(X)\,,$$

*which maps $(f_1,\cdots,f_m)$ into the set $K(f_1,\cdots,f_m)$, is continuous.*

*Proof.* Suppose that $g_j^{(n)}\to g_j$ as $n\to\infty$ in $\mathscr{W}_f(X)$ for $1\leq j\leq m$. Put $d^*=\operatorname{diam}(K(g_1,\cdots,g_m))$ for brevity. Let $G_\delta^\varepsilon$ be the closure of $\{(x,y);\ \varepsilon\leq x\leq d^*,\ y=\Omega_f(x)\geq x-\delta\}$ for $\varepsilon,\delta>0$. Then it follows that for any fixed $\varepsilon>0$, $G_\delta^\varepsilon=\phi$ for a sufficiently small $\delta=\delta(\varepsilon)$. Thus there exists $n(\varepsilon)$ such that $H^n(d^*)\leq\varepsilon$ for any $n\geq n(\varepsilon)$ where $H(x)=\Omega_f(x)+\delta$.

On the other hand, there exists $N=N(\varepsilon)\geq n(\varepsilon)$ such that

$$\sup_{x\in Q} d(g_j^{(n)}(x), g_j(x))\leq\delta(\varepsilon)\qquad\text{for any}\quad 1\leq j\leq m\,,\quad n\geq N$$

where $Q=K(g_1,\cdots,g_m)\in\mathscr{C}(X)$. Then

$$\begin{aligned}&d(g_{\alpha_1}^{(n)}\circ g_{\alpha_2}^{(n)}\circ\cdots, g_{\alpha_1}\circ g_{\alpha_2}\circ\cdots)\\ &\leq d(g_{\alpha_1}^{(n)}\circ g_{\alpha_2}^{(n)}\circ\cdots, g_{\alpha_1}^{(n)}\circ g_{\alpha_2}\circ\cdots)+d(g_{\alpha_1}^{(n)}\circ g_{\alpha_2}\circ\cdots, g_{\alpha_1}\circ g_{\alpha_2}\circ\cdots)\\ &\leq\Omega_f(d(g_{\alpha_2}^{(n)}\circ\cdots, g_{\alpha_2}\circ\cdots))+\delta(\varepsilon)=H(d(g_{\alpha_2}^{(n)}\circ\cdots, g_{\alpha_2}\circ\cdots))\,.\end{aligned}$$

Continuing in this way, we arrive at $d(g_{\alpha_1}^{(n)}\circ\cdots, g_{\alpha_1}\circ\cdots)\leq H^n(d^*)\leq\varepsilon$ for $n\geq N$; therefore $d_H(K(g_1^{(n)},\cdots,g_m^{(n)}), K(g_1,\cdots,g_m))\leq\varepsilon$. Since $\varepsilon$ is arbitrary, this completes the proof. □

For the case $\Lambda=N$, we have

THEOREM 3.4. *Suppose that $\{f_n\}_{n\geq1}$ is a family of weak contractions of $X$ satisfying $\lim_{n\to\infty}\Omega_{f_n}(t)=0$ for any $t>0$. Suppose further the set $\bigcup_{n\geq1}\operatorname{Fix}(f_n)$ is precompact. Then there exists a unique compact subset $K=K(f_1,f_2,\cdots)$ satisfying the equation* (2.3) *with $\Lambda=N$. Moreover, for any compact $Q\in\mathscr{C}(X)$, we have*

$$\lim_{n\to\infty}F^n(Q)=K(f_1,f_2,\cdots)\tag{3.3}$$

*where $F=\sum_{n\geq1}f_n^*\in\mathscr{F}(\mathscr{C}(X))$.*

*Proof.* We first show that the operator $F=\sum_{n\geq1}f_n^*$ is well-defined. It suffices to show the set $\bigcup_{n\geq 1} f_n(x)$ is pre-compact for any $x\in\mathscr{C}(X)$. We denote by $\gamma(M)$ Kuratowski's noncompactness measure [26, p. 412] of a bounded subset $M$ of $X$. For any fixed $x\in\mathscr{C}(X)$, put $Q=\sum_{n\geq 1}\operatorname{Fix}(f_n)\in\mathscr{C}(X)$ and $d^*=\sup\{d(p,q);\, p\in x,\, q\in Q\}$ for brevity. Then, for any $\varepsilon>0$, there exists $N=N(\varepsilon)$ such that $d(\operatorname{Fix}(f_n), f_n(p))\leq \Omega_{f_n}(d(\operatorname{Fix}(f_n),p))\leq \Omega_{f_n}(d^*)\leq\varepsilon$ for any $p\in x$, $n\geq N$. This implies $f_n(x)\subset N_\varepsilon(Q)$ and therefore

$$\gamma\Big(\bigcup_{n\geq 1} f_n(x)\Big)\leq \gamma\Big(\bigcup_{n\geq N} f_n(x)\Big)\leq \gamma(N_\varepsilon(Q))\leq \gamma(Q)+2\varepsilon=2\varepsilon\,.$$

Since $\varepsilon$ is arbitrary, it follows that $\bigcup_{n\geq 1} f_n(x)$ is pre-compact.

Now define $\Omega^*(t)=\sup_{n\geq 1}\Omega_{f_n}(t)$ for any $t>0$. Evidently we have $\Omega_F(t)\leq \Omega^*(t)$. Also it is easily verified that $\Omega^*$ is a non-decreasing right-continuous function satisfying $\Omega^*(t)<t$ for $t>0$. This implies that $F$ is a weak contraction of $\mathscr{C}(X)$; this completes the proof. □

Note that the symbol space $\Sigma=N^N$ is complete (not compact) with the metric (3.2). Then we have

THEOREM 3.5. *Suppose that $\{f_n\}_{n\geq 1}$ satisfies the same conditions as in Theorem 3.4. Then there exists a continuous mapping $\psi_\infty: N^N\to K(f_1,f_2,\cdots)$ such that*

$$K(f_1,f_2,\cdots)=\operatorname{closure}(\psi_\infty(N^N))$$
$$=\operatorname{closure}\Bigg(\bigcup_{\substack{i_1,\cdots,i_n\geq 1\\ n\geq 1}}\operatorname{Fix}(f_{i_1}\circ\cdots\circ f_{i_n})\Bigg). \tag{3.4}$$

The proof is similar to that of Theorem 3.2 and easily verified.

We now remark that it is quite interesting to take off the restriction that $\{f_j\}_{1\leq j\leq m}$ is a set of weak contractions in the equation (2.3). As an example, consider a rational function $R(z)$ on the Riemann sphere $\bar{C}$. The *Julia set* $J$ of $R(z)$ is defined by the set of $\bar{C}$ where the family of the iteration $\{R^n(z)\}$ is not normal. It is well-known that $J$ is a closed, perfect and completely invariant set under $R$; that is, $J=R(J)=R^{-1}(J)$ (see e.g. Brolin [5]). On the other hand one can easily show that a set $Y$ is completely invariant under $R$ if and only if the set $Y$ satisfies

$$Y=R(Y)+R^{-1}(Y)\,. \tag{3.5}$$

Then we conclude that the Julia set is the smallest closed solution of (3.5) which contains a repulsive periodic point, since Julia showed that $J$ is the closure of the set of all repulsive periodic points. (This will correspond to Williams' formula (1.3).)

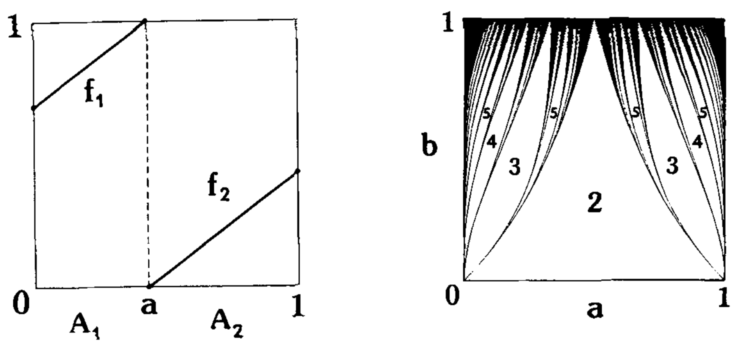
As a second example, consider the action in $\bar{C}$ of a discrete subgroup $G$ of Möbius transformations. For simplicity, we suppose that $G=\langle A, B\rangle$ is not elementary. Then the *limit set* $L$ of $G$ is defined by the closure of the set of points fixed by some elements of $G$. It is well-known that $L$ is a perfect and $G$-invariant set; that is, $L=V(L)$ for all $V$ in $G$ (see e.g. Beardon [1]). In other words, $L$ is the smallest non-empty closed set satisfying

$$L = A(L) + A^{-1}(L) + B(L) + B^{-1}(L)\,. \tag{3.6}$$

Finally we will give an interesting example of a set-equation different from (2.3'). Let $X$ be the unit interval [0, 1] with the usual Euclidean distance. Then we consider the set-equation

$$K = f_1(K \cdot A_1) + f_2(K \cdot A_2)\,, \tag{3.7}$$

where $A_1 = [0, a]$, $A_2 = [a, 1]$ and $f_1(s) = 1 + b(s-a)$, $f_2(s) = b(s-a)$ with two parameters $0 < a < 1$, $0 < b < 1$ (Fig. 1(a)). The equation (3.7) originates in the study of some discontinuous dynamical system done by the author [16]. In fact, the attractor of the dynamical system becomes a compact solution of (3.7) for almost all parameters. Moreover, the uniqueness of such a solution follows from the fact that the attractor is *minimal*. If $(a, b)$ belongs to the domain $D_n$ numbered by $n$ in Fig. 1(b),




(a) (b)

Fig. 1.

the solution $K$ of (3.7) consists of $n$ points. On the other hand, if $(a, b)$ belongs to the remainder set $R = (0, 1)^2 - \sum_{n \geq 2} \bar{D}_n$, $K$ becomes a Cantor set with zero Hausdorff dimension. Note that the mapping $F(x) = f_1^*(x \cdot A_1) + f_2^*(x \cdot A_2)$[†)] for $x \in \mathscr{C}(X)$ belongs to $\mathscr{F}(\mathscr{C}(X))$, while it is discontinuous in $\mathscr{C}(X)$. We will give a generalization of the above fact as follows:

THEOREM 3.6. *Suppose that $X$ consists of $m \geq 2$ closed convex subsets $A_1, \cdots, A_m$ of $R^p$ with the usual Euclidean distance. Let $f_j$: $A_j \to X$ be a weak contraction for $1 \leq j \leq m$. Then the equation*

$$K = \sum_{j=1}^{m} f_j(K \cdot A_j) \tag{3.8}$$

*has the maximal compact solution $K_M$; that is, every compact solution $K$ of* (3.8) *satisfies $K \leq K_M$. If in addition $K_m \cdot A_i \cdot A_j = \phi$ for any $i \neq j$, then $K_M$ is a unique compact solution of* (3.8) *if and only if $K_M$ is minimal; that is,*

[†)] Of course, we adopt the rule $f^*(\phi) = \phi$.

$$K_M=\prod_{i\geq 1}\sum_{j\geq i} F^j(\{q\}) \qquad \textit{for any} \quad q\in K_M\,, \tag{3.9}$$

*where* $F(x)=\sum_{j=1}^m f_j^*(x\cdot A_j)\in\mathscr{F}(\mathscr{C}(X))$.

*Proof.* It is known that there exists a retraction $r_j: R^p\to A_j$ such that $\mathrm{Lip}(r_j)\leq 1$ for $1\leq j\leq m$. Hence the extension $\tilde{f}_j=f_j\circ r_j$ of $f_j$ becomes a weak contraction of $X$. Put $Q=K(\tilde{f}_1,\cdots,\tilde{f}_m)$ for brevity. Then

$$Q=\sum_{j=1}^m \tilde{f}_j^*(Q)\geq\sum_{j=1}^m f_j^*(Q\cdot A_j)=F(Q)$$

and therefore there exists $Q_\infty=\lim_{n\to\infty}Q_n\equiv\lim_{n\to\infty}F^N(Q)\in\mathscr{C}(X)$ since $F$ is isotone. We now show $Q_\infty$ satisfies the equation (3.8). One can easily show that (i) if $Q_\infty\cdot A_j=\phi$, then $Q_N\cdot A_j=\phi$ for some $N$; (ii) if $Q_\infty\cdot A_j\neq\phi$, then $Q_n\cdot A_j\to Q_\infty\cdot A_j$ as $n\to\infty$ in $\mathscr{C}(X)$. Hence $Q_{n+1}=F(Q_n)=\sum f_j^*(Q_n\cdot A_j)\to\sum f_j^*(Q_\infty\cdot A_j)=F(Q_\infty)$ as required.

Put $\tilde{F}=\sum_{j=1}^m\tilde{f}_j^*$. Then for every compact solution $K$ of (3.8), we have

$$\tilde{F}(K)=\sum_{j=1}^m\tilde{f}_j^*(K)\geq\sum_{j=1}^m f_j^*(K\cdot A_j)=K\,,$$

and therefore $Q=\lim_{n\to\infty}\tilde{F}^n(K)\geq K$ by Theorem 3.1. Hence

$$Q_\infty=\lim_{n\to\infty}F^n(Q)\geq\lim_{n\to\infty}F^n(K)=K\,.$$

Thus $K_M=Q_\infty$ is the maximal solution of (3.8).

We now show the second part of the theorem. It suffices to deduce the minimality from the uniqueness of $K_M$. For any fixed $q\in K_M$, put $Q^\infty=\lim_{n\to\infty}Q^n$ where $Q^n=\sum_{j\geq n}F^j(\{q\})$. Then we have

$$F(Q^n)=\sum_{j=1}^m f_j(Q^n\cdot A_j)\longrightarrow\sum_{j=1}^m f_j(Q^\infty\cdot A_j)=F(Q^\infty)\qquad\text{as}\quad n\to\infty\,,$$

since $\{Q^n\}$ is a decreasing sequence in $\mathscr{C}(X)$. Hence $F(Q^n)\geq\sum_{j\geq n+1}F^j(\{q\})=Q^{n+1}$ and therefore $F(Q^\infty)\geq Q^\infty$. Since $Q^\infty\leq K_M$, $\{Q^\infty\cdot A_j\}_{1\leq j\leq m}$ are pairwise disjoint compact subsets by assumption. Therefore $F(Q^\infty)\leq Q^\infty$ and we get $Q^\infty=K_M$ by uniqueness. This completes the proof. □

REMARK. One can easily construct an example for which the set $K_M$ is a unique compact solution of (3.8) and satisfies $K_M\cdot A_i\cdot A_j\neq\phi$ for some $i\neq j$, while $K_M$ is not minimal.

## 4. Connectedness

In this section, we will discuss the connectedness of self-similar sets. Throughout this paper, $\dim_T(Q)$ denotes the (topological) dimension of a set $Q$ in the Menger-

Urysohn sense (see e.g. Hurewicz-Wallman [21, p. 24]); $W(n)$ denotes the set of all finite words with length $n$ on symbols $\{1, 2, \cdots, m\}$. First of all, we have

THEOREM 4.1 (Williams [54]). *Suppose that* $\{f_j\}_{1\le j\le m}$ *is a finite set of contractions of* $X$ *satisfying* $\sum_{j=1}^m \mathrm{Lip}(f_j) < 1$. *Then* $K = K(f_1, \cdots, f_m)$ *is totally disconnected and therefore* $\dim_T(K) = 0$.

It will be interesting to consider a higher dimensional version of this theorem; is it true or not that, if $\sum_{j=1}^m (\mathrm{Lip}(f_j))^p < 1$, then $\dim_T(K) \le p-1$? In connection with this, we have

THEOREM 4.2. *Suppose* $X \subset \boldsymbol{R}^p$ *and* $\{f_j\}_{1\le j\le m}$ *is a finite set of contractions of* $X$ *satisfying* $\sum_{j=1}^m (\mathrm{Lip}(f_j))^p < 1$. *Then Riemann's p-dimensional outer area of the set* $K(f_1, \cdots, f_m)$ *is zero. In particular, it also holds true for the p-dimensional Lebesgue measure.*

*Proof.* Consider a closed ball $B \subset \boldsymbol{R}^p$ containing the set $K = K(f_1, \cdots, f_m)$. The outer area in the sense of Riemann of a bounded set $Q$ will be denoted by $\bar{s}(Q)$. Then

$$\bar{s}(K) \le \sum_{w\in W(n)} \bar{s}(f_w(K)) \le \sum_{w\in W(n)} \bar{s}(f_w(B)) \le \bar{s}(B) \sum_{w\in W(n)} (\mathrm{Lip}(f_w))^p$$
$$\le \bar{s}(B)\left(\sum_{j=1}^m (\mathrm{Lip}(f_j))^p\right)^n \longrightarrow 0 \quad \text{as} \quad n\to\infty ,$$

where $f_w = f_{w_1} \circ \cdots \circ f_{w_n}$ for any $w = (w_1 \cdots w_n) \in W(n)$. Hence $\bar{s}(K) = 0$ as required. □

Using the mapping $\psi : \Sigma \to K(f_1, \cdots, f_m)$ defined in Theorem 3.2, we can get two theorems concerning the topological structures of the set $K$ for weak contractions.

THEOREM 4.3. *Suppose that* $\{f_j\}_{1\le j\le m}$ *is a finite set of one to one weak contractions of* $X$. *Suppose further that* $\mathrm{Fix}(f_i) \ne \mathrm{Fix}(f_j)$ *for some* $i \ne j$. *Then the set* $K = K(f_1, \cdots, f_m)$ *is perfect and therefore* $K$ *is uncountable.*

*Proof.* Suppose, on the contrary, that $\psi(\alpha)$ is an isolated point of $K$ for some $\alpha = (\alpha_n) \in \Sigma$. By the continuity of $\psi$, there exists a $\delta > 0$ such that $\psi(\alpha) = \psi(\beta)$ for any $\beta \in N_\delta(\alpha)$. Put $\bar{\alpha} = (\alpha_1 \cdots \alpha_N\, iii \cdots)$ and $\underline{\alpha} = (\alpha_1 \cdots \alpha_N\, jjj \cdots)$ for a sufficiently large $N$. Then $\psi(\bar{\alpha}) = \psi(\alpha) = \psi(\underline{\alpha})$ implies $\mathrm{Fix}(f_i) = \mathrm{Fix}(f_j)$, contrary to the assumption. □

THEOREM 4.4. *Suppose that* $\{f_j\}_{1\le j\le m}$ *is a finite set of one to one weak contractions of* $X$. *Suppose further that* $\{f_j(K)\}_{1\le j\le m}$ *are pairwise disjoint where* $K = K(f_1, \cdots, f_m)$. *Then the set* $K$ *is totally disconnected and perfect* (*therefore* $\dim_T(K) = 0$ *and* $K$ *is uncountable*).

*Proof.* By assumption, it easily follows that $\psi$ is one to one. Hence $\psi$ becomes a homeomorphism and the conclusion follows from the fact that $\Sigma$ is totally disconnected and perfect. □

EXAMPLE. Let $X = [0, 1]$ with the usual Euclidean distance and

$$f_1(x)=ax \quad \text{and} \quad f_2(x)=b(x-1)+1\ , \tag{4.1}$$

where $0<a<1$ and $0<b<1$ are two parameters. If $a+b<1$, the set $K=K(f_1,f_2)$ is totally disconnected and perfect by Theorems 4.1 and 4.3. This also follows from Theorem 4.4. In particular, if $a=b=1/3$, then $K$ becomes Cantor's ternary set. On the other hand, if $a+b\geq 1$, it is clear that $K=[0, 1]$ and therefore $\dim_T(K)=1$.

REMARK 1. There exist two weak contractions $f_1$, $f_2$ of $X=[0, 1]$ such that $\mathrm{Lip}(f_1)=\mathrm{Lip}(f_2)=1$ and $K(f_1,f_2)$ is totally disconnected and perfect. For example, put

$$f_1(x)=\frac{x}{1+2x} \quad \text{and} \quad f_2(x)=\frac{2-x}{3-2x}\ , \tag{4.2}$$

and apply Theorems 4.3 and 4.4. One can also construct $f_1, f_2$ for which $K(f_1,f_2)$ is totally disconnected, perfect and of positive measure.

REMARK 2. There exists a finite set of contractions $\{f_j\}_{1\leq j\leq m}$, $m\geq 3$, satisfying $\sum_{j=1}^{m}\mathrm{Lip}(f_j)<1$, for which the set $K(f_1, \cdots, f_m)$ is totally disconnected and perfect, and the mapping $\psi: \Sigma\to K$ is *not* a homeomorphism. For example, let $X=[0, 1]$ and

$$f_1(x)=\frac{x}{4}\ , \qquad f_2(x)=\frac{x}{4}+\frac{3}{5} \quad \text{and} \quad f_3(x)=\frac{x}{4}+\frac{3}{4}\ . \tag{4.3}$$

In fact, $K(f_1,f_2,f_3)$ has the required properties by Theorems 4.1 and 4.3, while $\psi$ is not a homeomorphism since $\mathrm{Fix}(f_2)=\mathrm{Fix}(f_3\circ f_1)$. One can easily construct such an example for any $m\geq 3$. This gives a counter-example for Williams' Theorem D. Indeed, $m=2$ is the only correct case and its proof will be given later.

We now state our main theorem in this section. We need some definitions.

DEFINITION 4.5. A set $Q\subset X$ is said to be *locally connected* at $p\in Q$ provided that for any neighborhood $U$ of $p$, there exists a neighborhood $V$ of $p$ such the $Q\cdot V$ lies in a single component of $Q\cdot U$ containing $p$. A set $Q$ which is locally connected at every point of $Q$ is said to be locally connected. A finite sequence of points $\{p_1, \cdots, p_n\}$ is said to be an *ε-chain* joining $p_1$ and $p_n$ provided that $d(p_i, p_{i+1})<\varepsilon$ for any $1\leq i\leq n-1$. A set $Q\subset X$ is said to be *well-chained* provided that for any $\varepsilon>0$, any two points $p, q\in Q$ can be joined by an ε-chain of points all lying in $Q$. A finite sequence of subsets $\{Q_1, \cdots, Q_n\}$ is said to be a *finite chain* joining $Q_1$ and $Q_n$ provided that $Q_i\cdot Q_{i+1}\neq\phi$ for any $1\leq i\leq n-1$.

THEOREM 4.6. *Let $\{f_j\}_{1\leq j\leq m}$ be a finite set of weak contractions of $X$. Then the set $K=K(f_1, \cdots, f_m)$ is a locally connected continuum if and only if for any $1\leq i<j\leq m$, there exists a sequence $\{r_1, \cdots, r_n\}\subset\{1, 2, \cdots, m\}$ such that $\{f_i(K), f_{r_1}(K), \cdots, f_{r_n}(K), f_j(K)\}$ is a finite chain.*

*Proof.* It suffices to show the condition is sufficient. Let $K(w)=f_{w_1}\circ\cdots\circ f_{w_n}(K)\in\mathscr{C}(X)$ for any $w=(w_1\cdots w_n)\in W(n)$. We first prove the following proposition by induction on $k$; for any finite words $u \neq v \in W(k)$, there exists a sequence $\{w^1, \cdots, w^n\} \subset W(k)$ such that $\{K(u), K(w^1), \cdots, K(w^n), K(v)\}$ is a finite chain. By assumption, this holds true for $k=1$. Suppose next that this holds true for $k=l$. Then we must show this is also valid for $k=l+1$. Suppose, on the contrary, that there exist $u \neq v \in W(l+1)$ for which there are no finite chains joining $K(u)$ and $K(v)$. Put $W' = \{w \in W(l+1)$; there exists a finite chain joining $K(u)$ and $K(w)\}$. Then $u \in W'$ and $v \in W'' \equiv W(l+1) - W'$. Thus we have a separation

$$K = \sum_{w \in W'} K(w) + \sum_{w \in W''} K(w) \equiv K' + K'' . \tag{4.4}$$

Therefore there exists a word $w^* \in W(l)$ satisfying $K(w^*) \cdot K' \neq \phi \neq K(w^*) \cdot K''$. Since $K(w^*) = K(w^* \circ 1) + \cdots + K(w^* \circ m)$[†], there exist $i \neq j$ satisfying $K(w^* \circ i) \cdot K' \neq \phi \neq K(w^* \circ j) \cdot K''$. Now let $\{K(i), K(r_1), \cdots, K(r_n), K(j)\}$ be a finite chain joining $K(i)$ and $K(j)$. Then it is clear that $\{K(w^* \circ i), K(w^* \circ r_1), \cdots, K(w^* \circ r_n), K(w^* \circ j)\}$ is a finite chain. This implies $w^* \circ j \in W'$ and therefore $K(w^* \circ j) \in K' \cdot K''$, contrary to (4.4). This completes the proof of our proposition.

Now for any $p, q \in K$, there exist $w^p, w^q \in W(n)$ such that $p \in K(w^p)$ and $q \in K(w^q)$. Then by our proposition, there exists a finite chain $\{K(w^p), K(w^1), \cdots, K(w^n), K(w^q)\}$. Choose a finite sequence of points $\{s_j\}$ satisfying $s_1 \in K(w^p) \cdot K(w^1), \cdots, s_{n+1} \in K(w^n) \cdot K(w^q)$. Since $\operatorname{diam}(K(w)) \leq \Omega^n(\operatorname{diam}(K))$ for any $w \in W(n)$ where $\Omega(t) = \max_{1 \leq j \leq m} \Omega_{f_j}(t)$, the sequence $\{p, s_1, \cdots, s_{n+1}, q\}$ becomes an $\varepsilon$-chain for a sufficiently large $n$. Since $\varepsilon$ is arbitrary, $K$ is well-chained and therefore $K$ is connected (Whyburn [53, p. 15]). Note that $K(w)$ is also connected and for any $\varepsilon > 0$, the set $K$ is the sum of a finite number of connected sets each of diameter less than $\varepsilon$. Hence $K$ is locally connected [53, p. 20]. This completes the proof. □

REMARK 1. The *structure matrix* for the set $K = K(f_1, \cdots, f_m)$ is defined by the $m \times m$ non-negative symmetric matrix $S = (s_{ij})$ where $s_{ij} = 1$ if $f_i(K) \cdot f_j(K) \neq \phi$ and $s_{ij} = 0$ otherwise. Then the above theorem runs as follows: $K(f_1, \cdots, f_m)$ is a locally connected continuum if and only if $S$ is *irreducible*.

REMARK 2. If the set $K = K(f_1, \cdots, f_m)$ is connected, $K$ has the following further properties: (i) $K$ is *semi-locally-connected*; that is, for any $p \in K$ and any $\varepsilon > 0$, there exists a neighborhood $U$ of $p$ of diameter less than $\varepsilon$ such that $K - U \cdot K$ has only a finite number of components [53, p. 20]; (ii) $K$ is *arcwise connected*; that is, any $p, q \in K$ can be joined in $K$ by a simple arc [53, p. 36]; (iii) there exists a continuous onto mapping $H$: $[0, 1] \to K$ (Hahn-Mazurkiewicz theorem [53, p. 33]).

REMARK 3. There exists a set of contractions $\{f_n\}_{n \geq 1}$ of $X = \boldsymbol{R}^2$ for which the set $K(f_1, f_2, \cdots)$ is *not* locally connected. For example, let $Q = Q_0 + \sum_{n \geq 1} K_n$, where $Q_0$ is the square with vertices (0, 0), (1, 0), (1, 1) and (0, 1), and $K_n$ is the straight line interval from $(1/n, 0)$ to $(1/n, 1)$ for $n \geq 1$ (Fig. 2). Then one can easily construct

[†] For any finite words $u = (u_1 \cdots u_r) \in W(r)$, $v = (v_1 \cdots v_s) \in W(s)$, the composite word $u \circ v$ is defined by $(u_1 \cdots u_r v_1 \cdots v_s) \in W(r+s)$.

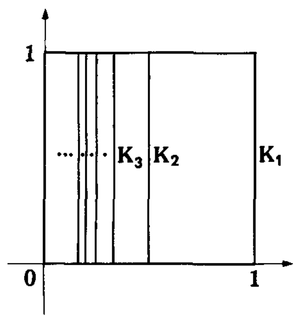



Fig. 2.

$\{f_n\}_{n\geq 1}$ such that $Q=K(f_1, f_2, \cdots)$ using compositions of a dilation, a rotation, a translation and $f(x, y)=x/2$.

Theorem 4.6 raises the following question: is it true or not that for any locally connected continuum $Q\subset X$, there exists a finite set of weak contractions $\{f_j\}_{1\leq j\leq m}$ of $X$ such that $Q=K(f_1, \cdots, f_m)$? In other words, is it possible to characterize locally connected continua by the self-similarity defined by (2.3)?

Note that for a fixed $m\geq 2$, there exists a locally connected continuum $Q\subset \boldsymbol{R}^m$ for which $Q\neq K(f_1, \cdots, f_m)$ for any $m$ weak contractions $\{f_j\}$ of $\boldsymbol{R}^m$. For example, an $(m-1)$-dimensional sphere in $\boldsymbol{R}^m$ has the required property by the Lusternik-Schnirelman-Borsuk theorem (Granas [15, p. 50]).

Finally, combining Theorems 4.1 and 4.6, we have immediately

COROLLARY 4.7 (Williams' Theorem D for $m=2$). *Let $f_1$ and $f_2$ be one to one contractions of $X$ satisfying* $\mathrm{Lip}(f_1)+\mathrm{Lip}(f_2)<1$ *and* $\mathrm{Fix}(f_1)\neq \mathrm{Fix}(f_2)$. *Then the mapping $\psi: \Sigma\to K(f_1, f_2)$ is a homeomorphism.*

## 5. Cut Points, End Points and Simple Links

To state further properties of self-similar sets, we need some definitions.

DEFINITION 5.1. A point $p$ of a connected set $Q$ is said to be a *cut point* of $Q$ provided that $Q-p$ is the sum of two mutually separated sets; an *end point* of $Q$ provided that there exist arbitrarily small neighborhoods of $p$ in $Q$ each of whose boundaries consists of a single point. Two points $p$, $q$ of a connected set $Q$ are said to be *conjugate* provided that no points separate $p$ and $q$ in $Q$. If $p$ is neither a cut point nor an end point of a connected set $Q$, the set consisting of $p$ together with all points of $Q$ conjugate to $p$ is called a *simple link* of $Q$. A continuum $Q$ is said to be an *acyclic curve* provided that it is locally connected and contains no simple links.

It is known that any simple link of a continuum $Q$ is a *nondegenerate* continuum; that is, it contains more than one point ([53, p. 64]). Every point of $Q$ is either a cut point, an end point or a point of a single simple link of $Q$. We now state our main theorem in this section.

THEOREM 5.2. *Suppose that* $\{f_j\}_{1\leq j\leq m}$ *is a finite set of one to one weak contractions of* $X$ *such that* $\mathrm{Fix}(f_i)\neq\mathrm{Fix}(f_j)$ *for some* $i\neq j$. *Suppose further that the set* $K=K(f_1,\cdots,f_m)$ *is an acyclic curve. Then either* $K$ *is a simple arc or* $K$ *has an infinite number of end points.*

*Proof.* Put $K(w)=f_{w_1}\circ\cdots\circ f_{w_n}(K)$ for any $w=(w_1\cdots w_n)\in W(n)$. Suppose that $K$ has a finite number of end points, say $e^1, e^2,\cdots,e^N$. Then it suffices to show $N=2$, since a continuum is a simple arc if and only if it has exactly two non-cut points ([53, p. 54]). Suppose, on the contrary, that $N\geq 3$. The remainder of the proof is devoted to demonstrating a contradiction.

*1st Step. There exists a finite word* $w\in W(n)$ *for some* $n$ *for which every point of* $K(w)$ *is a cut point of* $K$.

*Proof.* Suppose, on the contrary, that $K(u)$ contains at least one of the end points of $K$ for any $u\in W(n)$, $n\geq 1$. Take a sufficiently large integer $n$ so that

$$\mathrm{diam}(K(u))\leq\Omega^n(\mathrm{diam}(K))<\frac{1}{2}\min_{i\neq j} d(e^i, e^j), \tag{5.1}$$

for any $u\in W(n)$ where $\Omega(t)=\max_{1\leq j\leq m}\Omega_{f_j}(t)$. Obviously (5.1) contradicts the connectedness of $K$. Thus there exists a word $w\in W(n)$ possessing the required property. Put $F=f_{w_1}\circ\cdots\circ f_{w_n}$ and $p=\mathrm{Fix}(F)$ for brevity. Evidently $F(K)=K(w)$ has exactly $N$ end points $\{F(e^j)\}$. Note that $p$ is not an end point of $F(K)$. For otherwise, $p=F(e^j)$ for some $j$; hence $p=e^j$, contrary to the above definition of $K(w)$.

*2nd Step. There exists a simple arc* $A_j$ *joining* $e^j$ *and* $p$ *for* $1\leq j\leq N$ *such that* $A_i\cdot A_j=p$ *for any* $i\neq j$.

*Proof.* Since $F(e^j)$ is a cut point of $K$, we have a separation

$$K-F(e^j)=P(j)+Q(j), \tag{5.2}$$

where $\bar{P}(j)\cdot Q(j)=P(j)\cdot\bar{Q}(j)=\phi$ and $P(j)$ contains the connected set $F(K)-F(e^j)$. Then there exists a non-cut point $q^j$ of $Q(j)$ such that $q^j\neq F(e^j)$ since $\bar{Q}(j)=Q(j)+F(e^j)$ is a nondegenerate continuum. Evidently $q^j$ is an end point not only of $\bar{Q}(j)$ but also of $K$. We also have $q^i\neq q^j$ for any $i\neq j$ since $\bar{Q}(i)\cdot\bar{Q}(j)=\phi$ for any $i\neq j$. Therefore $\bar{Q}(j)$ has exactly two end points $q^j$ and $F(e^j)$; hence $\bar{Q}(j)$ is a simple arc. Thus this enables us to define the permutation $\pi$ on the set $\{1, 2,\cdots, N\}$ such that $e^j=q^{\pi(j)}$ (Fig. 3(a)).

Now we define

$$S_j(n)=\bar{Q}(\pi(j))+F(\bar{Q}(\pi^2(j)))+\cdots+F^{n-1}(\bar{Q}(\pi^n(j))). \tag{5.3}$$

Then (5.3) implies that $S_j(n)$ is a simple arc joining $e^j$ and $F^n(e^{\pi^n(j)})$, and that $S_j(1)\leq S_j(2)\leq\cdots$ (Fig. 3(b)). Put $A_j=\lim_{n\to\infty}S_j(n)$ in $\mathscr{C}(X)$. We first show that the set $A_j$ is a simple arc. For otherwise, $p\in\bar{Q}(i)$ for some $i$; hence $p$ is an end point of $F(K)$, contrary to the result in 1st Step. We next show that $A_i\cdot A_j=p$ for any $i\neq j$. For

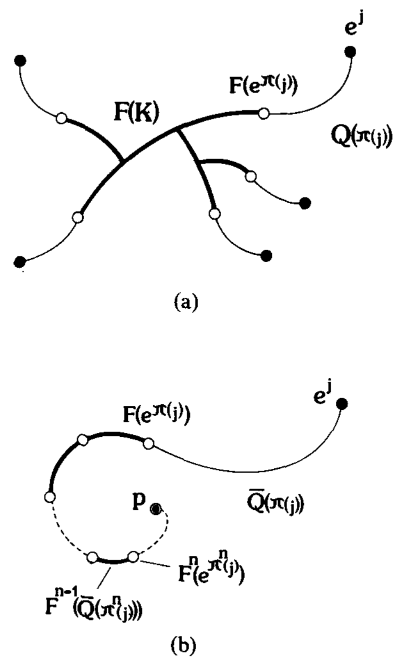



Fig. 3. (a) $F(K)$ is the heavy curve. The end points of $K$ and $F(K)$ are indicated by ● and ○ respectively.
(b) Simple arc $S_j(n)$.

otherwise, there exist two integers $r \geq s$ satisfying

$$F^{r-1}(\bar{Q}(\pi^r(i))) \cdot F^{s-1}(\bar{Q}(\pi^s(j))) \neq \phi . \tag{5.4}$$

Since $\bar{Q}(\pi^r(i)) \cdot \bar{Q}(\pi^r(j)) = \phi$, we have $r > s$; hence $F^{r-s}(\bar{Q}(\pi^r(i))) \cdot \bar{Q}(\pi^s(j)) \neq \phi$. Then it follows that $F(e^{\pi^s(j)}) \in F^{r-s}(\bar{Q}(\pi^r(i)))$ since $F(e^{\pi^s(j)}) = F(K) \cdot \bar{Q}(\pi^s(j))$. Hence $r - s = 1$ and $\pi^{r-1}(i) = \pi^s(j)$, contrary to $i \neq j$. Thus $A_i \cdot A_j = p$ for any $i \neq j$ as required. Note that $A_i + A_j$ is a simple arc joining $e^i$ and $e^j$ through the point $p$.

*3rd Step.* We are now ready to prove our theorem. Let $f_s$ be one of the weak contractions $\{f_j\}$ satisfying $p \neq \mathrm{Fix}(f_s)$. Note that every point of $f_s \circ F(K)$ is a cut point of $K$ since $f_s : K \to K(s)$ is a homeomorphism. By the same arguments as in 1st and 2nd Steps, we conclude that for any $i \neq j$, there exists a simple arc joining $e^i$ and $e^j$ through the point $p' = \mathrm{Fix}(f_s \circ F) \neq p$.

Consider now three end points $e^1$, $e^2$ and $e^3$. Since $A_i \cdot A_j = p$ for any $i \neq j$, there exist at least two simple arcs, say $A_1$ and $A_2$, such that $p' \notin A_1 + A_2$. Thus we have two different simple arcs joining $e^1$ and $e^2$. This contradicts the fact that a locally connected continuum is an acyclic curve if and only if there exists a unique simple arc joining any two points ([53, p. 89]). This completes the proof. □

A finite sequence of sets $\{Q_1, Q_2, \cdots, Q_n\}$ is said to be a *regular chain* provided that $Q_i \cdot Q_{i+1}$ consists of exactly one point for any $1 \leq i \leq n-1$ whereas $Q_i \cdot Q_j = \phi$ if

$|i-j|>1$. Then we have

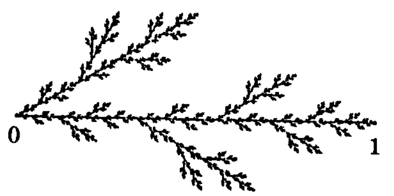
THEOREM 5.3. *Let $\{f_j\}_{1\le j\le m}$ be a finite set of one to one weak contractions of $X$ satisfying* $\mathrm{Fix}(f_i)\ne\mathrm{Fix}(f_j)$ *for some $i\ne j$. Let $K_j=f_j(K(f_1,\cdots,f_m))$, $1\le j\le m$, for brevity. For any $i\ne j$, suppose that the set $K_i\cdot K_j$ consists of at most one point and that there exists a unique regular chain joining $K_i$ and $K_j$. Then either $K=K(f_1,\cdots,f_m)$ is a simple arc or $K$ has an infinite number of end points.*

*Proof.* By Theorems 5.2, it suffices to show that $K$ is acyclic. Suppose, on the contrary, that $K$ has a simple link. Since any two conjugate points of a locally connected continuum lie together on a Jordan closed curve [53, p. 79], there exists a Jordan closed curve $J$ in $K$ such that $J\le K(w)$ and $J\cdot f_w(K_r')\ne\phi\ne J\cdot f_w(K_s')$ for some $w\in W(n)$ and some $r\ne s$, where $K_j'=K-\sum_{i\ne j}K_i$. Hence $J'=f_w^{-1}(J)$ satisfies $J'\cdot K_r'\ne\phi\ne J'\cdot K_s'$, contrary to the assumption. □

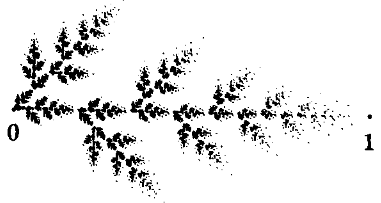
EXAMPLE 1. Let $X=\boldsymbol{C}$ with the usual Euclidean distance and put

$$f_1(z)=\alpha\bar{z} \quad\text{and}\quad f_2(z)=|\alpha|^2+(1-|\alpha|^2)\bar{z}\,, \tag{5.5}$$

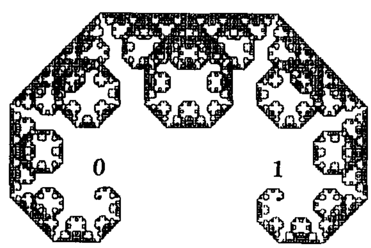
where $\alpha$ is a complex parameter satisfying $|\alpha|<1$, $|1-\alpha|<1$ and $\mathrm{Im}\,\alpha\ne0$. Then it is easily seen that $K=K(f_1,f_2)$ is not a simple arc and that $K_1\cdot K_2=|\alpha|^2$; hence $K$ has an infinite number of end points (Fig. 4(a) and (b)).




(a) $\alpha=\dfrac{1}{2}+\dfrac{\sqrt{3}}{6}i$.




(b) $\alpha=0.3+0.3i$.




(c) Lévy curve.

Fig. 4.

EXAMPLE 2. There exist two contractions $f_1, f_2$ such that the set $K=K(f_1,f_2)$ has an infinite number of simple links. For example, let $X=\boldsymbol{C}$ and put

$$f_1(z)=\alpha z \quad\text{and}\quad f_2(z)=(1-\alpha)z+\alpha\,, \tag{5.6}$$

where $\alpha$ is a complex parameter satisfying $|\alpha|<1$ and $|1-\alpha|<1$. It was pointed out by Lévy [28] that for $\alpha=1/2+i/2$, the measure of $K$ is positive and that the set of multiple points of $K$ is uncountable and dense in $K$ (Fig. 4(c)).

## 6. Parameterization

In this section, we will discuss the parameterizations of self-similar sets using some kind of functional equations. First of all, we have

THEOREM 6.1 (de Rham [42]). *Let $f_1$ and $f_2$ be two contractions of $X=R^p$. Then the functional equation*

$$G(t)=\begin{cases} f_1(G(2t)) & \text{for } 0\le t\le \dfrac{1}{2}, \\ f_2(G(2t-1)) & \text{for } \dfrac{1}{2}\le t\le 1, \end{cases} \tag{6.1}$$

*has a unique continuous solution if and only if*

$$f_1(\mathrm{Fix}(f_2))=f_2(\mathrm{Fix}(f_1)). \tag{6.2}$$

Note that de Rham's theorem gives a parameterization of the set $K(f_1, f_2)$ if the condition (6.2) is fulfilled. Indeed, we have

$$G([0,1])=G\left(\left[0,\frac{1}{2}\right]\right)+G\left(\left[\frac{1}{2},1\right]\right)=f_1(G([0,1])+f_2(G([0,1])$$

and therefore $G([0,1])=K(f_1, f_2)$ by Theorem 3.1.

We now generalize de Rham's Theorem 6.1. The following definitions are essentially taken from Milnor-Thurston [35]:

DEFINITION 6.2. A continuous function $h$ of $[a, b]$ is said to be *piecewise-monotone* provided that the interval $[a, b]$ is subdivided into finite subintervals so that the restriction of $h$ to each subinterval is strictly monotone.

DEFINITION 6.3. For any function $H$: $[0,1]\to[0,1]$, define the mapping $\nu_H$: $[0,1]\to\Sigma$ by setting

$$\nu_H(t)=(A(t), A(H(t)), \cdots, A(H^n(t)), \cdots) \tag{6.3}$$

where $A(t)=[mt]+[1-t]$ for $0\le t\le 1$. $\nu_H(t)$ is called the *itinerary* of a point $t$ under $H$.

Note that $\nu_H$ is discontinuous for any $H$ since $\Sigma$ is totally disconnected. However, for some kind of $H$, the mapping $\nu_H$ is 'almost continuous' in the following sense.

LEMMA 6.4. *Let $h_j$: $[(j-1)/m, j/m]\to[0,1]$ be piecewise-monotone for any $1\le j\le m$. Put $H(t)=h_{A(t)}(t)$ for brevity. Then there exist the limits $\nu_H(s\pm)$ in $\Sigma$ for any $0<s<1$. Moreover $\nu_H$ is continuous on*

$$\Gamma_H = \{t \in [0, 1];\ H^n(t) \neq j/m \text{ for any } n \geq 0 \text{ and } 1 \leq j \leq m-1\}\,, \tag{6.4}$$

*which is a dense set of* [0, 1].

*Proof.* For any fixed $s \in (0, 1)$ and $N \geq 1$, there exists a sufficiently small $\varepsilon > 0$ such that each of the functions

$$H(t),\ H^2(t), \cdots, H^N(t) \tag{6.5}$$

is strictly monotone, either increasing or decreasing on $(s, s+\varepsilon)$ and that each of

$$A(t),\ A(H(t)), \cdots, A(H^N(t)) \tag{6.6}$$

is independent of the choice of $t \in (s, s+\varepsilon)$. Obviously this implies that $v_H(s+)$ exists. Similarly $v_H(s-)$ exists for any $0 < s < 1$. Suppose now $s \in \Gamma_H$. Then it follows that each of the functions (6.6) is continuous in a sufficiently small neighborhood of $s$. Therefore $v_H(s+) = v_H(s-) = v_H(s)$; hence $v_H$ is continuous at $s$. Since each $h_j$ is piecewise-monotone, the set $\gamma_{n,j} = \{t;\ H^n(t) = j/m\}$ is finite; hence $\Gamma_H = [0, 1] - \sum_{n,j} \gamma_{n,j}$ is obviously dense in [0, 1]. □

Using this lemma, we can prove the following generalization of Theorem 6.1.

THEOREM 6.5. *Let* $\{f_j\}_{1 \leq j \leq m}$ *be a finite set of weak contractions of* $X$ *and* $\{h_j\}_{1 \leq j \leq m}$ *be the same functions as in Lemma* 6.4. *Then the functional equation*

$$G(t) = \begin{cases} f_1(G(h_1(t))) & \text{for } 0 \leq t \leq \dfrac{1}{m}\,, \\ \quad\vdots & \quad\vdots \\ f_m(G(h_m(t))) & \text{for } \dfrac{m-1}{m} \leq t \leq 1\,, \end{cases} \tag{6.7}$$

*has a unique continuous solution* $G$: $[0, 1] \to X$ *if and only if*

$$\psi \circ v_H\left(\frac{j}{m}+\right) = \psi \circ v_H\left(\frac{j}{m}-\right) \qquad \text{for any} \quad 1 \leq j \leq m-1\,, \tag{6.8}$$

*where* $\psi: \Sigma \to K(f_1, \cdots, f_m)$ *is the mapping defined in Theorem* 3.2. *If in addition each* $h_j$ *is onto, the continuous solution* $G$ *of* (6.7) *satisfies* $G([0, 1]) = K(f_1, \cdots, f_m)$.

*Proof.* Obviously the condition (6.8) is necessary, since we have

$$G(t) = f_{A(t)} \circ G \circ H(t) = f_{A(t)} \circ f_{A(H(t))} \circ \cdots = \psi \circ v_H(t)\,. \tag{6.9}$$

We now show the sufficiency. Put $F(t) = \psi \circ v_H(t)$ for brevity. Then $F$ is continuous on $\Gamma_H$ by Lemma 6.4. The condition (6.8) implies $F((j/m)+) = F((j/m)-)$ for $1 \leq j \leq m-1$. Since $F(t) = f_{A(t)} \circ F(H(t))$ for any $t$, it follows that $F(s+)$ as well as $F(s-)$ is equal to one of $f_{A(s)} \circ F((j/m)\pm)$ for any $s \in \gamma_{1,j}$. Therefore $F(s+) = F(s-)$. Similarly one can show that $F(s+) = F(s-)$ for any $s \in \gamma_{n,j}$, $n \geq 1$. Now define $\tilde{F}(t) = F(t)$ if $t \in \Gamma_H$ and $\tilde{F}(t) = F(t+)$ otherwise. Then it is easily seen that $\tilde{F}$ is continuous on [0, 1]. Since

$H(\Gamma_H) \subset \Gamma_H$, we have

$$\tilde{F}(t) = f_j(\tilde{F}(h_j(t))) \qquad \text{for} \quad t \in \Gamma_H \cdot \left(\frac{j-1}{m}, \frac{j}{m}\right).$$

Hence $\tilde{F}$ is a continuous solution of (6.7) since $\Gamma_H$ is dense in [0, 1]. The uniqueness of such a solution follows from (6.9). It is obvious that $G([0, 1]) = K(f_1, \cdots, f_m)$ if each $h_j$ is onto for the continuous solution $G$ of (6.7). This completes the proof. □

Applying the above theorem to the case $h_j = mt - j + 1$, we have

COROLLARY 6.6. *Let $\{f_j\}_{1 \le j \le m}$ be a finite set of weak contractions of $X$. Then the functional equation*

$$G(t) = \begin{cases} f_1(G(mt)) & \text{for} \quad 0 \le t \le \dfrac{1}{m}, \\ \quad \vdots & \qquad \vdots \\ f_m(G(mt - m + 1)) & \text{for} \quad \dfrac{m-1}{m} \le t \le 1, \end{cases} \tag{6.10}$$

*has a unique continuous solution if and only if*

$$f_2(\mathrm{Fix}(f_1)) = f_1(\mathrm{Fix}(f_m)), \cdots, f_m(\mathrm{Fix}(f_1)) = f_{m-1}(\mathrm{Fix}(f_m)). \tag{6.11}$$

The continuous solution $G$ of (6.10) gives a parameterization of $K(f_1, \cdots, f_m)$ since each $h_j$ is onto. The conditions (6.11) are frequently referred to as the *D-conditions*.

As applications of this kind of functional equations, Denny [10] gave an example of a uniformly continuous function $f$: $\boldsymbol{R}^m \to (0, 1)$ which is almost everywhere one to one; the author [18] showed the existence of periodic solutions of a certain functional equation, which are continuous and of bounded variation.

EXAMPLE. Consider the contractions defined by (5.5). Since $f_1^2(\mathrm{Fix}(f_2)) = f_2(\mathrm{Fix}(f_1))$, it is easily seen that the continuous solution $G$ of (6.7) for

$$h_1(t) = 1 - |(2 + \sqrt{2})t - 1| \quad \text{and} \quad h_2(t) = 2t - 1 \tag{6.12}$$

gives a parameterization of the set $K = K(f_1, f_2)$ illustrated in Fig. 4(a) and (b). Note that $h_1$ has two fixed points (Fig. 5) and the set $G(\sum_{n \ge 0} H^{-n}(1))$ gives all end points of $K$.

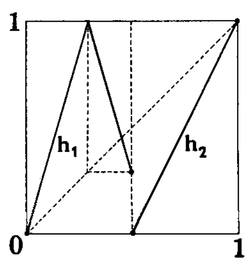



Fig. 5.

Finally we will study the case where $G$ is a homeomorphism. Compare with Theorem 5.3.

THEOREM 6.7. *Let $\{f_j\}_{1\le j\le m}$ be a finite set of one to one weak contractions of $X$ satisfying* $\mathrm{Fix}(f_i)\neq\mathrm{Fix}(f_j)$ *for some $i\neq j$. Suppose that the set $\{K_1,\cdots,K_m\}$ is a regular chain where $K_j=f_j(K(f_1,\cdots,f_m))$ for $1\le j\le m$. Then the set $K=K(f_1,\cdots,f_m)$ is a simple arc if and only if there exist linear homeomorphisms $h_j$: $[(j-1)/m, j/m]\to[0,1]$, $1\le j\le m$, such that $\psi\circ\nu_H$ satisfies the condition* (6.8).

*Proof.* We first show the condition is necessary. Suppose $K$ is a simple arc. Since each $K_j$ is also a simple arc, the point $K_i\cdot K_{i+1}$ is an end point of both $K_i$ and $K_{i+1}$. Let $g_j$: $[0,1]\to K_j$ be a homeomorphism satisfying $g_j(1)=g_{j+1}(0)$ for $1\le j\le m-1$. Then $G(t)=g_{A(t)}(mt-A(t)+1)$: $[0,1]\to K$ becomes a homeomorphism. Define

$$h_j(t)=G^{-1}\circ f_j^{-1}\circ G(t)\qquad\text{for}\quad \frac{j-1}{m}\le t\le\frac{j}{m}.$$

Then obviously $h_j$: $[(j-1)/m, j/m]\to[0,1]$ is a homeomorphism and $G$ satisfies the equation (6.7); hence $\psi\circ\nu_H$ satisfies the condition (6.8). It is obvious that each $h_j$ can be replaced by a linear homeomorphism $\tilde{h}_j$ such that

$$\tilde{h}_j\left(\frac{j-1}{m}\right)=h_j\left(\frac{j-1}{m}\right)\quad\text{and}\quad \tilde{h}_j\left(\frac{j}{m}\right)=h_j\left(\frac{j}{m}\right).$$

We now show the sufficiency. It suffices to show the solution $G$ of (6.7) is a homeomorphism. Suppose, on the contrary, that $G(t_1)=G(t_2)$ for some $t_1<t_2$. Let $\Xi=\{(s,t);\ G(s)=G(t)\text{ for }0\le s,t\le 1\}$. Without loss of generality, we can assume

$$|t_1-t_2|=\max_{(s,t)\in\Xi}|s-t|. \tag{6.13}$$

Then $A(t_1)<A(t_2)$. For otherwise, we have $(H(t_1),H(t_2))\in\Xi$ and $|H(t_1)-H(t_2)|=m|t_1-t_2|$, contrary to (6.13). Since $\{K_1,\cdots,K_m\}$ is a regular chain, it follows that $A(t_1)=A(t_2)-1$, say $l$. Thus $G(t_1)=G(t_2)=G(l/m)=K_l\cdot K_{l+1}$. Then $(t_1, l/m)\in\Xi$ implies $(h_l(t_1), h_l(l/m))\in\Xi$ and $|h_l(t_1)-h_l(l/m)|=m|t_1-l/m|$; hence $t_2-t_1\ge l-mt_1$. Similarly $(t_2, l/m)\in\Xi$ implies $(h_{l+1}(t_2), h_{l+1}(l/m))\in\Xi$ and $|h_{l+1}(t_2)-h_{l+1}(l/m)|=m|t_2-l/m|$; hence $t_2-t_1\ge mt_2-l$. Combining two inequalities, we have $t_1\ge l/m$ for $m\ge 3$, contrary to $A(t_1)=l$. For the case $m=2$, it is easily seen that $t_1+t_2=1$ and $\mathrm{Fix}(f_1)=\mathrm{Fix}(f_2)$, contrary to the assumption. This completes the proof. □

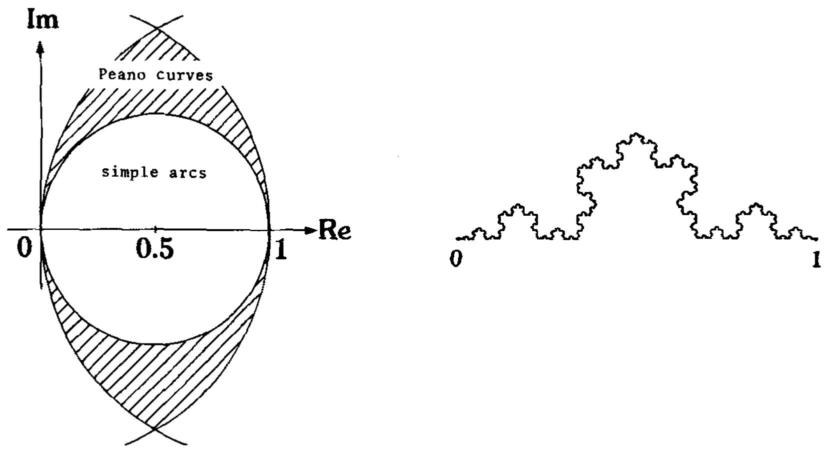
REMARK. The condition of the above theorem for $m=2$ takes the following form: $\{f_1,f_2\}$ satisfies at least one of the following four conditions:

(a) $f_1(\mathrm{Fix}(f_2))=f_2(\mathrm{Fix}(f_1))$;
(b) $f_1\circ f_2(\mathrm{Fix}(f_1))=f_2^2(\mathrm{Fix}(f_1))$;
(c) $f_1^2(\mathrm{Fix}(f_2))=f_2\circ f_1(\mathrm{Fix}(f_2))$;
(d) $f_1(\mathrm{Fix}(f_1\circ f_2))=f_2(\mathrm{Fix}(f_2\circ f_1))$.

EXAMPLE. Let $X=C$ with the usual Euclidean distance and

$$f_1(z)=\alpha\bar{z} \quad \text{and} \quad f_2(z)=(1-\alpha)\bar{z}+\alpha\,, \tag{6.14}$$

where $\alpha$ is a complex parameter satisfying $|\alpha|<1$ and $|1-\alpha|<1$. Since $\{f_1, f_2\}$ satisfies the $D$-condition (6.11) and $K_1 \cdot K_2=\alpha$ for any $|\alpha-1/2|<1/2$, it follows that $K=K(f_1, f_2)$ is a simple arc by Theorem 6.7; hence $\dim_T(K)=1$. Note that Riemann's outer area of $K$ is always zero by Theorem 4.2. Compare with the examples given by Osgood [39] and by Besicovitch-Schoenberg [2], which are simple arcs with positive area. On the other hand, if $|\alpha-1/2|\geq 1/2$, it is clear that $K(f_1, f_2)$ is a closed triangle with vertices 0, 1 and $\alpha$ (Fig. 6(a)); therefore $\dim_T(K)=2$. It was pointed out by de Rham [42] that for $\alpha=1/2+\sqrt{3}\,i/6$, the solution $G$ of (6.10) gives the curve studied by von Koch [25] (Fig. 6(b)) and that for $\alpha=1/2+e^{i\theta}/2$, $G$ gives the space-filling curve studied by Pólya [41].




(a) (b)

Fig. 6.

## 7. Regularity of $G$

In this section, we will discuss the regularity of the continuous solution $G$ of (6.10). Throughout this section let $X$ be a closed subset of a Banach space $E$ and $\{f_j\}_{1\leq j\leq m}$ be a finite set of weak contractions of $X$ satisfying $D$-conditions (6.11). First of all, we have

THEOREM 7.1. *The solution $G$ is Hölder-continuous with exponent* $\alpha = -\log\delta/\log m$, *where* $\delta=\max_{1\leq j\leq m}\mathrm{Lip}(f_j)$.

*Proof.* For any $t\neq s$, let $n$ be an integer such that $m^{-n-1}<|t-s|\leq m^{-n}$. Then it is easily seen that $G(t)$, $G(s)\in K(w)+K(w')$ with $K(w)\cdot K(w')\neq\phi$ for some $w, w'\in W(n)$. Therefore $\|G(t)-G(s)\|\leq \mathrm{diam}(K(w))+\mathrm{diam}(K(w'))$; hence

$$\frac{\|G(t)-G(s)\|}{|t-s|^\alpha}\leq 2m^\alpha\,\mathrm{diam}(K)(\delta m^\alpha)^n\leq 2m^\alpha\,\mathrm{diam}(K)\,. \quad \square$$

A mapping $f$: $[0, 1]\to E$ is said to be of *bounded p-variation* provided that

$$\sup_{\Delta}\left(\sum_i \|f(t_{i+1})-f(t_i)\|^p\right)^{1/p}<\infty\,, \tag{7.1}$$

where the supremum extends over all subdivisions $\Delta$: $0=t_0<t_1<\cdots<t_n=1$ of $[0, 1]$. For $p=1$, we usually say that $f$ is of bounded variation. Note that every Hölder-continuous mapping with exponent $\alpha$ is of bounded $1/\alpha$-variation. Then

THEOREM 7.2. *Suppose that each $f_j$ is one to one and* $\mathrm{Fix}(f_1)\neq\mathrm{Fix}(f_m)$. *If $\{f_j\}$ satisfies for some $\alpha>0$,*

$$\sum_{j=1}^{m}(\mathrm{Lip}(f_j^{-1}))^{-\alpha}>1\,, \tag{7.2}$$

*then the solution G is not of bounded $\alpha$-variation.*

*Proof.* Let $v(n,j)=\|G(j/m^n)-G((j-1)/m^n)\|$, $1\leq j\leq m^n$, $n\geq 1$ for brevity. Then it is easily verified that

$$\sum_{j=1}^{m^n}(v(n,j))^{\alpha}\geq\left(\sum_{j=1}^{m}(\mathrm{Lip}(f_j^{-1}))^{-\alpha}\right)\cdot\sum_{j=1}^{m^{n-1}}(v(n-1,j))^{\alpha}.$$

Therefore, by (7.2), it follows that $G$ is not of bounded $\alpha$-variation, since $v(0,1)=\|\mathrm{Fix}(f_1)-\mathrm{Fix}(f_m)\|\neq 0$. □

We now turn to the differentiability of $G$. In this respect, we have the following theorem by applying the same method as in Lax [27].

THEOREM 7.3. *Suppose that $\{f_j\}$ satisfies*

$$\prod_{j=1}^{m}\mathrm{Lip}(f_j)<m^{-m}. \tag{7.3}$$

*Then the Fréchet derivative $DG(t)$ of the solution G is equal to zero for almost every t.*

*Proof.* Since almost every number is normal in the scale of $m$ (Billingsley [3, p. 34]), it suffices to show that $DG(t)=0$ for every normal number $t\in[0, 1]$. Let $s\neq t$ be an arbitrary number and let $t=\sum_{n\geq 1}t_n m^{-n}$ and $s=\sum_{n\geq 1}s_n m^{-n}$. Let $N\geq 1$ be the smallest integer such that $t_N\neq s_N$ and let $M>N$ be the smallest integer such that $t_M\geq 1$ or $t_M\leq m-2$ according to whether $t>s$ or $t<s$ respectively. Then it is easily verified that

$$m^{-M}<|s-t|<m^{-N+1}\quad\text{and}\quad M=N+o(N)\qquad\text{as}\quad N\to\infty\,. \tag{7.4}$$

Note that (7.4) implies $s\to t$ if and only if $N\to\infty$.

On the other hand, we have from the equation (6.10),

$$\|G(s)-G(t)\|\leq\mathrm{diam}(K)\left(\prod_{j=1}^{m}a_j^{r_j}\right), \tag{7.5}$$

where $a_j = \mathrm{Lip}(f_j)$ and $r_j = \#\{1 \le i \le N-1;\ t_i = j\}$ for $1 \le j \le m$. Since $r_j = N/m + o(N)$ as $N \to \infty$, we have from (7.5),

$$\left\| \frac{G(s)-G(t)}{s-t} \right\| \le \mathrm{diam}(K) \exp\left( \frac{N}{m} \log\left( m^m \prod_{j=1}^{m} a_j \right) + o(N) \right).$$

Since $m^m \prod_{j=1}^{m} a_j < 1$, it follows that $DG(t) = 0$. This completes the proof. □

COROLLARY 7.4. *Suppose that each $f_j$ is a strictly monotone increasing function and* $\mathrm{Fix}(f_1) < \mathrm{Fix}(f_m)$. *Suppose further that* $\{f_j\}$ *satisfies* (7.3). *Then the solution $G$ is a strictly monotone increasing and purely singular function.*

*Proof.* Since $\{f_j\}$ satisfies (6.11), it is clear that the set $K = [\mathrm{Fix}(f_1), \mathrm{Fix}(f_m)]$ and $\{K_j\}$ is a regular chain. Hence $G$ is a homeomorphism by Theorem 6.7; therefore it is strictly monotone increasing. Also it follows from Theorem 7.3 that $G'(t) = 0$ almost everywhere. □

EXAMPLE. Consider the contractions defined by (4.1). If $a + b = 1$ (this is also a special case of (6.14)) and $a \ne 1/2$, $\{f_1, f_2\}$ satisfies the conditions of Corollary 7.4; therefore $G_a(t) = G(t)$ is a strictly monotone increasing and purely singular function with a parameter $a$. This function was studied by Salem [47]. It is known that $G_a(t)$ is the distribution function for the Bernoulli trials of unfair coin tossings. See also Lomnicki-Ulam [29] and de Rham [42, 43].

Concerning the non-differentiability of $G$, we have

THEOREM 7.5. *Suppose that each $f_j$ is one to one and that $\{f_j\}$ satisfies*

$$\prod_{j=1}^{m} \mathrm{Lip}(f_j^{-1}) < m^m. \tag{7.6}$$

*Then the solution $G$ is not Fréchet differentiable at almost every $t$. If in addition* $\mathrm{Lip}(f_j^{-1}) < m$ *for any $1 \le j \le m$, then $G$ is nowhere differentiable.*

*Proof.* We first show the non-differentiability of $G$ at every normal number $t$. Let $t = \sum_{n \ge 1} t_n m^{-n}$. For any $N \ge 1$, take a suitable number $s_N \in [0, 1]$ such that $\|G(s_N) - G(H^N(t))\| \ge (1/2)\,\mathrm{diam}(K)$ where $H(t) = mt - A(t) + 1$. Put $t^{(N)} = \sum_{j=1}^{N} t_j m^{-j} + s_N m^{-N}$. Then from the equation (6.10),

$$\|G(t^{(N)}) - G(t)\| \ge \frac{1}{2} \mathrm{diam}(K) \prod_{j=1}^{m} b_j^{r_j}, \tag{7.7}$$

where $b_j = (\mathrm{Lip}(f_j^{-1}))^{-1}$ and $r_j = \#\{1 \le i \le N;\ t_i = j\}$ for $1 \le j \le m$. Since $|t^{(N)} - t| \le 2m^{-N}$, we have

$$\left\| \frac{G(t^{(N)}) - G(t)}{t^{(N)} - t} \right\| \ge \frac{1}{4} \mathrm{diam}(K) \exp\left( \frac{N}{m} \log\left( m^m \prod_{j=1}^{m} b_j \right) + o(N) \right).$$

Hence (7.6) implies that $G$ is not differentiable at $t$.

Next assume that $mb_j > 1$ for $1 \le j \le m$ instead of (7.6). Then the same argument as above can be applied to an arbitrary $t$, since

$$\prod_{j=1}^{m} b_j^{r_j} \geq b_*^N,$$

where $b_* = \min_{1 \leq j \leq m} b_j > 1/m$; hence

$$\left\| \frac{G(t^{(N)}) - G(t)}{t^{(N)} - t} \right\| \geq \frac{1}{4} \operatorname{diam}(K)(mb_*)^N .$$

This completes the proof. □

EXAMPLE. Consider the contractions defined by (6.14). Then, by Theorem 7.1, the solution $G_\alpha(t) = G(t)$ of (6.10) has Hölder-exponent

$$-\frac{\log \max(|\alpha|, |1-\alpha|)}{\log 2}.$$

In particular, Koch's curve ($\alpha = 1/2 + \sqrt{3}\, i/6$) is Hölder-continuous with exponent $\log 3/\log 4$, which can not be replaced by any larger value by Theorem 7.2. For almost every $t$, $G_\alpha'(t) = 0$ or $G_\alpha(t)$ is not differentiable according to whether $|\alpha(1-\alpha)| < 1/4$ or $> 1/4$ by Theorems 7.3 and 7.5. Note that the boundary curve $|\alpha(1-\alpha)| = 1/4$ is a lemniscate (Fig. 7). Moreover, if $|\alpha| > 1/2$ and $|1-\alpha| > 1/2$, then $G_\alpha(t)$ is nowhere differentiable, as shown by de Rham [43]. For Pólya's case ($\alpha = 1/2 + e^{i\theta}/2$), the above results were shown by Lax [27].

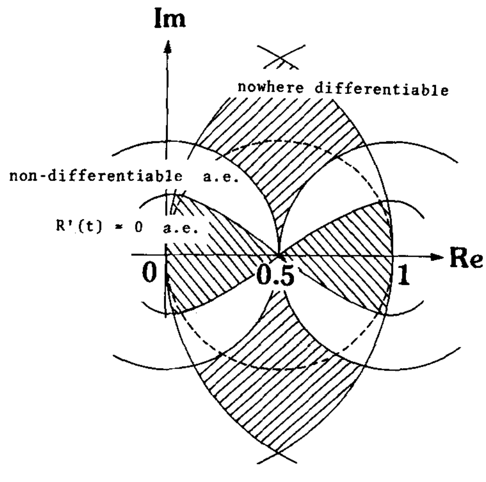



Fig. 7.

## 8. Reversed Equations

Recall that in Section 6 we obtained the continuous solution $G$ of (6.7) using the diagram:

$$\begin{array}{ccc} [0,1] & \xrightarrow{G} & K(f_1,\cdots,f_m) \\ & {\scriptstyle \nu_H} \searrow \quad \nearrow {\scriptstyle \psi} & \\ & \Sigma & \end{array} \tag{8.1}$$

Of course, such a solution does not exist if $K=K(f_1,\cdots,f_m)$ is not connected. Here we will discuss the existence of a non-trivial continuous mapping $R$ which maps $K(f_1,\cdots,f_m)$ into $[0,1]$. Let $\{g_j\}_{1\le j\le m}$ be a finite set of weak contractions of $X=[0,1]$ with the usual Euclidean distance and $\psi^*=\psi: \Sigma\to K(g_1,\cdots,g_m)$. Then the desired mapping $R$ will be obtained by the diagram:

$$\begin{array}{ccc} K(f_1,\cdots,f_m) & \xrightarrow{R} & K(g_1,\cdots,g_m)\subset[0,1] \\ & {\scriptstyle \psi^{-1}} \searrow \quad \nearrow {\scriptstyle \psi^*} & \\ & \Sigma & \end{array} \tag{8.2}$$

Indeed, we have

THEOREM 8.1. *Let $\{f_j\}_{1\le j\le m}$ and $\{g_j\}_{1\le j\le m}$ be two finite sets of weak contractions of $X$ and $[0,1]$ respectively. Then the functional equations*

$$\begin{cases} R(f_1(x))=g_1(R(x)) \\ \quad\vdots \qquad\quad \vdots \\ R(f_m(x))=g_m(R(x)) \end{cases} \tag{8.3}$$

*have a unique continuous onto solution $R$: $K(f_1,\cdots,f_m)\to K(g_1,\cdots,g_m)$ if and only if $\psi^*(\alpha)=\psi^*(\beta)$ whenever $\psi(\alpha)=\psi(\beta)$.*

*Proof.* It is clear that the condition is necessary, since

$$R(\psi(\alpha))=R\circ f_{\alpha_1}\circ\psi(\sigma(\alpha))=g_{\alpha_1}\circ\psi(\sigma(\alpha))=g_{\alpha_1}\circ g_{\alpha_2}\circ\cdots=\psi^*(\alpha) \tag{8.4}$$

for any $\alpha\in\Sigma$, where $\sigma: \Sigma\to\Sigma$ is the left-shift transformation.

We now show the sufficiency. Define the mapping $R$: $K(f_1,\cdots,f_m)\to K(g_1,\cdots,g_m)$ by $R(\psi(\alpha))=\psi^*(\alpha)$. The condition of the theorem implies that $R$ is well-defined. Then it is clear that $R$ satisfies the equations (8.3). We must show the continuity of $R$. Suppose, on the contrary, that $R$ is discontinuous at $\psi(\alpha)$ for some $\alpha\in\Sigma$. Then there exists a sequence $\{\alpha^{(n)}\}$ in $\Sigma$ such that

$$|\psi^*(\alpha)-\psi^*(\alpha^{(n)})|\ge\delta>0 \tag{8.5}$$

and $\psi(\alpha^{(n)})\to\psi(\alpha)$ as $n\to\infty$. Without loss of generality, we can assume $\alpha^{(n)}\to\bar\alpha$ as $n\to\infty$. Then we have $\psi(\bar\alpha)=\psi(\alpha)$ and therefore $\psi^*(\bar\alpha)=\psi^*(\alpha)$, contrary to (8.5). The uniqueness of such a solution is obvious from (8.4). □

As a corollary, we have immediately

COROLLARY 8.2. *Let $\{f_j\}_{1\le j\le m}$ be a finite set of weak contractions of $X$ such* *that* $\{f_j(K)\}_{1\le j\le m}$ *are pairwise disjoint where* $K=K(f_1,\cdots,f_m)$. *Then for any weak contractions* $\{g_j\}_{1\le j\le m}$ *of* [0, 1], *the reversed equations* (8.3) *have a unique continuous onto solution* $R\colon K(f_1,\cdots,f_m)\to K(g_1,\cdots,g_m)$.

EXAMPLE 1. Consider the contractions $f_1, f_2$ defined by (4.1) and put

$$g_1(t)=\frac{t}{2} \quad\text{and}\quad g_2(t)=\frac{t+1}{2}. \tag{8.6}$$

If $a+b<1$, the mapping $\psi\colon \Sigma\to K(f_1,f_2)$ becomes a homeomorphism by Theorem 4.4. Then there exists a unique continuous onto solution $R_{a,b}=R$: $K(f_1,f_2)\to K(g_1,g_2)=[0,1]$ by Corollary 8.2. Note that $R_{a,b}$ is monotone increasing and there exists a unique extension $\tilde{R}_{a,b}\colon [0,1]\to[0,1]$ of $R_{a,b}$, which is also monotone increasing and satisfies the equations (8.3) for *any* $x\in[0,1]$. In particular, if $a=b=1/3$, $\tilde{R}_{a,b}(t)$ is the well-known Cantor function. The functional equations for the Cantor function were studied by Sierpiński [48]. Note that, if $a=b$ $(<1/2)$, it is easily seen that

$$L_a(t)\equiv\int_0^1 e^{itx}d\tilde{R}_{a,a}(t)=e^{ti/2}\prod_{n\ge 0}\cos\left(\frac{1-t}{2}a^n\right). \tag{8.7}$$

It is known that $L_a(t)$ is not absolutely continuous (Kershner-Wintner [24]). Carleman [6, pp. 223–226] has shown that $L_a(t)$ does not tend to 0 as $|t|\to\infty$, if $a=q^{-1}$, where $q=3,4,5,\cdots$. Kershner [23] has shown that $L_a(t)=O((\log|t|)^{-\beta})$ if $a=p/q$, not the reciprocal of an integer, while $\beta$ is a positive function of $p$ and $q$. Note that this gives an example of a continuous function which is not absolutely continuous and satisfies the Riemann-Lebesgue lemma. See also Erdös [12].

EXAMPLE 2. De Rham [44] gave an example of a $C^1$-function $f(x,y)$ with two variables such that the set

$$f\left(\left\{(x,y);\frac{\partial f}{\partial x}=\frac{\partial f}{\partial y}=0\right\}\right)$$

contains an interval, which is analogous to Whitney's example [52]. De Rham's function $f$ is an extension of the solution $R$ of (8.3) for certain affine contractions of the plane satisfying the condition of Corollary 8.2.

EXAMPLE 3. If the continuous solution $G$ of (6.7) is a homeomorphism and each $h_j^{-1}$ is a weak contraction of [0, 1], then it is clear that $R=G^{-1}$: $K(f_1,\cdots,f_m)\to[0,1]$ satisfies the equations (8.3) for $g_j=h_j^{-1}$, $1\le j\le m$. For example, let $X=[0,1]$ with the usual Euclidean distance and put

$$f_1(x)=\frac{x}{1+x} \quad\text{and}\quad f_2(x)=\frac{1}{2-x}. \tag{8.8}$$

Then it is easily seen that the solution $R=G^{-1}\colon [0,1]\to[0,1]$ of (8.3) exists for the contractions $g_1, g_2$ defined by (8.6), which is known as Minkowski's function [36]. It was proved by Denjoy [8] that $R(t)$ is purely singular. See also Salem [47].

## 9. Space-Filling Curves

In this section, we will discuss various properties of the classical space-filling curves, which will be obtained by the continuous solution $G$ of the equation (6.7) for certain simple affine contractions.

We denote by $I^p$ the $p$-dimensional cube given by $[0, 1]^p$. The following theorem is a standard result. For the proof, see Vitushkin-Khenkin [51].

THEOREM 9.1. *Suppose that* $p<q$ *and* $f\colon I^p \to I^q$ *is an onto Hölder-continuous mapping with exponent* $\alpha$. *Then* $\alpha \le p/q$. *Moreover, there exists an onto Hölder-continuous mapping* $f\colon I^p \to I^q$ *with exponent* $p/q-\varepsilon$ *for any* $\varepsilon>0$. *If in addition p divides q, then one can take* $\varepsilon$ *to be zero.*

EXAMPLE 1. In 1890, Peano [40] gave the first example of a continuous planar curve $P_1(t)$ filling the unit square $I^2$ with vertices $0, 1, 1+i$ and $i$. It is easily seen that $P_1(t)$ is a continuous solution of (6.10) for the nine affine contractions:

$$\begin{cases} f_1(z)=\dfrac{z}{3}; & f_2(z)=-\dfrac{\bar{z}}{3}+\dfrac{1+i}{3}; \quad f_3(z)=\dfrac{z}{3}+\dfrac{2i}{3}; \\ f_4(z)=\dfrac{\bar{z}}{3}+\dfrac{1+3i}{3}; & f_5(z)=-\dfrac{z}{3}+\dfrac{2+2i}{3}; \quad f_6(z)=\dfrac{\bar{z}}{3}+\dfrac{1+i}{3}; \\ f_7(z)=\dfrac{z}{3}+\dfrac{2}{3}; & f_8(z)=-\dfrac{\bar{z}}{3}+\dfrac{3+i}{3}; \quad f_9(z)=\dfrac{z}{3}+\dfrac{2+2i}{3}. \end{cases} \tag{9.1}$$

Then, it follows that $P_1(t)$ is nowhere differentiable by Theorem 7.5 and satisfies

$$\|P_1(t)-P_1(s)\| \le 3\sqrt{5}\,|t-s|^{1/2}. \tag{9.2}$$

Note that the exponent $1/2$ in (9.2) can not be replaced by $1/2+\varepsilon$ for any $\varepsilon>0$ by Theorem 9.1. This also follows from Theorem 7.2. Cesàro [7] gave the analytic formula for $P_1$ and Moore [37] discussed a generalization of $P_1$ by geometrical observation. Using Moore's construction, Milne [34] gave an example of a mapping $f\colon I^1 \to I^p$, which is Hölder-continuous with exponent $p^{-1}$ and measure-preserving, that is, $\mu_p(A)=\mu_1(f^{-1}(A))$ for any Borel subset $A$ of $I^p$ where $\mu_p$ is the usual product measure on $I^p$.

EXAMPLE 2. In 1891, Hilbert [20] gave a simpler example of a continuous planar curve $P_2(t)$ filling $I^2$. It is easily seen that $P_2(t)$ is a continuous solution of (6.10) for the four affine contractions:

$$\begin{cases} f_1(z)=\dfrac{i}{2}\bar{z}; & f_2(z)=\dfrac{z}{2}+\dfrac{i}{2}; \\ f_3(z)=\dfrac{z}{2}+\dfrac{1+i}{2}; & f_4(z)=-\dfrac{i}{2}\bar{z}+\dfrac{2+i}{2}. \end{cases} \tag{9.3}$$

Then $P_2(t)$ is nowhere differentiable and satisfies

$$\|P_2(t)-P_2(s)\| \le 2\sqrt{5}\,|t-s|^{1/2}\,. \tag{9.4}$$

EXAMPLE 3. Sierpiński [49] gave a slightly different example of a planar curve $P_3(t)$ filling the square with vertices $1+i$, $-1+i$,$.-1-i$ and $1-i$. $P_3(t)$ is a unique continuous periodic solution with period 1 of the equation (6.7) for the four contractions:

$$\begin{cases} f_1(z)=\dfrac{i}{2}(z-1-i)\,; & f_2(z)=\dfrac{1}{2}(z-1-i)\,; \\ f_3(z)=-\dfrac{i}{2}(z-1-i)\,; & f_4(z)=-\dfrac{1}{2}(z-1-i) \end{cases} \tag{9.5}$$

and $h_j(t)=4t-1/8 \pmod 1$ (Fig. 8).

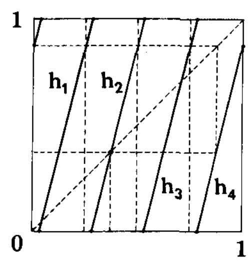



Fig. 8.

Note that $\{f_j\}$ satisfies

$$f_1\circ f_4(\mathrm{Fix}(f_2))=f_2\circ f_4(\mathrm{Fix}(f_2))=f_3\circ f_4(\mathrm{Fix}(f_2))=f_4^2(\mathrm{Fix}(f_2))\,.$$

We have $\sum_{j=1}^{m}(\mathrm{Lip}(f_j))^2=1$ in all examples.

## 10. Hausdorff Dimension of Self-Similar Sets

We begin with some definitions.

DEFINITION 10.1. For any $\alpha>0$ and $U\subset X$, we shall denote, for each $\varepsilon>0$, by $\Lambda_\alpha^\varepsilon(U)$ the lower bound of the sum $\sum_{n\ge 1}(\mathrm{diam}(S_n))^\alpha$ where $\{S_n\}_{n\ge 1}$ is an arbitrary covering of $U$ consisting of closed spheres of diameters less than $\varepsilon$. When $\varepsilon\to 0+$, $\Lambda_\alpha^\varepsilon(U)$ tends to a unique limit $\Lambda_\alpha(U)$ (finite or infinite), which we shall call the $\alpha$-*dimensional outer measure*. Then there exists a uniquely determined number such that

$$\sup\{\alpha;\ \Lambda_\alpha(U)=\infty\}=\inf\{\alpha;\ \Lambda_\alpha(U)=0\}\,,$$

which we shall call the *Hausdorff dimension* of $U$ and denote by $\dim_H(U)$.

The function of a set $\Lambda_\alpha(U)$ thus defined is an outer measure in the sense of Carathéodory. It is known that every Borel set is measurable and every set is regular with respect to this measure (Saks [46, p. 53]).

First of all, we have

THEOREM 10.2. *Suppose that* $\{f_j\}_{1\le j\le m}$ *is a finite set of weak contractions of* $X$. *Then* $\dim_H(K(f_1,\cdots,f_m))\le\lambda$ *where* $\lambda$ *is given by*

$$\sum_{j=1}^{m}(\mathrm{Lip}(f_j))^{\lambda}=1\,. \tag{10.1}$$

*Proof.* Fix an arbitrary $\kappa>\lambda$. Consider a closed sphere $S$ containing the set $K=K(f_1,\cdots,f_m)$. Put $f_w=f_{w_1}\circ\cdots\circ f_{w_n}$ for any word $w=(w_1\cdots w_n)$. Then we have

$$\Lambda_\kappa^\varepsilon(K)\le\sum_{w\in W(n)}\Lambda_\kappa^\varepsilon(f_w(K))\le\sum_{w\in W(n)}\Lambda_\kappa^\varepsilon(f_w(S))\le(2\,\mathrm{diam}(S))^\kappa\left(\sum_{j=1}^{m}(\mathrm{Lip}(f_j))^\kappa\right)^n,$$

where $\varepsilon=2\Omega^n(\mathrm{diam}(S))$ and $\Omega(t)=\max_{1\le j\le m}\Omega_{f_j}(t)$. Taking the limit as $n\to\infty$, it follows that $\Lambda_\kappa(K)=0$. This completes the proof. □

THEOREM 10.3. *Suppose that* $\{f_j\}_{1\le j\le m}$ *is a finite set of one to one weak contractions of* $X$. *Suppose further that* $\{f_j(K)\}_{1\le j\le m}$ *are pairwise disjoint where* $K=K(f_1,\cdots,f_m)$. *Then* $\dim_H(K)\ge\lambda$ *where* $\lambda$ *is given by*

$$\sum_{j=1}^{m}(\mathrm{Lip}(f_j^{-1}))^{-\lambda}=1\,. \tag{10.2}$$

*Proof.* Fix an arbitrary $\kappa<\lambda$. By assumption, we have $\mathrm{dist}(f_i(K),f_j(K))\ge\rho>0$ for any $i\ne j$. Consider now an arbitrary closed sphere $S$ satisfying $S\cdot K\ne\phi$.

Suppose first that $S\cdot K$ consists of more than one point. Then there exist an integer $n=n(S)\ge0$ and a word $w=w(S)\in W(n)$ such that $S\cdot K\le f_w(K)$ and $S\cdot f_{w\circ i}(K)\ne\phi\ne S\cdot f_{w\circ j}(K)$ for some $i\ne j$. Note that $\mathrm{diam}(S)\ge a_{w_1}\cdots a_{w_n}\rho$ where $a_j=(\mathrm{Lip}(f_j^{-1}))^{-1}$ for $1\le j\le m$.

Suppose next that $S\cdot K$ consists of exactly one point. Then we can take a sufficiently large integer $n=n(S)$ and a word $w=w(S)\in W(n)$ such that $S\cdot K=S\cdot f_w(K)$ and $\mathrm{diam}(S)\ge a_{w_1}\cdots a_{w_n}\rho$.

Thus, for any finite covering $\{S_j\}_{j\ge1}$ of $K$, we have $\sum_{w(S_j)}f_w(K)=K$. Therefore

$$\sum_{j\ge1}(\mathrm{diam}(S_j))^\kappa\ge\rho^\kappa\sum_{w(S_j)}(a_{w_1}\cdots a_{w_n})^\kappa\ge\rho^\kappa\sum_{w(S_j)}\mu(f_w(K))\ge\rho^\kappa\mu(K)=\rho^\kappa,$$

where $\mu$ is the probability measure such that $\mu(f_w(K))=(a_{w_1}\cdots a_{w_n})^\lambda$ for any $w$. Hence $\Lambda_\kappa(K)\ge\rho^\kappa$. This completes the proof. □

In the case $X=\boldsymbol{R}^p$ with the usual Euclidean distance, the following theorem is known. For the proof, see Falconer [13, pp. 118–124]. See also Moran [38], Marion [31], and Hutchinson [22].

THEOREM 10.4. *Let* $\{f_j\}_{1\le j\le m}$ *be a finite set of contractions of* $X=\boldsymbol{R}^p$ *satisfying* $\|f_j(x)-f_j(y)\|=\mathrm{Lip}(f_j)\|x-y\|$ *for any* $x,y\in X$. *Suppose that there exists a bounded open set* $V$ *such that* $\sum_{j=1}^{m}f_j(V)\subset V$ *and* $f_i(V)\cdot f_j(V)=\phi$ *for any* $i\ne j$. *Then*

$0<\Lambda_\lambda(K(f_1,\cdots,f_m))<\infty$; *therefore* $\dim_H(K)=\lambda$ *where* $\lambda$ *is given by*

$$\sum_{j=1}^{m}(\mathrm{Lip}(f_j))^\lambda=1\,. \tag{10.3}$$

EXAMPLE 1. Consider the contractions defined by (4.1). If $a+b<1$, it follows that $\dim_H(K(f_1,f_2))=\lambda$ where $a^\lambda+b^\lambda=1$ by Theorems 10.2 and 10.3. In particular, for Cantor's ternary set ($a=b=1/3$) we have $\dim_H(K)=\log 2/\log 3$. For the contractions defined by (5.5), one can easily verify that $\{f_1,f_2\}$ satisfies the condition of Theorem 10.4. Hence $\dim_H(K)=\lambda$ where $\lambda$ is given by $|\alpha|^\lambda+(1-|\alpha|^2)^\lambda=1$. Note that $\dim_H(K)$ is discontinuous at every real $\alpha$. On the other hand, the contractions defined by (5.6) does not presumably satisfy the condition of Theorem 10.4 for $\mathrm{Im}\,\alpha\neq 0$.

EXAMPLE 2. Let $X=[0,1]$ with the usual Euclidean distance and put

$$f_j(x)=\frac{1}{x+n_j} \qquad \text{for} \quad 1\leq j\leq m\,, \tag{10.4}$$

where $n_1,\cdots,n_m$ are $m$ distinct positive integers. Then $K(f_1,\cdots,f_m)$ is the set of all continued fractions each of whose partial quotients is either $n_1,\cdots,n_{m-1}$ or $n_m$, since

$$\psi(\alpha)=f_{\alpha_1}\circ f_{\alpha_2}\circ\cdots=\frac{1}{n_{\alpha_1}+}\,\frac{1}{n_{\alpha_2}+}\cdots \qquad \text{for any} \quad \alpha=(\alpha_n)\in\Sigma\,.$$

Using Theorems 10.2 and 10.3, one can easily obtain lower and upper estimates for $\dim_H(K)$. In this respect, see Good [14]. Moreover one can get better estimates using the following fact repeatedly:

$$K(\{f_j\}_{1\leq j\leq m})=K(\{f_{j_1}\circ f_{j_2}\}_{1\leq j_i\leq m})\,. \tag{10.5}$$

EXAMPLE 3. Let $X=\boldsymbol{C}$ with the usual Euclidean distance and put $R_a(z)=az(1-z)$ where $a$ is a real parameter satisfying $a>4$. It is known that the Julia set $J_a$ for $R_a(z)$ is totally disconnected and contained in $[0,1]$ (Brolin [5]). Then it is easily seen that if $a\geq 2+\sqrt{5}$, $J_a$ coincides with the set $K(f_1,f_2)$ where

$$f_1(x)=\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{x}{a}} \quad \text{and} \quad f_2(x)=\frac{1}{2}-\sqrt{\frac{1}{4}-\frac{x}{a}}\,. \tag{10.6}$$

From Theorems 10.2 and 10.3, it follows that if $a\geq 2+2\sqrt{2}$,

$$\frac{\log 2}{\log a}\leq \dim_H(K)\leq \frac{\log 4}{\log(a^2-4a)}\,.$$

Using (10.5), we also have the following asymptotic expansion:

$$\frac{\log 2}{\dim_H(K)}=\log a-\frac{1}{a}+O(a^{-2}) \qquad \text{as} \quad a\to\infty\,.$$

## 11. Williams' Problem

Throughout this section, we will restrict ourselves to the case $X=\boldsymbol{R}^p$ with the usual Euclidean distance. The following theorem has been shown by Williams. Compare with Theorem 4.1.

THEOREM 11.1 (Williams [54]). *Let $\{f_1, f_2\}$ be two one to one contractions of $\boldsymbol{R}$ such that* $\mathrm{Fix}(f_1)\neq\mathrm{Fix}(f_2)$ *and that*

$$\mathrm{Lip}(f_1^{-1})^{-1}+\mathrm{Lip}(f_2^{-1})^{-1}\geq 1\,. \tag{11.1}$$

*Then the set $K(f_1, f_2)$ is a closed line interval.*

Here we will give a simple proof for this, which is completely different from Williams' proof.

*Proof.* Let $L_0$ be the smallest closed interval containing the set $K=K(f_1, f_2)$. Then there exist $\alpha, \beta\in\Sigma$ such that $L_0=[\psi(\alpha), \psi(\beta)]$. Since $f_j\circ\psi(\alpha), f_j\circ\psi(\beta)\in K$, we have $L_0\geq f_j(L_0)$ for $j=1, 2$; therefore $L_0\geq L_1\geq L_2\geq\cdots$ where $F=f_1^*+f_2^*\in\mathscr{F}(\mathscr{C}(\boldsymbol{R}))$ and $L_n=F^n(L_0)$ for $n\geq 1$. Suppose now that $L_k$ is connected but $L_{k+1}$ is not for some $k\geq 0$. Since each $f_j(L_k)$ is a closed interval, it follows that $f_1(L_k)\cdot f_2(L_k)=\phi$; therefore

$$\begin{aligned}\mathrm{diam}(L_{k+1})&>\mathrm{diam}(f_1(L_k))+\mathrm{diam}(f_2(L_k))\\&\geq(\mathrm{Lip}(f_1^{-1})^{-1}+\mathrm{Lip}(f_2^{-1})^{-1})\mathrm{diam}(L_k)\geq\mathrm{diam}(L_k)\,,\end{aligned}$$

contrary to $L_{k+1}\leq L_k$. Therefore every $L_n$ is connected. Hence the set $\lim_{n\to\infty}L_n=K(f_1, f_2)$ is connected, as required. □

Note that the above theorem holds true even for weak contractions satisfying (11.1). We now give a generalization of Theorem 11.1 as follows:

THEOREM 11.2. *Let $\{f_j\}_{1\leq j\leq m}$ be a finite set of one to one weak contractions of $\boldsymbol{R}$ such that* $\mathrm{Fix}(f_i)\neq\mathrm{Fix}(f_j)$ *for some $i\neq j$ and that*

$$\sum_{j=1}^{m}(\mathrm{Lip}(f_j^{-1}))^{-1}\geq m-1\,. \tag{11.2}$$

*Then the set $K(f_1,\cdots,f_m)$ is a closed line interval.*

*Proof.* It suffices to show the connectedness of $K=K(f_1,\cdots,f_m)$ since $K$ is perfect by Theorem 4.3. Suppose, on the contrary, that $K$ is not connected. By Theorem 4.6, there exist two positive integers $r$ and $s$ such that $r+s=m$ and that

$$K_j\cdot K_{r+i}=\phi\qquad\text{for any}\quad 1\leq j\leq r\quad\text{and}\quad 1\leq i\leq s\,, \tag{11.3}$$

where $K_n=f_n(K)$ for $1\leq n\leq m$. Put $a_n=\mathrm{Lip}(f_n^{-1})^{-1}$ for $1\leq n\leq m$. Then we get $a_j+a_{r+i}<1$ for any $1\leq j\leq r$ and $1\leq i\leq s$. For otherwise, the set $K^*=K(f_j, f_{r+i})$ is connected by Theorem 11.1; therefore, by Theorem 4.6, $K_j\cdot K_{r+i}\geq f_j(K^*)\cdot f_{r+i}(K^*)\neq\phi$, contrary to (11.3). Thus we have

$$s\sum_{j=1}^{r} a_j + r\sum_{i=1}^{s} a_{r+i} < rs\,. \tag{11.4}$$

On the other hand,

$$s\sum_{j=1}^{r} a_j + r\sum_{i=1}^{s} a_{r+i} \geq \min(r, s)\sum_{j=1}^{m} a_j \geq \min(r, s)\cdot(m-1) \geq rs\,,$$

contrary to (11.4). This completes the proof. □

REMARK. The constant $m-1$ in (11.2) can not be replaced by any smaller number. For example, for an arbitrary $\varepsilon > 0$, consider the contractions

$$f_1(x) = \frac{\varepsilon}{m}x\,, \qquad f_j(x) = (1-\varepsilon)x + \frac{j}{m}\varepsilon \qquad \text{for} \quad 2 \leq j \leq m\,. \tag{11.5}$$

Then it is clear that $f_1([0, 1])\cdot f_j([0, 1]) = \phi$ for $2 \leq j \leq m$; therefore $K(f_1, \cdots, f_m)$ is not connected by Theorem 4.6, while

$$\sum_{j=1}^{m} (\mathrm{Lip}(f_j^{-1}))^{-1} > m - 1 - \varepsilon m\,.$$

In connection with Theorems 4.1 and 11.1, Williams gave the following problem: what is the structure of $K(f_1, f_2)$ for $f_1, f_2$: $\boldsymbol{R}^2 \to \boldsymbol{R}^2$, affine contractions satisfying (11.1)? Here we will give a partial answer for this. In fact, more generally we have

THEOREM 11.3. *Let $\{f_1, f_2\}$ be two one to one weak contractions of $\boldsymbol{R}^p$ such that* $\mathrm{Fix}(f_1) \neq \mathrm{Fix}(f_2)$ *and that*

$$\mathrm{Lip}(f_1^{-1})^{-p} + \mathrm{Lip}(f_2^{-1})^{-p} > 1\,. \tag{11.6}$$

*Then the set $K = K(f_1, f_2)$ is a nondegenerate locally connected continuum; therefore* $\dim_T(K) \geq 1$.

*Proof.* Suppose, on the contrary, that $K$ is not connected. Then, by Theorem 4.6, we have $f_1(K)\cdot f_2(K) = \phi$. Therefore it follows that $\dim_H(K) > p$ by Theorem 10.3. This contradiction completes the proof. □

As a corollary, we have immediately

COROLLARY 11.4. *Theorem* 11.3 *holds true for two one to one weak contractions $\{f_1, f_2\}$ satisfying* $\mathrm{Fix}(f_1) \neq \mathrm{Fix}(f_2)$ *and*

$$\mathrm{Lip}(f_1^{-1})^{-1} + \mathrm{Lip}(f_2^{-1})^{-1} > 2^{(p-1)/p}\,. \tag{11.7}$$

REMARK. For $p=2$, the constant $\sqrt{2}$ in (11.7) can not be replaced by any smaller number. For example, consider the contractions

$$f_1(z) = \left(s + \frac{i}{2}\right)\bar{z} \quad \text{and} \quad f_2(z) = \left(s - \frac{i}{2}\right)(\bar{z} - 1) + 1\,, \tag{11.8}$$

where $s$ is a real parameter satisfying $0<s<1/2$. We denote by $Q_s$ the closed quadrangle with vertices 0, 1, $1-s+i/2$ and $s+i/2$. Then it is easily seen that $f_1(Q_s)+f_2(Q_s)\leq Q_s$ and $f_1(Q_s)\cdot f_2(Q_s)=\phi$. Therefore the set $K(f_1,f_2)$ is totally disconnected by Theorem 4.4 (Fig. 9), while

$$\mathrm{Lip}(f_1^{-1})^{-1}+\mathrm{Lip}(f_2^{-1})^{-1}=\sqrt{1+4s^2}\,. \tag{11.9}$$

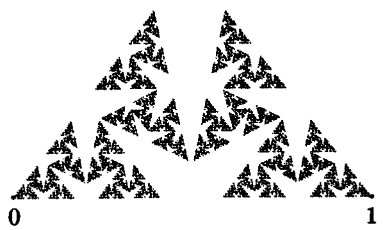



Fig. 9.

Finally, we will present the following problem: is it true or not that if one to one weak contractions $\{f_j\}_{1\leq j\leq m}$ of $R^p$ satisfy

$$\sum_{j=1}^{m}(\mathrm{Lip}(f_j^{-1}))^{-p}\geq 1\,,$$

then the set $K(f_1,\cdots,f_m)$ contains a nondegenerate component?

## References

[1] A. F. Beardon, The Geometry of Discrete Groups. Discrete Groups and Automorphic Functions. (ed. W. J. Harvey), Academic Press, New York, 1977, 47–72.
[2] A. S. Besicovitch and I. J. Schoenberg, On Jordan arcs and Lipschitz classes of functions. Acta Math., **106** (1961), 113–136.
[3] P. Billingsley, Ergodic Theory and Information. J. Wiley and Sons, New York, 1965.
[4] G. Birkhoff, Lattice Theory. 3rd Ed., Amer. Math. Soc., 1967.
[5] H. Brolin, Invariant sets under iteration of rational functions. Ark. Mat., **6** (1965), 103–144.
[6] T. Carleman, Sur les équations intégrales singulières à noyau réel et symétrique. Uppsala, 1923.
[7] E. Cesàro, Sur la représentation analytique des régions, et des courbes qui les remplissent. Bull. Sci. Math., **21** (1897), 257–266.
[8] A. Denjoy, Sur une fonction réele de Minkowski. J. Math. Pures Appl., **17** (1938), 105–151.
[9] F. M. Dekking, Recurrent sets. Adv. in Math., **44** (1982), 78–104.
[10] J. L. Denny, A continuous real-valued function on $E^n$ almost everywhere 1-1. Fund. Math., **55** (1964), 95–99.
[11] H. G. Eggleston, On closest packing by equilateral triangles. Proc. Camb. Philos. Soc., **49** (1953), 26–30.
[12] P. Erdös, On a family of symmetric Bernoulli convolutions. Amer. J. Math., **61** (1939), 974–976.
[13] K. J. Falconer, The Geometry of Fractal Sets. Cambridge, 1985.
[14] I. J. Good, The fractional dimensional theory of continued fractions. Proc. Camb. Philos. Soc., **37** (1941), 199–228.
[15] A. Granas, Introduction to Topology of Functional Spaces. Univ. Chicago, Math. Lecture Notes, 1961.
[16] M. Hata, Dynamics of Caianiello's equation. J. Math. Kyoto Univ., **22** (1982), 155–173.

[17] M. Hata, Scrambled sets on compact metric spaces. J. Math. Kyoto Univ., **24** (1984), 689–698.

[18] M. Hata, On the functional equation

$$\frac{1}{p}\left\{f\left(\frac{x}{p}\right)+\cdots+f\left(\frac{x+p-1}{p}\right)\right\}=\lambda f(\mu x).$$

J. Math. Kyoto Univ., **25** (1985), 357–364.

[19] M. Hata, On some properties of set-dynamical systems. Proc. Japan Acad., **61** (1985), Ser. A, 99–102.

[20] D. Hilbert, Über die stetige Abbildung einer Linie auf ein Flächenstuck. Math. Ann., **38** (1891), 459–460.

[21] W. Hurewicz and H. Wallman, Dimension Theory. Princeton, 1948.

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

[23] R. Kershner, On singular Fourier-Stieltjes transforms. Amer. J. Math., **58** (1936), 450–452.

[24] R. Kershner and A. Wintner, On symmetric Bernoulli convolutions. Amer. J. Math., **57** (1935), 541–548.

[25] H. von Koch, Sur une courbe continue sans tangente obtenue par une construction géométrique élémentaire. Ark. Mat. Astronom. Fys., **1** (1904), 681–702.

[26] K. Kuratowski, Topology. Vol. 1. Academic Press, New York, 1966.

[27] P. D. Lax, The differentiability of Pólya's function. Adv. in Math., **10** (1973), 456–464.

[28] P. Lévy, Les courbes planes ou gauches et les surfaces composées de parties semblables au tout. J. Ecole Poly., 1939, 227–292.

[29] Z. Lomnicki and S. Ulam, Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités I: Variables indépendantes. Fund. Math., **23** (1934), 237–278.

[30] B. B. Mandelbrot, The Fractal Geometry of Nature. W. H. Freeman, San Francisco, 1982.

[31] J. Marion, Le calcul de la mesure Hausdorff des sous-ensembles parfaits isotypiques de $\boldsymbol{R}^m$. C. R. Acad. Sci., Paris, **289** (1979), A65–68.

[32] P. Mattila, On the structure of self-similar fractals. Ann. Acad. Sci. Fenn., **7** (1982), 189–195.

[33] E. Michael, Topologies on spaces of subsets. Trans. Amer. Math. Soc., **71** (1951), 152–182.

[34] S. C. Milne, Peano curves and smoothness of functions. Adv. in Math., **35** (1980), 129–157.

[35] J. Milnor and W. Thurston, On iterated maps of the interval (I). Preprint, Princeton, 1977.

[36] H. Minkowski, Zur Geometrie der Zahlen. Gesammelte Abhandlungen, II, 50–51.

[37] E. H. Moore, On certain crinkly curves. Trans. Amer. Math. Soc., **1** (1900), 72–90.

[38] P. A. P. Moran, Additive functions of intervals and Hausdorff measure. Proc. Camb. Phil. Soc., **42** (1946), 15–23.

[39] W. F. Osgood, A Jordan curve of positive area. Trans. Amer. Math. Soc., **4** (1903), 107–112.

[40] G. Peano, Sur une courbe qui remplit toute une aire plane. Math. Ann., **36** (1890), 157–160.

[41] G. Pólya, Über eine Peanosche Kurve. Bull. Acad. Sci. Cracovie, A, 1913, 305–313.

[42] G. de Rham, Sur quelques courbes définies par des équations fonctionnelles. Rend. Sem. Mat. Torino, **16** (1957), 101–113.

[43] G. de Rham, Sur certaines équations fonctionnelles. Œuvres Mathématiques. l'Enseignement Math., 1981, 690–695.

[44] G. de Rham, Sur quelques fonctions différentiables dont toutes les valeurs sont des valeures critiques. Œuvres Mathématiques. l'Enseignement Math., 1981, 744–748.

[45] C. A. Rogers, Hausdorff Measures. Cambridge, 1970.

[46] S. Saks, Theory of the Integral. Warszawa, 1937.

[47] R. Salem, On some singular monotonic functions which are strictly increasing. Trans. Amer. Math. Soc., **53** (1943), 427–439.

[48] W. Sierpiński, Sur un système d'équations fonctionnelles, définissant une fonction avec un ensemble dense d'intervalles d'invariabilité. Œuvres Choisies, II. Warszawa, 1975, 44–48.

[49] W. Sierpiński, Sur une nouvelle courbe continue qui remplit toute une aire plane. Œuvres Choisies, II. Warszawa, 1975, 52–66.

[50] W. Sierpiński, Sur une courbe dont tout point est un point de ramification. Œuvres Choisies, II. Warszawa, 1975, 99–106.

[51] A. G. Vitushkin and G. M. Khenkin, Linear superpositions of functions. Uspehi Mat. Nauk, **22** (1967), 77–124.
[52] H. Whitney, A function not constant on a connected set of critical points. Duke Math. J., **1** (1935), 514–517.
[53] G. T. Whyburn, Analytic Topology. Amer. Math. Soc. Colloq. Pub., Vol. 28, 1942.
[54] R. F. Williams, Composition of contractions. Bol. Soc. Brasil. Mat., **2** (1971), 55–59.