Analysis on Fractals

Jun Kigami
Graduate School of Informatics
Kyoto University
Kyoto 606-8501, Japan
e-mail:kigami@i.kyoto-u.ac.jp

January 13, 2000

Contents

Chapter 1

Geometry of Self Similar Sets

In this chapter, we will review basics on geometry of self-similar sets which will be needed later. Namely, we will explain what is a self-similar set (in §1.1), how we can understand the structure of a self-similar set (in §1.2 and §1.3) and how to calculate the Hausdorff dimension of a self-similar set (in §1.5). The key notion is "self-similar structure" introduced in §1.3, which is a description of a self-similar set from a purely topological viewpoint. As we will explain in §1.3, topological structure of a self-similar set is essential in constructing analytical structure like Laplacians and Dirichlet forms. More precisely, if two self-similar sets are topologically same (i.e. homeomorphic), then analytical structure on one self-similar set can be transferred to another self-similar set through the homeomorphism.

§1.1 Construction of self similar sets

In this section, we will formulate self-similar sets on a metric space and show an existence and uniqueness theorem for self-similar sets. First we will introduce the notion of contractions on a metric space.

Notation. Let (X, d) be a metric space. For $x \in X$ and $r > 0$,

$$B_r(x) = \{y : y \in X, d(x, y) \leq r\}$$

Definition 1.1.1. Let (X, d_X) and (Y, d_Y) be metric spaces. A map $f : X \to Y$ is said to be (uniformly) Lipschitz continuous on X with respect to d_X, d_Y if

$$L = \sup_{x,y \in X, x \neq y} \frac{d_Y(f(x), f(y))}{d_X(x, y)} < \infty.$$

The above constant L is called the Lipschitz constant of f and denoted by $L = \mathrm{Lip}(f)$.

Obviously by the above definition, a Lipschitz continuous map is continuous.

Definition 1.1.2 (Contraction). Let (X, d) be a metric space. If $f : X \to X$ is Lipschitz continuous on X with respect to d, d and $\mathrm{Lip}(f) < 1$, then f is called a contraction with respect to the metric d with a contraction ratio $\mathrm{Lip}(f)$. In particular, a contraction f with a contraction ratio r is called a similitude if $d(f(x), f(y)) = rd(x, y)$ for all $x, y \in X$.

Remark. If f is a similitude on $(\mathbf{R}^n, d_E)$, then there exist $a \in \mathbf{R}^n$, $U \in O(n)$ and $r > 0$ such that $f(x) = rUx + a$ for all $x \in \mathbf{R}^n$. (Exercise 1.1)

The following theorem is called the "contractive mapping theorem".

Theorem 1.1.3. *Let (X, d) be a complete metric space and let $f : X \to X$ be a contraction with respect to the metric d. Then there exists a unique fixed point of f, in other words, there exists a unique solution of an equation $f(x) = x$. Moreover if x_* is the fixed point of f, then $\{f^n(a)\}_{n\geq 0}$ converges to x_* for all $a \in X$ where f^n is the n-th iteration of f.*

Proof. If r is the ratio of contraction of f, then for $m > n$,

$$\begin{aligned} d(f^n(a), f^m(a)) &\leq d(f^n(a), f^{n+1}(a)) + \cdots + d(f^{m-1}(a), f^m(a)) \\ &\leq (r^n + \cdots + r^{m-1})d(a, f(a)) \leq \frac{r^n}{(1-r)} d(a, f(a)) \end{aligned}$$

Hence $\{f^n(a)\}_{n\geq 0}$ is a Cauchy sequence. As (X, d) is complete, there exists $x_* \in X$ such that $f^n(a) \to x_*$ as $n \to \infty$. Using the fact that $f^{(n+1)}(a) = f(f^n(a))$, we can easily deduce that $x_* = f(x_*)$.

Now, if $f(x) = x$ and $f(y) = y$, then $d(x, y) = d(f(x), f(y)) \leq rd(x, y)$. Therefore $d(x, y) = 0$ and $x = y$. So we have uniqueness of fixed points. □

Remark. In general, for a self mapping f from an set to itself, a solution of $f(x) = x$ is called a fixed point or an equilibrium point of f.

We now state the main theorem of this section, which ensures uniqueness and existence of self-similar sets.

Theorem 1.1.4. *Let (X, d) be a complete metric space. If $f_i : X \to X$ is a contraction with respect to the metric d for $i = 1, 2, \cdots, N$,then there exists a unique non-empty compact subset K of X that satisfies*

$$K = f_1(K) \cup \cdots \cup f_N(K).$$

K is called the self-similar set with respect to $\{f_1, f_2, \cdots, f_N\}$.

Remark. In other literature, the word "self-similar set" is used in a more restricted sense. For example, Huchinson [57] uses the word "self-similar set" only if all the contractions are similitudes. Also, in case all the contractions are affine function on $\mathbf{R}^n$, the associated set may be called a self-affine set.

The contractive mapping theorem is a special case of Theorem 1.1.4 where $N = 1$.

In the rest of this section, we will give a proof to Theorem 1.1.4. Define

$$F(A) = \bigcup_{1 \le j \le N} f_j(A)$$

for $A \subseteq X$. The main idea is to show existence of a fixed point of F. In order to do so, first we choose a good domain for F, which is defined by

$$\mathcal{C}(X) = \{A : A \text{ is a non-empty compact subset of } X\}.$$

Obviously F is a mapping from $\mathcal{C}(X)$ to itself. Next we define a metric δ on $\mathcal{C}(X)$, which is called the Hausdorff metric on $\mathcal{C}(X)$.

Proposition 1.1.5. *For $A, B \in \mathcal{C}(X)$, define*

$$\delta(A, B) = \inf\{r > 0 : U_r(A) \supseteq B \text{ and } U_r(B) \supseteq A\},$$

where $U_r(A) = \{x \in X : d(x, y) \le r \text{ for some } y \in A\} = \cup_{y \in A} B_r(y)$. Then δ is a metric on $\mathcal{C}(X)$. Moreover if (X, d) is complete, then $(\mathcal{C}(X), \delta)$ is also complete.

Before giving a proof of the above proposition, we recall some standard definitions in general topology.

Definition 1.1.6. Let (X, d) be a metric space and let K be a subset of X.

(1) A finite set $A \subset K$ is called an r-net of K for $r > 0$ if and only if $\cup_{x \in A} B_r(x) \supseteq K$.

(2) K is said to be totally bounded if and only if there exists an r-net of K for any $r > 0$

It is well-known that a metric space is compact if and only if it is complete and totally bounded.

Proof of Proposition 1.1.5. Obviously $\delta(A, B) = \delta(B, A) \ge 0$ and $\delta(A, A) = 0$.

$\delta(A, B) = 0 \Rightarrow A = B$: For any n, $U_{1/n}(B) \supseteq A$. Therefore for any $x \in A$, we can choose $x_n \in B$ such that $d(x, x_n) \le 1/n$. As B is closed, $x \in B$. Hence we have $A \subseteq B$. One can obtain $B \subseteq A$ in exactly the same way.

Triangle inequality : If $r > \delta(A, B)$ and $s > \delta(B, C)$, then $U_{r+s}(A) \supseteq C$ and $U_{r+s}(C) \supseteq A$. Hence $r + s \ge \delta(A, C)$. This implies $\delta(A, B) + \delta(B, C) \ge \delta(A, C)$.

Next we proof that $(\mathcal{C}(X), \delta)$ is complete if (X, d) is complete. For a Cauchy sequence $\{A_n\}_{n \ge 1}$ in $(\mathcal{C}(X), \delta)$, define $B_n = \overline{\cup_{k \ge n} A_k}$. First we will show that B_n is compact. As B_n is a monotonically decreasing sequence of closed sets, it is enough to show that B_1 is compact. For any $r > 0$, we can choose m so that $U_{r/2}(A_m) \supseteq A_k$ for all $k \ge m$. As A_m is compact, there exists a $r/2$-net P of A_m. We can immediately verify that $\cup_{x \in P} B_r(x) \supseteq U_{r/2}(A_m) \supseteq \cup_{k \ge m} A_k$. As $\cup_{x \in P} B_r(x)$ is closed, it is easy to see that P is a r-net of B_m. Adding r-nets of $A_1, A_2, \cdots, A_{m-1}$ to P, we can obtain an r-net of B_1. Hence B_1 is totally bounded. Also, B_1 is complete because it is a closed subset of the complete metric space X. Thus it follows that B_n is compact.

Now as $\{B_n\}$ is a monotonically decreasing sequence of non-empty compact sets, $A = \cap_{n\geq 1} B_n$ is compact and non-empty. For any $r > 0$, we can choose m so that $U_r(A_m) \supseteq A_k$ for all $k \geq m$. Then $U_r(A_m) \supseteq B_m \supseteq A$. On the other hand, we see that $U_r(A) \supseteq B_m \supseteq A_m$ for sufficiently large m. Thus we have $\delta(A, A_m) \leq r$ for sufficiently large m. Hence $A_m \to A$ as $m \to \infty$ in the Hausdorff metric. So we can see that $(\mathcal{C}(X), \delta)$ is complete. □

Now, Theorem 1.1.4 can be stated in the following way using the Hausdorff metric $(\mathcal{C}(X), \delta)$.

Theorem 1.1.7. *Let (X, d) be a complete metric space. Define $F : \mathcal{C}(X) \to \mathcal{C}(X)$ by $F(A) = \cup_{1\leq i\leq N} f_i(A)$, where $f_i : X \to X$ is a contraction for every $i = 1, 2, \cdots, N$. Then F has a unique fixed point K. Moreover, for any $A \in \mathcal{C}(X)$, $F^n(A)$ converges to K as $n \to \infty$ in the sense of the Hausdorff metric.*

Lemma 1.1.8. *For $A_1, A_2, B_1, B_2 \in \mathcal{C}(X)$,*

$$\delta(A_1 \cup A_2, B_1 \cup B_2) \leq \max\{\delta(A_1, B_1), \delta(A_2, B_2)\}$$

Proof. If $r > \max\{\delta(A_1, B_1), \delta(A_2, B_2)\}$, then $U_r(A_1) \supseteq B_1$ and $U_r(A_2) \supseteq B_2$. Hence $U_r(A_1 \cup A_2) \supseteq B_1 \cup B_2$. A similar argument implies $U_r(B_1 \cup B_2) \supseteq A_1 \cup A_2$. Hence $r \geq \delta(A_1 \cup A_2, B_1 \cup B_2)$. This completes the proof. □

Lemma 1.1.9. *If f is a contraction with a contraction ratio r, then $\delta(f(A), f(B)) \leq r\delta(A, B)$ for any $A, B \in \mathcal{C}(X)$.*

Proof. If $U_s(A) \supseteq B$ and $U_s(B) \supseteq A$, $U_{sr}(f(A)) \supseteq f(U_s(A)) \supseteq f(B)$. Also the same discussion implies $U_{sr}(f(B)) \supseteq f(A)$. Therefore, $\delta(f(A), f(B)) \leq rs$ and this completes the proof. □

Proof of Theorem 1.1.7. Using Lemma 1.1.8 repeatedly, we obtain

$$\delta(F(A), F(B)) = \delta(\cup_{1\leq j\leq N} f_j(A), \cup_{1\leq j\leq N} f_j(B)) \leq \max_{1\leq j\leq N} \delta(f_j(A), f_j(B)).$$

By Lemma 1.1.9, $\delta(f_i(A), f_i(B)) \leq r_i \delta(A, B)$, where r_i is the contraction ratio of f_i. Define $r = \max_{1\leq i\leq N} r_i$, it follows that $\delta(F(A), F(B)) \leq r\delta(A, B)$. Therefore F turns out to be a contraction with respect to the Hausdorff metric. By Proposition 1.1.5, we see that $(\mathcal{C}(X), \delta)$ is complete. Now the contractive mapping theorem (Theorem 1.1.3) implies Theorem 1.1.7 immediately. □

§1.2 Shift space and self similar sets

In this section, we will introduce the shift space, which is the key to understand topological structures of self-similar sets. In fact, Theorem 1.2.3 will show that every self-similar set is a quotient space of a shift space.

Definition 1.2.1. Let N be a natural number.

(1) For $m \geq 1$, we define

$$W_m^N = \{1, 2, \cdots, N\}^m = \{w_1 w_2 \cdots w_m : w_i \in \{1, 2, \cdots, N\}\}.$$

$w = w_1 w_2 \cdots w_m \in W_m^N$ is called a word of length m with symbols $\{1, 2, \cdots, N\}$. Also, for $m = 0$, we define $W_0^N = \{\emptyset\}$ and call $\emptyset$ the empty word. Moreover, set $W_*^N = \cup_{m \geq 0} W_m^N$ and denote the length of $w \in W_*^N$ by $|w|$.

(2) The collection of one-sided infinite sequence of symbols $\{1, 2, \cdots, N\}$ is denoted by Σ^N, which is called the shift space with N-symbols. More precisely,

$$\Sigma^N = \{1, 2, \cdots, N\}^{\mathbb{N}} = \{\omega_1 \omega_2 \omega_3 \cdots : \omega_i \in \{1, \cdots, N\} \text{ for } i \in \mathbf{N}\}.$$

For $k \in \{1, 2, \cdots, N\}$, define a map $\sigma_k : \Sigma^N \to \Sigma^N$ by $\sigma_k(\omega_1 \omega_2 \omega_3 \cdots) = k\omega_1 \omega_2 \omega_3 \cdots$. Also define $\sigma : \Sigma^N \to \Sigma^N$ by $\sigma(\omega_1 \omega_2 \omega_3 \cdots) = \omega_2 \omega_3 \omega_4 \cdots$. σ is called the shift map.

Remark. The two sided infinite sequence of $\{1, 2, \cdots, N\}$,

$$\{1, 2, \cdots, N\}^{\mathbb{Z}} = \{\cdots \omega_{-2} \omega_{-1} \omega_0 \omega_1 \omega_2 \cdots : \omega_i \in \{1, 2, \cdots, N\} \text{ for } i \in \mathbf{Z}\}$$

may also called the shift space with N-symbols. If one want to distinguish those two, the above Σ^N should be called the one-sided shift space with N-symbols. In this book, however, we will not treat the two-sided symbol space.

For ease of notation, we write W_m, W_* and Σ instead of W_m^N, W_*^N and Σ^N

It is obvious that σ_k is a branch of the inverse of σ for any $k \in \{1, 2, \cdots, N\}$. If we take an adequate distance, it turns out that σ_k is a contraction and the shift space Σ is the self-similar set with respect to $\{\sigma_1, \sigma_2, \cdots, \sigma_N\}$.

Theorem 1.2.2. *For $\omega, \tau \in \Sigma$ with $\omega \neq \tau$ and $0 < r < 1$, define $\delta_r(\omega, \tau) = r^{s(\omega,\tau)}$, where $s(\omega, \tau) = \min\{m : \omega_m \neq \tau_m\} - 1$. (i.e. $n = s(\omega, \tau)$ if and only if $\omega_i = \tau_i$ for $1 \leq i \leq n$ and $\omega_{n+1} \neq \tau_{n+1}$.) Also define $\delta_r(\omega, \tau) = 0$ if $\omega = \tau$. δ_r is a metric on Σ and (Σ, δ_r) is a compact metric space. Furthermore, σ_k is a similitude with $\mathrm{Lip}(\sigma_k) = r$ and Σ is the self-similar set with respect to $\{\sigma_1, \sigma_2, \cdots, \sigma_N\}$.*

Proof. It is obvious by the definition that $\delta_r(\omega, \tau) \geq 0$ and $\delta_r(\omega, \tau) = 0$ implies $\omega = \tau$. As $\min\{s(\omega, \tau), s(\tau, \kappa)\} \leq s(\omega, \kappa)$ for $\omega, \tau, \kappa \in \Sigma$, we can see that $\delta_r(\omega, \tau) + \delta_r(\tau, \kappa) \geq \delta_r(\omega, \kappa)$.

Now for every $w = w_1 w_2 \cdots w_m \in W_*$, we define

$$\Sigma_w = \{\omega = \omega_1 \omega_2 \omega_3 \cdots \in \Sigma : \omega_1 \omega_2 \cdots \omega_m = w_1 w_2 \cdots w_m\}.$$

Let $\{\omega^n\}_{n \geq 1}$ be a sequence in Σ. By using an induction on m, we can choose $\tau \in \Sigma$ so that $\{n \geq 1 : (\omega^n)_j = \tau_j \text{ for } j = 1, 2, \cdots, m\}$ becomes a infinite set for any $m \geq 1$. So there exists a subsequence of $\{\omega^n\}$ that converges to τ as $n \to \infty$. Hence (Σ, δ_r) is compact.

Finally it is obvious that σ_k is a similitude with $\mathrm{Lip}(\sigma_k) = r$. Also we can easily see that $\Sigma = \sigma_1(\Sigma) \cup \cdots \cup \sigma_N(\Sigma)$. This implies that Σ is the self-similar set with respect to $\{\sigma_1, \sigma_2, \cdots, \sigma_N\}$. ∎

Σ is called the (topological) Cantor set with N-symbols. See Example 1.2.6.

For the rest of this section, we assume that (X, d) is a complete metric space, $f_i : X \to X$ is a contraction with respect to (X, d) for every $i \in \{1, 2, \cdots, N\}$ and that K is the self-similar set with respect to $\{f_1, f_2, \cdots, f_N\}$.

The following theorem shows that every self-similar set is a quotient space of a shift space by a certain equivalence relation.

Theorem 1.2.3. *For $w = w_1 w_2 \cdots w_m \in W_*$, set $f_w = f_{w_1} \circ f_{w_2} \circ \cdots \circ f_{w_m}$ and $K_w = f_w(K)$. Then for any $\omega = \omega_1\omega_2\omega_3 \cdots \in \Sigma$, $\cap_{m\geq 1} K_{\omega_1\omega_2\cdots\omega_m}$ contains only one point. If we define $\pi : \Sigma \to K$ by $\{\pi(\omega)\} = \cap_{m\geq 1} K_{\omega_1\omega_2\cdots\omega_m}$, then π is a continuous surjective map. Moreover, for any $i \in \{1, 2, \cdots, N\}$, $\pi \circ \sigma_i = f_i \circ \pi$.*

Proof. Note that

$$K_{\omega_1\omega_2\cdots\omega_m\omega_{m+1}} = f_{\omega_1\omega_2\cdots\omega_m}(f_{\omega_{m+1}}(K)) \subseteq f_{\omega_1\omega_2\cdots\omega_m}(K) = K_{\omega_1\omega_2\cdots\omega_m}.$$

As $K_{\omega_1\omega_2\cdots\omega_m}$ is compact, $\cap_{m\geq 1} K_{\omega_1\omega_2\cdots\omega_m}$ is a non-empty compact set. For $A \subseteq X$, the diameter of A, $\mathrm{diam}(A)$, is defined by $\mathrm{diam}(A) = \sup_{x,y\in A} d(x, y)$. Set $R = \max_{1\leq i\leq N} \mathrm{Lip}(f_i)$. Then it follows that $\mathrm{diam}(f_i(A)) \leq R\mathrm{diam}(A)$. Hence $\mathrm{diam}(K_{\omega_1\omega_2\cdots\omega_m}) \leq R^m \mathrm{diam}(K)$. So $\mathrm{diam}(\cap_{m\geq 1} K_{\omega_1\omega_2\cdots\omega_m}) = 0$. Therefore $\cap_{m\geq 1} K_{\omega_1\omega_2\cdots\omega_m}$ should contain only one point.

If $\delta_r(\omega, \tau) \leq r^m$, then $\pi(\omega), \pi(\tau) \in K_{\omega_1\omega_2\cdots\omega_m} = K_{\tau_1\tau_2\cdots\tau_m}$. Therefore $d(\pi(\omega), \pi(\tau)) \leq R^m \mathrm{diam}(K)$. This immediately implies that π is continuous.

By using

$$\{\pi(\sigma_i(\omega))\} = \cap_{m\geq 1} K_{i\omega_1\omega_2\cdots\omega_m} = \cap_{m\geq 1} f_i(K_{\omega_1\omega_2\cdots\omega_m}) = \{f_i(\pi(\omega))\},$$

we can easily verify that $\pi \circ \sigma_i = f_i \circ \pi$.

Finally we would show that π is surjective. Note that $\pi(\Sigma) = \pi(\sigma_1(\Sigma) \cup \cdots \cup \sigma_N(\Sigma)) = \pi(\sigma_1(\Sigma)) \cup \cdots \cup \pi(\sigma_N(\Sigma)) = f_1(\pi(\Sigma)) \cup \cdots \cup f_N(\pi(\Sigma))$. As $\pi(\Sigma)$ is a non-empty compact set, uniqueness of self-similar sets (Theorem 1.1.4) implies that $\pi(\Sigma) = K$. □

Proposition 1.2.4. *Define $\dot{w} = www\cdots$ if $w \in W_*$ and $w \neq \emptyset$. Then $\pi(\dot{w})$ is the unique fixed point of f_w.*

Proof. As f_w is a contraction, it has a unique fixed point. By Theorem 1.2.3, $\pi(\dot{w}) = \pi(w \cdot \dot{w}) = f_w(\pi(\dot{w}))$. Hence $\pi(\dot{w})$ is the fixed point of f_w. □

By using the above proposition, we can see that $\pi(v_1 v_2 \cdots v_k \dot{w}) = f_v(p_w)$ where $w \in W_*$, $w \neq \emptyset$, $v = v_1 v_2 \cdots v_k \in W_*$ and p_w is the fixed point of f_w. This relation helps us to understand π in many examples. Moreover, since periodic sequences are dense in Σ, we have

$$K = \overline{\{p_w : w \in W_*, w \neq \emptyset\}}.$$

In fact, π determines a topological structure on a self-similar sets.

Proposition 1.2.5. *Suppose f_i is injective for every $i \in \{1, 2, \cdots, N\}$. Then, $\pi(\omega) = \pi(\tau)$ for $\omega \neq \tau \in \Sigma$ if and only if $\pi(\sigma^m\omega) = \pi(\sigma^m\tau)$, where $m = s(\omega, \tau)$.*

Proof. If $w = \omega_1\omega_2\cdots\omega_m = \tau_1\tau_2\cdots\tau_m$, then $\pi(\omega) = f_w(\pi(\sigma^m\omega))$ and $\pi(\tau) = f_w(\pi(\sigma^m\tau))$. As f_w is injective, we have $\pi(\omega) = \pi(\tau)$ for $\omega, \tau \in \Sigma$ if and only if $\pi(\sigma^m\omega) = \pi(\sigma^m\tau)$. The other direction is obvious. □

Note that if $\pi(\omega) = \pi(\tau)$, then $\pi(\sigma^m(\omega)) = \pi(\sigma^m(\tau)) \in C_K$, where $m = s(\omega,\tau)$ and $C_K = \cup_{1\leq i<j\leq N}(K_i \cap K_j)$.

Example 1.2.6 (Cantor set). Let $X = [0,1]$. Choose positive real numbers a and b so that $a + b < 1$. Define $f_1(x) = ax$ and $f_2(x) = b(x-1)+1$. If K is the self-similar set with respect to $\{f_1, f_2\}$, $K_1 \subset [0,a]$ and $K_2 \subset [1-b,1]$. Hence $C_K = K_1 \cap K_2 = \emptyset$. Therefore $\pi : \Sigma \to K$ is injective. By Theorem 1.2.3, π is also surjective and hence it is a homeomorphism between Σ and K. In particular, if $a = b = 1/3$, K is called the Cantor's ternary set or the Cantor's middle third set.

Example 1.2.7 (Koch curve). Let $X = \mathbf{C}$. Suppose that $a \in \{z : |z|^2 + |1-z|^2 < 1\}$. Set $f_1(z) = \alpha\bar{z}$ and $f_2(z) = (1-\alpha)(\bar{z}-1)+1$. Let D be a triangle domain with vertices $\{0,\alpha,1\}$, including the boundary. Then it follows that $f_1(D) \cup f_2(D) \subseteq D$ and $f_1(D) \cap f_2(D) = \{\alpha\}$. Hence $K(\alpha) \subseteq D$, where $K(\alpha)$ is the self-similar set with respect to $\{f_1, f_2\}$. Also note that $f_1(0) = 0$, $f_2(1) = 1$ and $f_2(0) = f_1(1) = \alpha$. These facts imply that $\pi_\alpha(\dot{1}) = 0, \pi_\alpha(\dot{2}) = 1$ and $\pi_\alpha(1\dot{2}) = \pi_\alpha(2\dot{1}) = \alpha$. Moreover, $C_K = K_1 \cap K_2 = \{\alpha\}$. Hence we can deduce that $\pi_\alpha(\omega) = \pi_\alpha(\tau)$ and $\omega \neq \tau$ if and only if there exists $w \in W_*$ such that $\{\omega,\tau\} = \{w1\dot{2}, w2\dot{1}\}$. In particular, $K(1/2) = [0,1]$ and $K(\alpha)$ is called the Koch curve if $\alpha = \frac{1}{2} + \frac{i}{2\sqrt{3}}$. Note that $\pi_\alpha \circ {\pi_{1/2}}^{-1}$ is a homeomorphism between $[0,1]$ and $K(\alpha)$.

Example 1.2.8 (Sierpinski gasket). Let $X = \mathbf{C}$ and let $\{p_1, p_2, p_3\}$ be a set of vertices of a regular triangle. Define $f_j(z) = 1/2(z - p_j) + p_j$ for $j = 1,2,3$. The self-similar set with respect to $\{f_1, f_2, f_3\}$ is called the Sierpinski gasket. It is easy to see that $\pi(\dot{j}) = p_j$ for $j = 1,2,3$. Let T be the regular triangle with vertices $\{p_1, p_2, p_3\}$, including the boundary. Then $f_1(T) \cup f_2(T) \cup f_3(T) \subseteq T$. Hence $K \subset T$. Also $f_1(K) \cap f_2(K) = f_1(T) \cap f_2(T)$ and this set contains only one point, which is denoted by $\{q_3\}$. Then $\pi^{-1}(q_3) = \{2\dot{1}, 1\dot{2}\}$. In the same way, if $f_2(K) \cap f_3(K) = \{q_1\}$ and $f_3(K) \cap f_1(K) = \{q_2\}$ then $\pi^{-1}(q_1) = \{2\dot{3}, 3\dot{2}\}$ and $\pi^{-1}(q_2) = \{3\dot{1}, 1\dot{3}\}$. By those facts, if $\pi(\omega) = \pi(\tau)$ and $\omega \neq \tau$, there exists $w \in W_*$ such that $\{\omega,\tau\} = \{w1\dot{2}, w2\dot{1}\}$ or $\{w2\dot{3}, w3\dot{2}\}$ or $\{w3\dot{1}, w1\dot{3}\}$.

Example 1.2.9 (Hata's tree-like set). Let $X = \mathbf{C}$. Set $f_1(z) = c\bar{z}$, $f_2(z) = (1-|c|^2)\bar{z} + |c|^2$, where $|c|, |1-c| \in (0,1)$. The self-similar set with respect to $\{f_1, f_2\}$ is called Hata's tree-like set. Let $A = \{t : 0 \leq t \leq 1\} \cup \{tc, 0 \leq t \leq 1\}$. Then it follows that $f_1(A) \cup f_2(A) \supset A$. Hence if $A_m = \cup_{w\in W_m} f_w(A)$, then $\{A_m\}_{m\geq 0}$ is an increasing sequence and $K = \overline{\cup_{m\geq 0} A_m}$. Also we can easily observe that $f_1(K) \cap f_2(K) = \{|c|^2\}$, $f_1(0) = 0, f_2(1) = 1$ and $f_1(f_1(1)) = f_2(0) = |c|^2$. Hence $\pi^{-1}(0) = \{\dot{1}\}$, $\pi^{-1}(1) = \{\dot{2}\}$, $\pi^{-1}(\bar{c}) = \{1\dot{2}\}$ and $\pi^{-1}(|c|^2) = \{11\dot{2}, 2\dot{1}\}$. Moreover, if $\pi(\omega) = \pi(\tau)$ and $\omega \neq \tau$, there exists $w \in W_*$ such that $\{\omega,\tau\} = \{w11\dot{2}, w2\dot{1}\}$.

§1.3 Self similar structure

From the viewpoint of analysis, only the topological structure of a self-similar set is important. For example, suppose you want to know what is analysis on the Koch curve. Recall Example 1.2.7, there exists a natural homeomorphism between $[0, 1]$ and the Koch curve. Through this homeomorphism, any kind of analytical structure on $[0, 1]$ can be translated to its counterpart on the Koch curve. So it is easy to construct analysis on the Koch curve.

The notion of self-similar structure has been introduced to give a topological description of self-similar sets.

Definition 1.3.1. Let K be a compact metrizable topological space and let S be a finite set. Also, let F_i be a continuous injection from K to itself for any $i \in S$. Then, $(K, S, \{F_i\}_{i \in S})$ is called a self-similar structure if there exists a continuous surjection $\pi : \Sigma \to K$ such that $F_i \circ \pi = \pi \circ \sigma_i$ for every $i \in S$, where $\Sigma = S^{\mathbb{N}}$ is the one-sided shift space and $\sigma_i : \Sigma \to \Sigma$ is defined by $\sigma_i(w_1 w_2 w_3 \cdots) = i w_1 w_2 w_3 \cdots$ for each $w_1 w_2 w_3 \cdots \in \Sigma$.

Σ is called the shift space with symbols S. We will define $W_m = S^m$, $W_* = \cup_{m \geq 0} W_m$, $\sigma : \Sigma \to \Sigma$ and so on in exactly the same way as in §1.2. Also the topology of Σ is given by exactly the same way as in §1.2. If we need to specify the symbols S, we use $\Sigma(S)$, $W_m(S)$ and $W_*(S)$ in place of Σ, W_m and W_* respectively. In many cases, we think of $S = \{1, 2, \cdots, N\}$.

Obviously if K is the self-similar set with respect to injective contractions $\{f_1, f_2, \cdots, f_N\}$, then $(K, \{1, 2, \cdots, N\}, \{f_i\})$ is a self-similar structure. It is possible that two different self-similar sets have same topological structure. For example, the self-similar structures corresponding to the self-similar sets $K(\alpha)$ in Example 1.2.7 are all essentially same. More precisely, they are isomorphic in the following sense.

Definition 1.3.2. Let $\mathcal{L}_1 = (K_1, S_1, \{F_i\}_{i \in S_1})$ and $\mathcal{L}_2 = (K_2, S_2, \{G_i\}_{i \in S_2})$ be self-similar structures. Also let $\pi_i : \Sigma(S_i) \to K_i$ be the continuous surjection associated with $\mathcal{L}_i$ for $i = 1, 2$. We say that $\mathcal{L}_1$ and $\mathcal{L}_2$ are isomorphic if there exists a bijective map $\rho : S_1 \to S_2$ such that $\pi_2 \circ \iota_\rho \circ {\pi_1}^{-1}$ becomes a well-defined homeomorphism between K_2 and K_1, where ι_ρ is the natural bijective map induced by γ, i.e. $\iota(\omega_1 \omega_2 \cdots) = \rho(\omega_1)\rho(\omega_2) \cdots$.

We say that two self-similar structures are same if they are isomorphic.

Proposition 1.3.3. *If $(K, S, \{F_i\}_{i \in S})$ is a self-similar structure, then π is unique. In fact,*

$$\{\pi(\omega)\} = \bigcap_{m \geq 0} F_{\omega_1 \omega_2 \cdots \omega_m}(K)$$

for any $\omega = \omega_1 \omega_2 \cdots \in \Sigma$.

Proof. By the above definition, we have $F_{w_1 w_2 \cdots w_m} \circ \pi = \pi \circ \sigma_{w_1 w_2 \cdots w_m}$ for any $w \in W_*$. Hence, $\pi(\omega) \in \cap_{m \geq 0} F_{\omega_1 \omega_2 \cdots \omega_m}(K)$. For $x \in \cap_{m \geq 0} F_{\omega_1 \omega_2 \cdots \omega_m}(K)$, there exists $x_m \in \Sigma_{\omega_1\omega_2\cdots\omega_m}$ such that $\pi(x_m) = x$. Note that π is continuous. Since $x_m \to \omega$ as $m \to \infty$, it follows that $x = \pi(x_m) \to \pi(\omega)$ as $m \to \infty$. Hence $x = \pi(\omega)$. □

Definition 1.3.4. Let $\mathcal{L} = (K, S, \{F_i\}_{i\in S})$ be a self-similar structure. We define $C_{\mathcal{L}} = \cup_{i,j\in S, i\neq j}(F_i(K) \cap F_j(K))$, $\mathcal{C}_{\mathcal{L}} = \pi^{-1}(C_{\mathcal{L}})$ and $\mathcal{P}_{\mathcal{L}} = \cup_{n\geq 1}\sigma^n(\mathcal{C}_{\mathcal{L}})$. $\mathcal{C}_{\mathcal{L}}$ is called the critical set of $\mathcal{L}$ and $\mathcal{P}_{\mathcal{L}}$ is called the post critical set of $\mathcal{L}$. Also we define $V_0(\mathcal{L}) = \pi(\mathcal{P}_{\mathcal{L}})$.

For ease of notations, we use $\mathcal{C}$, $\mathcal{P}$ and V_0 instead of $\mathcal{C}_{\mathcal{L}}$, $\mathcal{P}_{\mathcal{L}}$ and $V_0(\mathcal{L})$ as far as it may not cause any confusion.

The critical set and the post critical set play an important role in determining the topological structure of a self-similar set. For example, if $\mathcal{C} = \emptyset$, (and hence $\mathcal{P}$, V_0 are all empty sets), then K is homeomorphic to the (topological) Cantor set Σ.

Also V_0 is thought as a "boundary" of K. For example, define $F_1(x) = \frac{1}{2}x$ and $F_2(x) = \frac{1}{2}x + \frac{1}{2}$ and recall Example 1.2.7. Then we find that $\mathcal{C} = \{1\dot{2}, 2\dot{1}\}$ and $\mathcal{P} = \{\dot{1}, \dot{2}\}$. Hence $V_0 = \{0, 1\}$. See also Exercise 1.3 for another example.

Proposition 1.3.5. *Let $\mathcal{L} = (K, S, \{F_i\}_{i\in S})$ be a self-similar structure. Then*

(1) $\pi^{-1}(V_0) = \mathcal{P}$.

(2) If $\Sigma_w \cap \Sigma_v = \emptyset$ for $w, v \in W_*$, then $K_w \cap K_v = F_w(V_0) \cap F_v(V_0)$, where $K_w = F_w(K)$.

(3) $\mathcal{C} = \emptyset$ if and only if π is injective.

Proof. (1) If $\pi(\omega) \in V_0$, then there exist $\tau \in \mathcal{C}$ and $m \geq 1$ such that $\sigma^m\tau = \omega$. Set $\omega' = \tau_1\tau_2\cdots\tau_m \cdot \omega$, then $\pi(\omega') = F_{\tau_1\tau_2\cdots\tau_m}(\pi(\omega)) = F_{\tau_1\tau_2\cdots\tau_m}(\pi(\sigma^m\tau)) = \pi(\tau) \in C_{\mathcal{L}}$. Hence $\omega' \in \mathcal{C}$ and $\omega \in \mathcal{P}$.

(2) It is obvious that $F_w(V_0) \cap F_v(V_0) \subseteq K_w \cap K_v$. For $x \in K_w \cap K_v$, we can choose $\omega, \tau \in \Sigma$ so that $x = \pi(w\omega) = \pi(v\tau)$. As $\Sigma_w \cap \Sigma_v = \emptyset$, there exists $k < \min\{|w|, |v|\}$ such that $w_1w_2\cdots w_k = v_1v_2\cdots v_k$ and $w_{k+1} \neq v_{k+1}$. As $F_{w_1w_2\cdots w_k}$ is injective, it follows that $\pi(\sigma^k(w\omega)) = \pi(\sigma^k(v\tau))$ Hence we can conclude that $\sigma^k(w\omega), \sigma^k(v\tau) \in \mathcal{C}$ and therefore $\omega, \tau \in \mathcal{P}$.

(3) If π is injective, then K is homeomorphic to Σ and hence $\mathcal{C} = \emptyset$. Conversely, if π is not injective, we can use the same discussion as in Proposition 1.2.5 to show that $\mathcal{C} \neq \emptyset$. □

A self-similar structure $(K, S, \{F_i\}_{i\in S})$ may contain an unnecessary symbol. For example, let $K = [0, 1]$ and define $S = \{1, 2, 3\}$, $F_1(x) = \frac{1}{2}x$, $F_2(x) = \frac{1}{2}x + \frac{1}{2}$ and $F_3(x) = \frac{1}{2}x + \frac{1}{4}$. Then obviously $K = F_1(K) \cup F_2(K)$ and we don't need F_3 to describe K. This example may be a little artificial but there are more natural examples. To explain such examples, we need to introduce some notations.

Let $\mathcal{L} = (K, S, \{F_i\}_{i\in S})$ be a self-similar structure. Let W be a finite subset of $W_*\backslash W_0$. Then $\Sigma(W) = W^{\mathbb{N}}$ can be identified as a subset of $\Sigma(S) = S^N$ in the natural manner. Set $K(W) = \pi(\Sigma(W))$. Then $(K(W), W, \{F_w\}_{w\in W})$ becomes a new self-similar structure. We denote this self-similar structure by $\mathcal{L}(W)$.

By using these notations, we can rephrase the above example as $K(\{1,2\}) = K(S)$. The following is a more natural example.

Example 1.3.6. Let $K = [0,1]$ and define $S = \{1,2\}$, $F_1(x) = \frac{3}{4}x$ and $F_2(x) = \frac{3}{4}x + \frac{1}{4}$. Then $\mathcal{L} = (K, S, \{F_1, F_2\})$ is a self-similar structure. Set $W = \{11, 22\}$, then $K(W) = K$ because $K = F_{11}(K) \cup F_{22}(K)$. This means that to describe K, we don't need the words $\{12, 21\}$.

You may notice that this kind of unnecessary symbols (or words) occurs when the overlap set $C_{\mathcal{L}}$ (or equivalently $\mathcal{C}_{\mathcal{L}}$) is "large". The following theorem justifies such an intuition.

Theorem 1.3.7. *Let $\mathcal{L} = (K, S, \{F_i\}_{i\in S})$ be a self-similar structure. The following conditions are equivalent. If $\mathcal{L}$ satisfies one of the following conditions, we say that $\mathcal{L}$ is minimal.*

(Mi1) *If $\pi(A) = K$ for a closed set $A \subseteq \Sigma$, then $A = \Sigma$.*

(Mi2) *For any $w \in W_*$, K_w is not contained in $\cup_{v\in W_m\setminus\{w\}} K_v$, where $m = |w|$.*

(Mi3) *If $K(W) = K$ for $W \subseteq W_m$, then $W = W_m$.*

(Mi4) *K_w does not contained in $C_{\mathcal{L}}$ for any $w \in W_*$.*

(Mi5) $\text{int}(\mathcal{C}_{\mathcal{L}}) = \emptyset$.

(Mi6) $\text{int}(\mathcal{P}_{\mathcal{L}}) = \emptyset$. (Mi6*) $\mathcal{P}_{\mathcal{L}} \neq \Sigma$.

(Mi7) $\text{int}(V_0) = \emptyset$. (Mi7*) $V_0 \neq K$.

As we can see from (Mi3), a minimal self-similar structure does not have any unnecessary symbol (or word). It is easy to see that the self-similar structures corresponding to the self-similar sets in §1.2 are all minimal.

Proof.

(Mi1) ⇒ (Mi4) Assume that $C \supset K_w$ for some $w \in W_*$. Let $k \in S$ be the first symbol of w. Then for any $x \in K_w$, there exists some $j \neq k$ such that $x \in K_j$. If $m = |w|$ and $A = \cup_{v\in W_m\setminus\{w\}} \Sigma_v$, then A is closed and $\pi(A) = K$.

(Mi4) ⇒ (Mi5) Assume that $\text{int}(\mathcal{C}) \neq \emptyset$. Then $\mathcal{C} \supset \Sigma_w$ for some $w \in W_*$. Hence $C \supset K_w$.

(Mi5) ⇒ (Mi6*) Assume that $\mathcal{P} = \Sigma$. Then as $\mathcal{P} = \cup_{m\geq 1}\sigma^m\mathcal{C}$, Baire's category argument shows that $\text{int}(\sigma^m\mathcal{C}) \neq \emptyset$ for some m. (See, for example, [147] about Baire's category argument.) Hence, $\sigma^m\mathcal{C} \supseteq \Sigma_w$ for some $w \in W_*$. Therefore $\sigma^k\mathcal{C} = \Sigma$ for $k = m + |w|$. Now $\sigma^k\mathcal{C} = \cup_{v\in W_k}\sigma^k(\Sigma_v \cap \mathcal{C})$. Again using Baire's category argument, it follows that $\sigma^k(\Sigma_v \cap \mathcal{C}) \supseteq \Sigma_u$ for some $v \in W_k$ and $u \in W_*$. Therefore $\mathcal{C} \supseteq \Sigma_{vu}$.

(Mi6*)⇒ (Mi6) Assume that $\text{int}(\mathcal{P}) \neq \emptyset$. Then $\mathcal{P} \supset \Sigma_w$ for some $w \in W_*$. Since $\sigma^m\mathcal{P} \subset \mathcal{P}$ for $m = |w|$, we have $\Sigma = \mathcal{P}$.

(Mi6) ⇒ (Mi7) As $\pi^{-1}(V_0) = \mathcal{P}$, we have $\pi^{-1}(\text{int}(V_0)) \subseteq \text{int}(\mathcal{P})$.

(Mi7) ⇒ (Mi7*) ⇒ (Mi6*) This is obvious by the fact that $\pi^{-1}(V_0) = \mathcal{P}$.

(Mi6*) $\Rightarrow$ (Mi2) Assume that $K_w \subseteq \cup_{v \in W_m \backslash \{w\}} K_v$ for some m and $w \in W_m$. Then for any $\omega \in \Sigma$, there exist $v \in W_m \backslash \{w\}$ and $\tau \in \Sigma$ such that $\pi(w\omega) = \pi(v\tau)$. As $w \neq v$, we can choose $k \leq m$ so that $w_1 w_2 \cdots w_{k-1} = v_1 v_2 \cdots v_{k-1}$ and $w_k \neq v_k$. Since $F_{w_1 w_2 \cdots w_{k-1}}$ is injective, we see that $\sigma^k(w\omega) \in \mathcal{C}$. Therefore $\omega \in \mathcal{P}$. So $\mathcal{P} = \Sigma$.

(Mi2) $\Rightarrow$ (Mi1) If there exists a closed subset $A \subset \Sigma$ with $\pi(A) = K$, then A^c is a non-empty open set and so it should contain Σ_w for some $w \in W_*$. Since $A \supset \cup_{v \in W_m \backslash \{w\}} \Sigma_v$, where $m = |w|$, we have $K_w \in \cup_{v \in W_m \backslash \{w\}} K_v$.

(Mi2) $\Rightarrow$ (Mi3) Let W be a proper subset of W_m and assume $K(W) = K$. Then for $w \in W_m \backslash W$, $K_w \subset K = \cup_{v \in W} K_v$. Hence (Mi2) does not hold.

(Mi3) $\Rightarrow$ (Mi2) If $K_w \subset \cup_{v \in W_m \backslash \{w\}} K_v$, where $m = |w|$, then $K = \cup_{v \in W} F_v(K)$, where $W = W_m \backslash \{w\}$. Hence, for any $x \in K$, there exists $\omega \in \Sigma(W)$ such that $\pi(\omega) = x$. Therefore $K(W) = K$. □

Remark. It seems quite possible that the condition $\text{int}(C_{\mathcal{L}}) = \emptyset$ is also equivalent to those conditions in Theorem 1.3.7 as well. Unfortunately this is not true. In fact, there is an example where $\text{int}(C_{\mathcal{L}}) = \emptyset$ but $\text{int}(\mathcal{C}_{\mathcal{L}}) \neq \emptyset$. See Exercise 1.5.

Definition 1.3.8. Let S be a finite set. We say that a finite subset $\Lambda \subset W_*(S)$ is a partition of $\Sigma(S)$ if $\Sigma_w \cap \Sigma_v = \emptyset$ for any $w \neq v \in \Lambda$ and $\Sigma = \cup_{w \in \Lambda} \Sigma_w$. A partition Λ_1 is said to be a refinement of a partition Λ_2 if and only if either $\Sigma_w \subseteq \Sigma_v$ or $\Sigma_w \cap \Sigma_v = \emptyset$ for any $(w, v) \in \Lambda_1 \times \Lambda_2$.

W_m is a partition for any $m \geq 0$ and W_n is a refinement of W_m if (and only if) $n \geq m$.

Lemma 1.3.9. *Let $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$ be a self-similar structure. Define $V(\Lambda, \mathcal{L}) = \cup_{w \in \Lambda} F_w(V_0)$ if Λ is a partition of Σ. Then $V(\Lambda_1, \mathcal{L}) \supseteq V(\Lambda_2, \mathcal{L})$ if Λ_1 is a refinement of Λ_2.*

Proof. Assume that Λ_1 is a refinement of Λ_2. Set $x = \pi(w\omega)$ for $w \in \Lambda_2$ and $\omega \in \mathcal{P}$. Then there exists $v \in \Lambda_1$ such that $v = w\omega_1 \cdots \omega_k$. As $\omega_{k+1}\omega_{k+2} \cdots \in \mathcal{P}$, we can see that $x = \pi(w\omega) \in V(\Lambda_1, \mathcal{L})$. □

Lemma 1.3.10. *Let $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$ be a self-similar structure. Define $V_m(\mathcal{L}) = V(W_m, \mathcal{L})$. Then $V_m(\mathcal{L}) \subseteq V_{m+1}(\mathcal{L})$ and*

$$V_{m+1}(\mathcal{L}) = \cup_{i \in S} F_i(V_m(\mathcal{L})).$$

Furthermore, set $V_*(\mathcal{L}) = \cup_{m \geq 0} V_m(\mathcal{L})$. If $V_0 \neq \emptyset$, then $V_*(\mathcal{L})$ is dense in K.

Proof. The proof of the first statement is straight forward from Lemma 1.3.9. If $x = \pi(\omega) \in K$, then for $\tau \in \mathcal{P}$, $x_n = \pi(\omega_1 \cdots \omega_n \tau)$ converges to x as $n \to \infty$. Hence $V_*(\mathcal{L})$ is dense in K. □

We write V_m instead of $V_m(\mathcal{L})$ if no confusion may occur.

Let Λ be a partition of $\Sigma(S)$. If $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$ is a self-similar structure, then we can define a self-similar structure $\mathcal{L}(\Lambda) = (K(\Lambda), \Lambda, \{F_w\}_{w \in \Lambda})$ as before except $\Lambda = W_0$. Immediately by Definition 1.3.8, it follows that $K(\Lambda) = K$ and $\Sigma(S) = \Sigma(\Lambda)$. Of course, the topological structures of K and $K(\Lambda)$ should be same since they are virtually the same self-similar structures.

Proposition 1.3.11. *Let Λ be a partition of $\Sigma(S)$ and let $\mathcal{L} = (K, S, \{F_i\}_{i\in S})$ be a self-similar structure. Then $\mathcal{P}_{\mathcal{L}} \supseteq \mathcal{P}_{\mathcal{L}(\Lambda)}$, where we identify $\Sigma(S)$ and $\Sigma(\Lambda)$ through the natural mapping. Furthermore, if $\Lambda = W_m(S)$ for $m \geq 1$, then $\mathcal{P}_{\mathcal{L}} = \mathcal{P}_{\mathcal{L}(\Lambda)}$.*

Proof. Let $\alpha = \alpha_1\alpha_2\cdots \in \mathcal{P}_{\mathcal{L}(\Lambda)}$, where $\alpha_i \in \Lambda$. Then there exists $\beta_1\beta_2\cdots\beta_m \in W_*(\Lambda)\backslash W_0(\Lambda)$ and $\gamma = \gamma_1\gamma_2\cdots \in \Sigma(\Lambda)$ such that $\pi(\beta) = \pi(\gamma)$ and $\beta_1 \neq \gamma_1$, where $\beta = \beta_1\beta_2\cdots\beta_m\alpha \in \Sigma(\Lambda)$. Hence, if $\beta_1 = w_1w_2\cdots w_m \in W_m(S)$ and $\gamma_1 = v_1v_2\cdots v_n \in W_n(S)$, we can find k so that $w_1w_2\cdots w_k = v_1v_2\cdots v_k$ and $w_{k+1} \neq v_{k+1}$. Therefore as elements in $\Sigma(S)$, $\pi(\sigma^k\beta) = \pi(\sigma^k\gamma)$ and hence $\sigma^k\beta \in \mathcal{C}_{\mathcal{L}}$. This implies that $\alpha \in \mathcal{P}_{\mathcal{L}}$.

Next let $\Lambda = W_m$ for $m \geq 1$. For $\omega = \omega_1\omega_2\cdots \in \mathcal{P}_{\mathcal{L}}$, there exists $w \in W_*(S)\backslash W_0(S)$ and $\tau \in \Sigma(S)$ such that $\pi(w\omega) = \pi(\tau)$ and $w_1 \neq \tau_1$. Now we can choose $v \in W_*(S)$ so that $vw = \beta_1\beta_2\cdots\beta_j$ and $v\tau = \gamma_1\gamma_2\cdots$ with $\beta_i, \gamma_i \in \Lambda$ and $\beta_1 \neq \gamma_1$. If $\omega = \alpha_1\alpha_2\cdots$, where $\alpha_i \in \Lambda$, then it follows that $\beta_1\beta_2\cdots\beta_j\alpha_1\alpha_2\cdots \in \mathcal{C}_{\mathcal{L}(\Lambda)}$. Therefore $\omega = \alpha_1\alpha_2\cdots \in \mathcal{P}_{\mathcal{L}(\Lambda)}$. □

Even if $\Lambda \neq W_m(S)$, $\mathcal{P}_{\mathcal{L}(\Lambda)}$ often coincides with $\mathcal{P}_{\mathcal{L}}$. In general, however, this is not true. See Exercise 1.6 and 1.7 for examples.

Finally, we will give the definition of post critically finite (p. c. f. for short) self-similar structure, which is one of the key notions in this book.

Definition 1.3.12. Let $\mathcal{L} = (K, S, \{F_i\}_{i\in S})$ be a self-similar structure. $\mathcal{L}$ is said to be post critically finite or p. c. f. in short if and only if the post critical set $\mathcal{P}_{\mathcal{L}}$ is a finite set.

If $\mathcal{L} = (K, S, \{F_i\}_{i\in S})$ is post critically finite, V_m is a finite set for all m. In particular, $K_w \cap K_v$ is a finite set for any $w \neq v \in W_m$. Such a self-similar set is often called a finitely ramified self-similar set. Obviously, a p. c. f. self-similar set is finitely ramified. The converse is, however, not true.

Later in Chapter 3, we will mainly study analysis on post critically finite self-similar sets.

Example 1.3.13 (Sierpinski gasket). Let K be the Sierpinski gasket defined in Example 1.2.8. Then $\mathcal{L} = (K, S, \{f_i\}_{i\in S})$, where $S = \{1, 2, 3\}$ and f_i are the same maps as in Example 1.2.8, is a post critically finite self-similar structure. In fact, $C_{\mathcal{L}} = \{q_1, q_2, q_3\}$, $\mathcal{C}_{\mathcal{L}} = \{1\dot{2}, 2\dot{1}, 2\dot{3}, 3\dot{2}, 3\dot{1}, 1\dot{3}\}$ and $\mathcal{P}_{\mathcal{L}} = \{\dot{1}, \dot{2}, \dot{3}\}$. Also $V_0 = \{p_1, p_2, p_3\}$.

Example 1.3.14 (Hata's tree-like set). Let f_1 and f_2 be the same as in Example 1.2.9. Also let K be the Hata's tree-like set. Then $\mathcal{L} = (K, \{1, 2\}, \{f_1, f_2\})$ is a p. c. f. self-similar structure. In fact, $C_{\mathcal{L}} = \{|c^2|\}$, $\mathcal{C}_{\mathcal{L}} = \{11\dot{2}, 2\dot{1}\}$ and $\mathcal{P}_{\mathcal{L}} = \{1\dot{2}, \dot{2}, \dot{1}\}$. Hence $V_0 = \{c, 0, 1\}$. Note that self-similar structures are isomorphic for all c with $|c|, |1 - c| \in (0, 1)$.

Of course there are numerous examples of non-p. c. f. self-similar structure. One easy example is the unit square. (See Exercise 1.3.) Another famous example is the Sierpinski carpet, which may be thought of as the simplest non-trivial non-p. c. f. self-similar structure.

Example 1.3.15 (Sierpinski carpet). Let $p_1 = 0, p_2 = 1/2, p_3 = 1, p_4 = 1 + \sqrt{-1}/2, p_5 = 1 + \sqrt{-1}, p_6 = 1/2 + \sqrt{-1}, p_7 = \sqrt{-1}$ and $p_8 = \sqrt{-1}/2$. Set $f_i(z) = (z - p_i)/3 + p_i$ for $i = 1, 2, \cdots, 8$. The self-similar set K with respect to $\{f_i\}_{i=1,2,\cdots,8}$ is called the Sierpinski carpet. Let $\mathcal{L}$ be the corresponding self-similar structure. The $\mathcal{L}$ is not post critically finite. In fact, $C_{\mathcal{L}}$, $\mathcal{C}_{\mathcal{L}}$ and $P_{\mathcal{L}}$ are infinite sets. In particular, V_0 equals to the boundary of the unit square $[0, 1] \times [0, 1]$.

§1.4 Self similar measure

In this section, we will introduce an important class of measures on a self-similar structure, that is, self-similar measures. First we will recall some of fundamental definitions in measure theory.

$(X, \mathcal{M})$ is called a measurable space if X is a set and $\mathcal{M}$ is a σ-algebra whose elements are subsets of X. A measure μ on a measurable space $(X, \mathcal{M})$ is a non-negative σ-additive function defined on $\mathcal{M}$.

Definition 1.4.1. Let (X, d) be a metric space and let μ be a measure on a measurable space $(X, \mathcal{M})$.

(1) The Borel σ-algebra, $\mathcal{B}(X, d)$, is the minimal σ-algebra which contains all open subset of X. An element of $\mathcal{B}(X, d)$ is called a Borel set. If no confusion may occur, we write $\mathcal{B}(X)$ instead of $\mathcal{B}(X, d)$.

(2) μ is called a Borel measure if $\mathcal{M}$ contains $\mathcal{B}(X)$.

(3) μ is called a Borel regular measure if it is a Borel measure and, for any $A \in \mathcal{M}$, there exists $B \in \mathcal{B}(X)$ such that $\mu(A) = \mu(B)$ and $A \subseteq B$.

(4) We say that μ is complete if any subset of a null set is measurable, i.e. $B \in \mathcal{M}$ if $B \subseteq A \in \mathcal{M}$ and $\mu(A) = 0$.

(5) μ is called a probability measure if and only if $\mu(X) = 1$.

The following proposition is one of the most important fact about a Borel regular measures.

Proposition 1.4.2. *Let (X, d) be a metric space and let μ be a Borel regular measure on $(X, \mathcal{M})$. Assume that $\mu(X) < \infty$. Then for any $A \in \mathcal{M}$,*

$$\begin{aligned} \mu(A) &= \inf\{\mu(U) : U \text{ is a open set that contains } A\} \\ &= \sup\{\mu(F) : F \text{ is a closed set that is contained in } A\} \end{aligned}$$

Proposition 1.4.3 (Bernoulli measure). *Let S be a finite set. If $p = (p_i)_{i \in S}$ satisfies that $\sum_{i \in S} p_i = 1$ and that $0 < p_i < 1$ for any $i \in S$, then there* *exists a unique complete Borel regular measure μ^p on $(\Sigma, \mathcal{M}^p)$, where $\Sigma = S^{\mathbb{N}}$, that satisfies $\mu^p(\Sigma_w) = p_{w_1}p_{w_2}\cdots p_{w_m}$ for any $w = w_1w_2\cdots w_m \in W_*$. This measure μ^p is called the Bernoulli measure on Σ with weight p.*

Remark. In this book, all the measures we will encounter are supposed to be complete unless otherwise stated.

Also the Bernoulli measure with weight p is characterized as the unique Borel regular probability measure on Σ that satisfies

$$\mu(A) = \sum_{i \in S} p_i \mu(\sigma_i^{-1}(A))$$

for any Borel set $A \subset \Sigma$.

Proposition 1.4.4 (Self-similar measures). *Let $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$ be a self-similar structure and let π be the natural map from Σ to K associated with $\mathcal{L}$. If $p = (p_i)_{i \in S} \in \mathbf{R}^S$ satisfies $\sum_{i \in S} p_i = 1$ and $0 < p_i < 1$ for any $i \in S$, then we define ν^p by $\nu^p(A) = \mu^p(\pi^{-1}(A))$ for $A \in \mathcal{N}^p = \{A : A \subseteq K, \pi^{-1}(A) \in \mathcal{M}^p\}$. Then, ν^p is a Borel regular measure on $(K, \mathcal{N}^p)$. ν^p is called the self-similar measure on K with weight p.*

It is known that ν^p is the unique Borel regular probability measure on K that satisfies

$$\nu(A) = \sum_{i \in S} p_i \nu(F_i^{-1}(A))$$

for any Borel set $A \subset K$.

By definition, we see that

$$\nu^p(K_w) \geq p_{w_1}p_{w_2}\cdots p_{w_m} \tag{1.4.1}$$

for any $w = w_1w_2\cdots w_m \in W_*$. Intuitively, it seems that equality holds in (1.4.1) rather than inequality if the overlapping set $C_{\mathcal{L}}$ is small enough. Precisely we have the following theorem.

Theorem 1.4.5. *Let $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$ be a self-similar structure and let π be the natural map from Σ to K associated with $\mathcal{L}$. Also let $p = (p_i)_{i \in S}$ satisfy $\sum_{i \in S} p_i = 1$ and $0 < p_i < 1$ for any $i \in S$. Then*

$$\nu^p(K_w) = p_{w_1}p_{w_2}\cdots p_{w_m}$$

for any $w = w_1w_2\cdots w_m \in W_*$ if and only if $\mu^p(\mathcal{I}_\infty) = 0$, where $\mathcal{I}_\infty = \{\omega \in \Sigma : \#(\pi^{-1}(\pi(\omega))) = +\infty\}$.

Remark. We will show that $\mathcal{I}_\infty \in \mathcal{M}^p$.

Lemma 1.4.6. *For any $A \in \mathcal{M}^p$, define*

$$A_\circ = \{\omega \in \Sigma : \sigma^m\omega \in A \text{ for infinitely many } m \in \mathbf{N}.\}.$$

Then $A_\circ \in \mathcal{M}^p$ and $\mu^p(A_\circ) \geq \mu^p(A)$. In particular, if $A \in \mathcal{B}(\Sigma)$ then $A_\circ \in \mathcal{B}(\Sigma)$.

Proof. Set $A_m = \cup_{w \in W_m} \sigma_w(A)$, where $\sigma_w = \sigma_{w_1} \circ \cdots \sigma_{w_m}$ for $w = w_1 w_2 \cdots w_m \in W_m$. Then $A_\circ = \limsup_{m\to\infty} A_m$. Hence $A_\circ \in \mathcal{M}^p$ and by the Fatou's lemma, we have $\mu^p(A_\circ) \geq \limsup_{m\to\infty} \mu^p(A_m)$. (Note that μ^p is a finite measure.) On the other hand, $\mu^p(A_m) = \sum_{w\in W_m} \mu^p(\sigma_w(A)) = \sum_{w\in W_m} p_w \mu^p(A) = \mu^p(A)$, where $p_w = p_{w_1} p_{w_2} \cdots p_{w_m}$. Hence it follows that $\mu^p(A_\circ) \geq \mu^p(A)$. □

Lemma 1.4.7. *Define* $\mathcal{I} = \{\omega \in \Sigma : \#(\pi^{-1}(\pi(\omega))) > 1\}$. *Then* $\mathcal{I} \in \mathcal{B}(\Sigma)$, $\mathcal{I}_\infty \in \mathcal{M}^p$, $\mathcal{I}_\circ \subseteq \mathcal{I}_\infty \subseteq \mathcal{I}$ *and* $\mu^p(\mathcal{I}_\circ) = \mu^p(\mathcal{I}_\infty) = \mu^p(\mathcal{I})$.

Proof. Set $I_m = \cup_{w\neq v \in W_m}(K_w \cap K_v)$. Then I_m is closed set and $\mathcal{I} = \cup_{m\geq 1}\pi^{-1}(I_m)$. Hence $\mathcal{I} \in \mathcal{B}(\Sigma)$. Now if $\omega \in \mathcal{I}_\circ$, by using inductive argument, we can choose $\{m_k\}_{k\geq 1}$, $\{n_k\}_{k\geq 1}$ and $\{\omega^{(k)}\}_{k\geq 1}, \{\tau^{(k)}\}_{k\geq 1} \subset \Sigma$ so that

$$1 \leq m_1 < n_1 < m_2 < n_2 < \cdots < m_k < n_k < m_{k+1} < \cdots,$$

$\sigma^{m_k}\omega \in \mathcal{I}, \sigma^{m_k}\omega \neq \tau^{(k)}, \pi(\sigma^{m_k}\omega) = \pi(\tau^{(k)}), \omega^{(k)} = \omega_1\omega_2\cdots\omega_{m_k}\tau^{(k)}, \omega_1\omega_2\cdots\omega_{n_{k-1}} = {\omega^{(k)}}_1{\omega^{(k)}}_2\cdots{\omega^{(k)}}_{n_{k-1}}$ and $\omega_{n_k} \neq {\omega^{(k)}}_{n_k}$. This implies that $\pi(\omega^{(k)}) = \pi(\omega)$ and hence $\omega \in \mathcal{I}_\infty$. Thus we have shown that $\mathcal{I}_\circ \subseteq \mathcal{I}_\infty \subseteq \mathcal{I}$. By Lemma 1.4.6, $\mu^p(\mathcal{I}) \leq \mu^p(\mathcal{I}_\circ)$, we can see that $\mu^p(\mathcal{I}) = \mu^p(\mathcal{I}_\circ)$. As μ^p is complete, $\mathcal{I}_\infty \in \mathcal{M}^p$ and $\mu^p(\mathcal{I}_\infty) = \mu^p(\mathcal{I})$. □

Proof of Theorem 1.4.5. By the definition of $\mathcal{I}$, we can easily see that $\mu^p(\mathcal{I}) = 0$ if and only if $\mu^p(\Sigma_w) = \nu^p(K_w) = p_w$ for any $w \in W_*$. This along with Lemma 1.4.7 implies the theorem. □

Remark. It is well-known that μ^p is ergodic with respect to the shift map σ. This means that if $A \in \mathcal{M}^p$ and $\sigma^{-1}(A) = A$ then $\mu^p(A) = 0$ or 1. Since $\sigma^{-1}(\mathcal{I}_\circ) = \mathcal{I}_\circ$, $\mu^p(\mathcal{I}_\circ) = \mu^p(\mathcal{I}_\infty) = \mu^p(\mathcal{I}) = 0$ or 1.

Corollary 1.4.8. *If* $\pi^{-1}(x)$ *is a finite set for any* $x \in K$, *then* $\nu^p(K_w) = p_w$ *for all* $w \in W_*$.

Since $\mathcal{I} = \cup_{w\in W_*}\sigma_w(\mathcal{C}_\mathcal{L})$, $\mu^p(\mathcal{C}_\mathcal{L}) > 0$ implies $\mu^p(\mathcal{I}) > 0$. Hence by Theorem 1.3.7, we have the following corollary.

Corollary 1.4.9. *If* $\nu^p(K_w) = p_w$ *for any* $w \in W_*$, *then* $\mathcal{L}$ *is minimal.*

Although the next theorem does not directly related to self-similar measures, it tells us a useful fact: two Borel regular measures on a self-similar sets are comparable if they are comparable on K_w for all $w \in W_*$.

Theorem 1.4.10. *Let* $\mathcal{L} = (K, S, \{F_i\}_{i\in S})$ *be a self-similar structure. Let* μ *and* ν *be Borel regular measures on* $(K, \mathcal{M}(\mu))$ *and* $(K, \mathcal{M}(\nu))$ *respectively. Assume that* $\nu(K) < \infty$ *and* $\nu(\mathcal{I}) = 0$. *If there exists* $c > 0$ *such that* $\mu(K_w) \leq c\nu(K_w)$ *for any* $w \in W_*$, *then* $\mu(A) \leq c\nu(A)$ *for any* $A \in \mathcal{M}(\mu) \cap \mathcal{M}(\nu)$. *In particular,* $\mu(\mathcal{I}) = 0$.

Proof. Let U be an open subset of K. Set $W(U) = \{w \in W_* : K_w \subset U\}$. For $w, v \in W(U)$, we define $w \geq v$ if and only if $\Sigma_w \supseteq \Sigma_v$. Then $\geq$ is a partial order on $W(U)$. If $W^+(U)$ is the collection of maximal elements in $W(U)$ with respect to this order, then $U = \cup_{w \in W^+(U)} K_w$ and $K_w \cap K_v \subset \mathcal{I}$ for $w \neq v \in W^+(U)$. Therefore

$$\mu(U) \leq \sum_{w \in W^+(U)} \mu(K_w) \leq c \sum_{w \in W^+(U)} \nu(K_w) = c\nu(U).$$

Now by Proposition 1.4.2, for any $A \in \mathcal{M}(\mu) \cap \mathcal{M}(\nu)$, there exists a decreasing sequence of open sets $\{O_k\}_{k \geq 1}$ such that $A \subseteq O_k$ for any k, $\mu(\cap_{k \geq 1} O_k) = \mu(A)$ and $\nu(\cap_{k \geq 1} O_k) = \nu(A)$. As $\mu(O_k) \leq c\nu(O_k)$, we have $\mu(A) \leq c\nu(A)$. □

§1.5 Dimension of self similar sets

In this section, we will introduce the notion of Hausdorff dimension of metric spaces and show how to calculate a Hausdorff dimension of self-similar sets.

Definition 1.5.1. Let (X, d) be a metric space. For any bounded set $A \subset X$, we define

$$\mathcal{H}^s_\delta(A) = \inf\{\sum_{i \geq 1} \mathrm{diam}(E_i)^s : A \subset \cup_{i \geq 1} E_i, \mathrm{diam}(E_i) \leq \delta\},$$

where $\mathrm{diam}(E)$ is a diameter of a set E defined by $\mathrm{diam}(E) = \sup_{x,y \in E} d(x, y)$. Also, we define $\mathcal{H}^s(A) = \limsup_{\delta \downarrow 0} \mathcal{H}^s_\delta(A)$.

It is well-known that $\mathcal{H}^s$ become a complete Borel regular measure for any $s > 0$. $\mathcal{H}^s$ is called the s-dimensional Hausdorff measure of (X, d).

Lemma 1.5.2. *Let (X, d) be a metric space. For $0 \leq s < t$,*

$$\mathcal{H}^t_\delta(E) \leq \delta^{t-s} \mathcal{H}^s_\delta(E)$$

for any $E \subseteq X$.

Proof. If $E \subseteq \cup_{i \geq 1} E_i$ and $\mathrm{diam}(E_i) \leq \delta$ for any i, then

$$\sum_{i \geq 1} \mathrm{diam}(E_i)^t \leq \sum_{i \geq 1} \mathrm{diam}(E_i)^{t-s} \mathrm{diam}(E_i)^s \leq \delta^{t-s} \sum_{i \geq 1} \mathrm{diam}(E_i)^s.$$

□

By Lemma 1.5.2, we can see the following proposition.

Proposition 1.5.3. *For any $E \subseteq X$,*

$$\sup\{s : \mathcal{H}^s(E) = \infty\} = \inf\{s : \mathcal{H}^s(E) = 0\}. \tag{1.5.1}$$

Proof. By Lemma 1.5.2, if $s < t$, then $\mathcal{H}^s(E) < \infty$ implies $\mathcal{H}^t(E) = 0$ and also $\mathcal{H}^t(E) > 0$ implies $\mathcal{H}^s(E) = \infty$. Now it is easy to see (1.5.1). □

Definition 1.5.4 (Hausdorff dimension). The value given by (1.5.1) is called the Hausdorff dimension of E, which is denoted by $\dim_{\mathrm{H}} E$.

Remark. The Hausdorff measures and the Hausdorff dimension depend on a metric d. In this sense, if one should specify which metric we are looking at, we would use the notation of $\dim_{\mathrm{H}}(E, d)$ instead of $\dim_{\mathrm{H}} E$.

The following lemma is often useful to calculate a Hausdorff dimension of a metric space. It is often called "Frostman's lemma". See, for example, Mattila [99]. It is also called the "mass distribution principle" in Falconer [28].

Lemma 1.5.5. *Let (K, d) be a compact metric space. If $\mathcal{H}^\alpha(K) < \infty$ and there exist positive constants c, l_0 and a probability measure μ on K such that*

$$\mu(B_l(x)) \leq c l^\alpha$$

for all $x \in K$ and any $l \in (0, l_0)$, then

$$\mu(A) \leq c\mathcal{H}^\alpha(A)$$

for any Borel set $A \subset K$. In particular, $0 < \mathcal{H}^\alpha(K) < \infty$.

Remark. According to the discussion of Moran [107], the converse of the above lemma is true : If $0 < \mathcal{H}^\alpha(K) < \infty$, then there exists a probability measure μ on K such that, for some $c > 0$,

$$\mu(B_l(x)) \leq c l^\alpha$$

for all $x \in K$ and $l > 0$. Moran proved this fact if K was a compact subset of Eudlidean space. His argument, however, can be easily extended to this case.

Proof. For $U \subset K$ and $x \in U$, note that $U \subset B_{\mathrm{diam}(U)}(x)$. Hence, if $A \subseteq \cup_i U_i$, then

$$\mu(A) \leq \sum_i \mu(B_{\mathrm{diam}(U_i)}(x_i)) \leq c \sum_i \mathrm{diam}(U_i)^\alpha,$$

where $x_i \in U_i$. Therefore $\mu(A) \leq c\mathcal{H}^\alpha_l(A)$. Letting $l \to 0$, it follows that $\mu(A) \leq c\mathcal{H}^\alpha(A)$. □

Now let $(K, \{1, 2, \cdots, N\}, \{F_i\}_{1 \leq i \leq N})$ be a self-similar structure and let d be a metric on K which is compatible with the original topology of K. In general, it is not easy to evaluate the Hausdorff dimension $\dim_{\mathrm{H}}(K, d)$. Moran [107] introduced what is now called "the open set condition", which ensures that the intersections $K_i \cap K_j$ for $i \neq j \in \{1, 2, \cdots, N\}$ are "small". Under this condition, he gave a formula for the Hausdorff dimension of K when K is a subset of $\mathbf{R}^k$, d is the Euclidean metric on $\mathbf{R}^k$ and F_i are similitudes with respect to d. See Proposition 1.5.8 and Corollary 1.5.9 for the Moran's result. His result is useful to calculate Hausdorff dimensions of many well-known examples of self-similar sets. See Exercise 1.9.

Remark. It was well before the notion of "fractal" when Moran published his paper [107]. Of course, he didn't use the terminology "self-similar set" but he had exactly the same notion of self-similar sets as we have today. Hutchinson [57] rediscovered Moran's result about 40 years later and introduced the name "open set condition".

Unfortunately we can apply Moran's result only when K is a subset of $\mathbf{R}^k$, d is the Euclidean metric on $\mathbf{R}^k$ and F_i are similitudes with respect to d. Later, a metric called an effective resistance metric, which satisfies none of those requirements, will become important from the analytical point of view. Here, we will introduce an extended version of Moran's theorem (Theorem 1.5.7) that can be applied in more general situations.

Definition 1.5.6. For $\mathbf{r} = (r_1, r_2, \cdots, r_N)$ where $0 < r_i < 1$ and for $0 < a < 1$,

$$\Lambda(\mathbf{r}, a) = \{w : w = w_1 w_2 \cdots w_m \in W_*, r_{w_1 w_2 \cdots w_{m-1}} > a \geq r_w\},$$

where $r_v = r_{v_1} r_{v_2} \cdots r_{v_k}$ for $v = v_1 v_2 \cdots v_k \in W_k$

Remark. $\Lambda(\mathbf{r}, a)$ becomes a partition of Σ.

The following is our main theorem. This theorem was introduced in Kigami [68]. The essential ideas are, however, essentially the same as in Moran [107].

Theorem 1.5.7. *Assume that there exist* $\mathbf{r} = (r_1, r_2, \cdots, r_N)$ *where* $0 < r_i < 1$ *and positive constants* c_1*,* c_2*,* c_* *and* M *such that*

$$\mathrm{diam}(K_w) \leq c_1 r_w \tag{1.5.2}$$

for all $w \in W_*$ *and*

$$\#\{w : w \in \Lambda(\mathbf{r}, a), d(x, K_w) \leq c_2 a\} \leq M \tag{1.5.3}$$

for any $x \in K$ *and any* $a \in (0, c_*)$*, where* $d(x, K_w) = \inf_{y \in K_w} d(x, y)$*. Then there exist constants* $c_3, c_4 > 0$ *such that for all* $A \in \mathcal{B}(K, d)$*,*

$$c_3 \nu(A) \leq \mathcal{H}^\alpha(A) \leq c_4 \nu(A), \tag{1.5.4}$$

where ν *is a self-similar measure on* K *with weight* $({r_i}^\alpha)_{1 \leq i \leq N}$ *and* α *is the unique positive number that satisfies*

$$\sum_{i=1}^N {r_i}^\alpha = 1. \tag{1.5.5}$$

In particular, $0 < \mathcal{H}^\alpha(K) < \infty$ *and* $\dim_{\mathrm{H}}(K, d) = \alpha$*.*

Remark. Under the assumption (1.5.3), it is easy to see that $\#(\pi^{-1}(x)) \leq M$ for any $x \in K$. Hence by Corollary 1.4.8,

$$\nu(K_w) = {r_w}^\alpha$$

for any $w \in W_*$. Also $\nu(\mathcal{I}) = 0$.

Proof. We write $\Lambda_a = \Lambda(\mathbf{r}, a)$. First we will show that $\mathcal{H}^\alpha(K_w) \leq (c_1)^\alpha \nu(K_w)$ for all $w \in W_*$. For $w = w_1 w_2 \cdots w_m \in W_*$, we define $\Lambda_a(w) = \{v = v_1 v_2 \cdots v_k : wv \in \Lambda_a\}$, where $wv = w_1 w_2 \cdots w_m v_1 v_2 \cdots v_k$. Then we can see that $\Lambda_a(w)$ is a partition for sufficiently small a. Hence

$$r_w{}^\alpha = \sum_{v \in \Lambda_a(w)} r_{wv}{}^\alpha. \tag{1.5.6}$$

By (1.5.2), it follows that $\mathrm{diam}(K_{wv}) \leq c_1 r_{wv} \leq c_1 a$ for $v \in \Lambda_a(w)$. Also note that $K_w = \cup_{v \in \Lambda_a(w)} K_{wv}$. Then we see that

$$\mathcal{H}^\alpha_{c_1 a}(K_w) \leq c_1{}^\alpha \sum_{v \in \Lambda_a(w)} r_{wv}{}^\alpha = (c_1 r_w)^\alpha.$$

Letting $a \to 0$, we obtain

$$\mathcal{H}^\alpha(K_w) \leq (c_1)^\alpha r_w{}^\alpha = (c_1)^\alpha \nu(K_w).$$

Next we show that $\nu(K_w) \leq M c_2{}^{-\alpha} \mathcal{H}^\alpha(K_w)$. Let μ be the Bernoulli measure on Σ with weight $(r_i{}^\alpha)_{1 \leq i \leq N}$. For every $x \in K$,

$$\pi^{-1}(B_{c_2 a}(x)) \subset \bigcup_{w \in \Lambda_{a,x}} \Sigma_w,$$

where $\Lambda_{a,x} = \{w : w \in \Lambda_a, d(x, K_w) \leq c_2 a\}$. Hence it follows that

$$\nu(B_{c_2 a}(x)) \leq \sum_{w \in \Lambda_{a,x}} \mu(\Sigma_w).$$

Since $\mu(\Sigma_w) = r_w^\alpha \leq a^\alpha$ and $\#(\Lambda_{a,x}) \leq M$ by (1.5.3), we have

$$\nu(B_{c_2 a}(x)) \leq M c_2{}^{-\alpha} (c_2 a)^\alpha.$$

Lemma 1.5.5 implies that

$$\nu(A) \leq M c_2{}^{-\alpha} \mathcal{H}^\alpha(A).$$

for any $A \in \mathcal{B}(K, d)$. Hence there exist $c_3, c_4 > 0$ such that

$$c_3 \nu(K_w) \leq \mathcal{H}^\alpha(K_w) \leq c_4 \nu(K_w)$$

By Theorem 1.4.10, we can verify (1.5.4). □

In the rest of this section, we show that the open set condition implies (1.5.2) and (1.5.3) of Theorem 1.5.7.

Proposition 1.5.8. *Suppose K is a subset of $\mathbf{R}^k$, d is the Euclidean metric of $\mathbf{R}^k$ and $F_i : \mathbf{R}^k \to \mathbf{R}^k$ is an r_i-similitude for $i = 1, 2, \cdots, N$ with respect to* *d. If the open set condition holds: there exists an bounded non-empty open set* $O \subset \mathbf{R}^k$ *such that*

$$\bigcup_{i=1}^{N} F_i(O) \subset O \quad \textit{and} \quad F_i(O) \cap F_j(O) = \emptyset \quad \textit{for } i \neq j,$$

then there exist constants $c_1, c_2, M > 0$ *such that*

$$\mathrm{diam}(K_w) \leq c_1 r_w$$

for all $w \in W_*$ *and*

$$\#\{w : w \in \Lambda(\mathbf{r}, a), d(x, K_w) \leq c_2 a\} \leq M$$

for all $0 < a < 1$ *and* $x \in K$.

Proof. We can see that $K_w \subset \overline{O}_w$ for any $w \in W_*$, where $O_w = F_w(O)$. (By Exercise 1.2, it follows that $\overline{O} \supseteq K$.) Without loss of generality, we may assume that $\mathrm{diam}(O) \leq 1$. Then, for all $w \in W_*$, $\mathrm{diam}(K_w) \leq \mathrm{diam}(\overline{O}_w) \leq r_w$. Let m be the k-dimensional Lebesgue measure and let $\Lambda_{a,x} = \{w : w \in \Lambda(\mathbf{r}, a), d(x, K_w) \leq a\}$. Then $\cup_{w \in \Lambda_{a,x}} O_w \subset B_{2a}(x)$. Since O_w are mutually disjoint, we have $\sum_{w \in \Lambda_{a,x}} \mathsf{m}(O_w) \leq \mathsf{m}(B_{2a}(x))$. Hence it follows that $\#(\Lambda_{a,x}) {r_w}^k \mathsf{m}(O) \leq 2^k C a^k$, where $C = \mathsf{m}(\text{unit ball})$. Since $r_w \geq aR$ where $R = \min\{r_1, r_2, \cdots, r_N\}$, we see that $\#(\Lambda_{a,x}) \leq 2^k C R^{-k} \mathsf{m}(O)^{-1}$. □

Corollary 1.5.9 (Moran's theorem). *If* K *satisfies the open set condition, then* $\dim_{\mathrm{H}}(K, d) = \alpha$, *where* α *is given by* (1.5.5) *with* $r_i = \mathrm{Lip}(F_i)$.

§1.6 Connectivity of self similar sets

Let $(K, S, \{F_i\}_{i \in S})$ be a self-similar structure. In this section we will give a simple condition for connectivity of K and also show that K is connected if and only if it is arcwise connected. For a reminder, the definition of connectivity is as follows.

Definition 1.6.1. Let (X, d) be a metric space.

(1) (X, d) is said to be connected if and only if any closed and open subset of X is X or the empty set. Also a subset A of X is said to be connected if and only if the metric space $(A, d|_A)$ is connected.

(2) A subset A of X is said to be arcwise connected if and only if there exists a path between x and y for any $x, y \in A$: there exists a continuous map $p : [0, 1] \to A$ such that $p(0) = x$ and $p(1) = y$.

Of course, arcwise connectivity implies connectivity, but the converse is not true in general. Now we come to the main theorem of this section.

Theorem 1.6.2. *The followings are equivalent.*

(1) *For any* $i, j \in S$*, there exists* $\{i_k\}_{k=0,1,\cdots,n} \subseteq S$ *such that* $i_0 = i$*,* $i_n = j$ *and* $K_{i_k} \cap K_{i_{k+1}} \neq \emptyset$ *for any* $k = 0, 1, \cdots, n-1$*.*

(2) K *is arcwise connected.*

(3) K *is connected.*

Proof. Obviously (2) $\Rightarrow$ (3). So let us show (3) $\Rightarrow$ (1). Choose $i \in S$ and define $A \subseteq S$ by

$$A = \{j \in S : \text{there exists } \{i_k\}_{k=0,1,\cdots,n} \subseteq S \text{ such that } i_0 = i,\ i_n = j \text{ and } K_{i_k} \cap K_{i_{k+1}} \neq \emptyset \text{ for any } k = 0, 1, \cdots, n-1\}$$

If $U = \cup_{j \in A} K_j$ and $V = \cup_{j \notin A} K_j$, then $U \cap V = \emptyset$ and $U \cup V = K$. Also both U and V are closed sets because K_i is closed and A is a finite set. Hence U is an open and closed set. Hence $U = K$ or $U = \emptyset$. Obviously $K_i \subseteq U$ and hence $U = K$. Therefore $V = \emptyset$ and hence $A = S$.

To prove (1) $\Rightarrow$ (2), we need the following lemma.

Lemma 1.6.3. *For a map* $u : [0, 1] \to K$ *and for* $t \in [0, 1]$*, we define*

$$D(u, t) = \sup\{\limsup_{n\to\infty} d(u(t_n), u(s_n)) : \lim_{n\to\infty} t_n = \lim_{n\to\infty} s_n = t\}$$

If $f_n : [0, 1] \to K$ *is uniformly convergent to* $f : [0, 1] \to K$ *as* $n \to \infty$ *and* $\lim_{n\to\infty} D(f_n, s) = 0$*, then* f *is continuous at* s*.*

Proof of Lemma 1.6.3. Let d be a metric on K that is compatible with the original topology of K. If $t_n \to s$ and $s_n \to s$ as $n \to \infty$, then

$$d(f(t_n), f(s_n)) \leq d(f(t_n), f_m(t_n)) + d(f_m(t_n), f_m(s_n)) + d(f_m(s_n), f(s_n)).$$

Set $r_m = \sup\{d(f_m(t), f(t)) : 0 \leq t \leq 1\}$. Then the above inequality implies $D(f, s) \leq 2r_m + D(f_m, s)$. Now letting $m \to \infty$, we can see that $D(f, s) = 0$. Hence f is continuous at s. $\square$

Now we return to the proof of (1) $\Rightarrow$ (2).

Define

$$P = \{f : K^2 \times [0, 1] \to K : f(p, q, 0) = p \text{ and } f(p, q, 1) = q \text{ for any } (p, q) \in K^2\}.$$

Also for $f, g \in P$, set

$$d_P(f, g) = \sup\{d(f(p, q, t), g(p, q, t)) : (p, q, t) \in K^2 \times [0, 1]\}.$$

Then (P, d_p) is a complete metric space. By (1), for any $(p, q) \in K^2$, we can choose $n(p, q)$, $\{i_k(p, q)\}_{0 \leq k \leq n(p,q)-1} \subseteq S$ and $\{x_k(p, q)\}_{0 \leq k \leq n(p,q)} \subseteq K$ so that $x_0(p, q) = p$, $x_{n(p,q)}(p, q) = q$ and $x_k(p, q), x_{k+1}(p, q) \in K_{i_k(p,q)}$ for $k = 0, \cdots, n(p, q) - 1$. For $f \in P$, define $Gf \in P$ by, for $\frac{k}{n(p,q)} \leq t \leq \frac{k+1}{n(p,q)}$,

$$(Gf)(p, q, t) = F_{i_k(p,q)}(f(y_k(p, q), z_k(p, q), n(p, q)t - k)),$$

where $y_k(p,q) = F^{-1}_{i_k(p,q)}(x_k(p,q))$ and $z_k(p,q) = F^{-1}_{i_k(p,q)}(x_{k+1}(p,q))$. Then it follows that $d_P(G^m f, G^m g) \le r_m$, where $r_m = \max_{w \in W_m} \mathrm{diam}(K_w)$. Since $r_m \to 0$ as $m \to \infty$, we see that there exists $f_* \in P$ such that $G^m f \to f_*$ as $m \to \infty$ in P. Also set $D(f) = \sup\{D(f_{(p,q)}, t) : (p,q,t) \in K^2 \times [0,1]\}$ for $f \in P$, where $f_{(p,q)}(t) = f(p,q,t)$. Then $D(G^m f) \le r_m D(f)$. Hence by Lemma 1.6.3, $f_*(p,q,t)$ is continuous with respect to t. As $f_*(p,q,t)$ is a continuous path between p and q, we see that K is arcwise connected. □

Corollary 1.6.4. *If $(K, S, \{F_i\}_{i\in S})$ is post critically finite, then K is connected if and only if, for any $p, q \in V_1$, there exist $\{p_i\}_{0\le i\le m} \subset V_1$ and $\{k_i\}_{0\le i\le m-1} \subseteq S$ such that $p_0 = p$, $p_m = q$ and $p_i, p_{i+1} \in F_{k_i}(V_0)$ for $i = 0, \cdots, m-1$.*

In the rest of this section, we show a proposition which will be used in the following sections.

Proposition 1.6.5. *Let $(K, S, \{F_i\}_{i\in S})$ be a connected post critically finite self-similar structure. Let p be the fixed point of F_i. If J is a connected component of $K\backslash\{p\}$, then $J \cap V_0 \ne \emptyset$. In particular, the number of connected components of $K\backslash\{p\}$ is finite. Moreover, let $\{J_j\}_{j=1,\cdots,m}$ be the collection of all connected components of $K\backslash\{p\}$. Then there exists a permutation of $\{1, \cdots, m\}$, ρ, such that $F_i(J_k) = J_{\rho(k)} \cap K_i$.*

Proof. Suppose that $U_1, \cdots, U_l$ are connected components of $K\backslash\{p\}$. Then, we may choose n so that U_j is not contained in ${F_i}^n(K)$ for all $j = 1, \cdots, l$. By Proposition 1.3.5-(2), $U_j \cap {F_i}^n(V_0) \ne \emptyset$ for any $j = 1, \cdots, l$. Therefore $l \le \#(V_0)$. This implies that the number of connected components of $K\backslash\{p\}$ is finite.

Now, let $J_1, \cdots, J_m$ be the collection of all connected components of $K\backslash\{p\}$. Note that $F_i(J_j)$ is connected and $\cup_{j=1}^m F_i(J_j) = K_i\backslash\{p\}$. Therefore, there exits $\rho(j)$ such that $F_i(J_j) \subset J_{\rho(j)}$. Since $J_k \cap K_i \ne \emptyset$ for any k, we may find j that satisfies $J_j \cap {F_i}^{-1}(J_k) \ne \emptyset$. This implies that ρ is a permutation of $\{1, 2, \cdots, m\}$. As $K_i\backslash\{p\} = \cup_{j=1}^m (K_i \cap J_i)$, we see that $K_i \cap J_{\rho(j)} = F_i(J_j)$ for any j.

Next choose n so that J_j is not contained in ${F_i}^n(K)$ for any j. Then, it follows that $J_{\rho^n(k)} \cap {F_i}^n(V_0) \ne \emptyset$ for any k. Since $J_{\rho^n(k)} \cap {F_i}^n(V_0) = {F_i}^n(V_0 \cap J_k)$, we see that $J_k \cap V_0 \ne \emptyset$ for any k. □

Next proposition also concerns a connected p. c. f. self-similar structure. It gives an sufficient condition for $K\backslash V_0$ being connected.

Proposition 1.6.6. *Let $(K, S, \{F_i\}_{i\in S})$ be a connected post critically finite self-similar structure. Assume that, for any $p, q \in V_0$, there exists a homeomorphism $g : K \to K$ such that $g(V_0) = V_0$ and $g(p) = q$. Then $K\backslash V_0$ is connected.*

If a connected p. c. f. self-similar structure satisfies the assumption of the above proposition, we say that the self-similar structure is weakly symmetric.

To prove the above proposition, we need the following lemmas.

Lemma 1.6.7. *Assume the conditions in Proposition 1.6.6. Let J be a connected component of $K\backslash V_0$. Then $\#(\overline{J} \cap V_0) \ge 2$.*

Proof. Suppose that there exists a connected component of $K\backslash V_0$ satisfying $\#(\overline{J}\cap V_0) = 1$. Let p_0 be the unique $p_0 \in V_0$ with $p_0 \in \overline{J}$. Then by applying g_{xy} for $x \neq y \in V_0$, we see that, for any $p \in V_0$, there exists a connected component of J_p of $K\backslash V_0$ such that $\overline{J}_p \cap V_0 = \{p\}$. Then it follows that J_p is a connected component of $K\backslash\{p\}$.

Now, since $\mathcal{L}$ is post critically finite, there exits $p \in V_0$ such that p is a fixed point of F_w for some $w \in W_*\backslash W_0$. By exchanging the self-similar structure $\mathcal{L}$ with $\mathcal{L}_m = (K, W_m, \{F_v\}_{v\in W_m})$, we can use Proposition 1.6.5 and obtain that $J_p \cap V_0 \neq \emptyset$. This contradicts to the fact that $\overline{J}_p \cap V_0 = \{p\}$. □

Lemma 1.6.8. *For $p \in V_0$, Define*

$$k(p, V_0) = \#\{C : C \text{ is a connected component of } K\backslash V_0, p \in \overline{C}\}$$

and $k(p)$ = the number of connected components of $K\backslash\{p\}$. Then $k(p, V_0) = k(p)$.

Proof. First we show that $k(p, V_0)$ is finite. Let $\{C_i\}_{i\geq 1}$ be the connected components of $K\backslash V_0$ with $p \in \overline{C}_i$. Suppose that $C_i \neq C_j$ if $i \neq j$. Set $\alpha = \min\{|p - q| : q \in V_0, q \neq p\}/2$. Then by Lemma 1.6.7, there exists $x_i \in C_i$ such that $|x_i - p| = \alpha$. Since K is compact, there exists a subsequence $\{x_{i_k}\}_{k\geq 1}$ that converges to some $x \in K$ as $k \to \infty$. Note that $x \notin V_0$. Therefore, $K_{l,x} = \cup_{w\in W_l, x\in K_w} K_w$ is contained in $K\backslash V_0$ for sufficiently large l. Moreover $K_{l,x}$ is connected because K_w is connected for any $w \in W_*$. Hence $K_{l,x}$ is a subset of a connected component of $K\backslash V_0$. Since $K_{l,x}$ is a neighborhood of x, it follows that $x_{i_k} \in K_{l,x}$ for sufficiently large k. Hence $C_{i(k)}$ equals to the connected component of $K\backslash V_0$ containing x for sufficiently large k. This contradicts to the fact that $C_i \neq C_j$ if $i \neq j$. Thus we have show that $k(p, V_0)$ is finite.

Since $\mathcal{L}$ is an affine nested self-similar structure, we see that $k(p)$ and $k(p, V_0)$ is independent of the choice of $p \in V_0$. Hence, as in the proof of the last lemma, we may assume that $F_w(p) = p$ for some $w \in W_*\backslash W_0$.

Now, let $k = k(p, V_0)$ and let $\{C_i\}_{i=1,\cdots,k}$ be the collection of all the connected components of $K\backslash V_0$ whose closure contains p. Let $U = (\cup_{i=1,2,\cdots,k} C_k) \cup \{p\}$. Then U is a neighborhood of p. Hence, if $w(n) = \underbrace{w \cdots w}_{n \text{ times}}$, then $K_{w(n)} \subset U$ for sufficiently large n. Therefore if k' is the number of connected components of $K_{w(n)}\backslash\{p\}$, then $k' \geq k$. Since $k(p) = k'$, we see that $k(p) \geq k(p, V_0)$.

On the other hand, a connected component of $K\backslash\{p\}$ contains at least one C_i. Hence $k(p) \leq k(p, V_0)$. □

Proof of Proposition 1.6.6. Let $\mathcal{J}$ be the collection of connected components of $K\backslash V_0$. Define $V = V_0 \cup \mathcal{J}$ and $E = \{(p, J) : p \in V_0, J \in \mathcal{J}, p \in \overline{J}\}$. Let $G = (V, E)$ be the non-directed graph, where V is the set of vertices and E is the set of edges.

First we show that this graph G does not contain any loop. Suppose that there exists a loop in G : there exist $\{p_i\}_{i=1,\cdots,n} \subset V_0$ and $\{J_i\}_{i=1,2,\cdots,n} \subset \mathcal{J}$

such that $(p_i, J_i), (p_{i+1}, J_i) \in E$ for $i = 1, 2, \cdots, n$, where $p_{n+1} = p_n$. Then J_i and J_{i+1} are connected components of $K \backslash V_0$ whose closures contain p_i. On the other hand, J_i and J_{i+1} are contained in the same connected component of $K \backslash \{p_i\}$. This contradicts to Lemma 1.6.8.

Since G does not contain any loop, G is a tree : for any $x, y \in V$, there exists a unique sequence of edges from x to y. Hence G has an end point. Namely, there exists $x \in V$ such that $\#\{y \in V : (x, y) \in E \text{ or } (y, x) \in E.\} = 1$. By Lemma 1.6.7, we see that $x \in V_0$. Hence $k(x, V_0) = 1$. Therefore, $k(p, V_0) = k(x, V_0) = 1$ for any $p \in V_0$. On the other hand, if $\#\mathcal{J} \geq 2$, then there exists $p \in V_0$ such that $k(p, V_0) \geq 2$. Hence we see that $\#\mathcal{J} = 1$. $\square$

Exercise

Exercise 1.1. Let $f : \mathbf{R}^n \to \mathbf{R}^n$ be a similitude with a Lipschitz constant r. Show that there exist $a \in \mathbf{R}^n$ and $U \in O(n)$ such that $f(x) = rUx + a$ for all $x \in \mathbf{R}^n$.

(Hint: If $g(x) = (f(x) - f(0))/r$, one can see that $|g(x) - g(y)| = |x - y|$. This may imply that the natural inner product of $\mathbf{R}^n$ is invariant under g. Also one should show that g is a linear map.)

Exercise 1.2. Let (X, d) be a complete metric space and let $f_i : X \to X$ be a contraction for $i = 1, 2, \cdots, N$. For $A \subseteq X$, define $F(A) = \cup_{1 \leq i \leq N} f_i(A)$. Let K be the self-similar set with respect to $\{f_1, f_2, \cdots, f_N\}$. Then

(1) Suppose $A \neq \emptyset$. Show that $A \supseteq F(A)$ implies $\overline{A} \supseteq K$.

(2) Show that for any $x \in X$, $B_r(x) \supseteq F(B_r(x))$ for sufficiently large r.

Exercise 1.3. Define $F_i(z) = \frac{1}{2}(z - p_i) + p_i$ for $i \in \{1, 2, 3, 4\}$, where $p_1 = 0, p_2 = 1, p_3 = (1 + \sqrt{-1})$ and $p_4 = \sqrt{-1}$. Let K be the self-similar set with respect to $\{F_1, F_2, F_3, F_4\}$. Prove that V_0 coincides with the topological boundary of K.

Exercise 1.4. Let $K = [0, 1]$ and let $S = \{1, 2, \cdots, N\}$. Set $F_i(x) = a_i x + b_i$ for $i \in S$. Assume that $0 < a_i < 1$ for any $i \in S$ and that $K = \cup_{i \in S} F_i(K)$. Prove that $(K, S, \{F_i\}_{i \in S})$ is minimal if and only if $\sum_{i=1}^N a_i = 1$.

Exercise 1.5. Define $f_1(x) = x/3$ and $f_2(x) = x/3 + 2/3$. Let K be the self-similar set with respect to $\{f_1, f_2\}$. (K is the Cantor's middle third set.) Set $g_i = f_i \circ f_i$ for $i = 1, 2$. Let K' be the self-similar set with respect to $\{g_1, g_2\}$. The natural map from $\Sigma(\{1, 2\}) \to K$ (resp. $\Sigma(S) \to K'$) is denoted by π (resp. π'). Note that both π and π' are homeomorphism. Set $f_3 = f_1 \circ \pi' \circ \pi^{-1}$.

(1) Show that f_3 is a contraction on K.

(2) Let $\mathcal{L} = (K, \{1, 2, 3\}, \{f_1, f_2, f_3\})$. Show that $\text{int}(\mathcal{C}_{\mathcal{L}}) \neq \emptyset$ and $\text{int}(C_{\mathcal{L}}) = \emptyset$.

Exercise 1.6. Prove that $\mathcal{P}_{\mathcal{L}(\Lambda)} = \mathcal{P}_{\mathcal{L}}$ for any partition Λ for the self-similar structures corresponding to Example 1.2.7, 1.2.8, 1.2.9 in the last section.

Exercise 1.7. Let $S = \{1, 2, 3\}$. Set $\omega \sim \tau$ if and only if $\{\omega, \tau\} \subseteq \{w121\dot{2}, w3\dot{1}\}$ for some $w \in W_*(S)$ or $\omega = \tau$.

(1) Let $K = \Sigma(S)/\sim$ with the quotient topology. Also define $F_i : K \to K$ by $F_i(x) = \pi(\sigma_i(\pi^{-1}(x)))$ for $x \in K$. Then prove that $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$ is a self-similar structure.

(2) Let $\Lambda = \{1, 21, 22, 23, 3\}$. Prove that $\mathcal{P}_{\mathcal{L}(\Lambda)}$ is a proper subset of $\mathcal{P}_{\mathcal{L}}$.

Exercise 1.8. Let $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$ is a self-similar structure and let Λ be a partition of $\Sigma(S)$. Show that $\mathcal{L}$ is post critically finite if and only if $\mathcal{L}(\Lambda)$ is post critically finite.

Exercise 1.9. Evaluate the Hausdorff dimensions of the self-similar sets introduced in Examples 1.2.6–1.2.9 under Euclidean metrics.

Chapter 2

Analysis on Limits of Networks

In this chapter, we will discuss limits of discrete Laplacians (or equivalently Dirichlet forms) on a increasing sequence of finite sets. The results in this chapter will play a fundamental role in constructing a Laplacian (or equivalently a Dirichlet form) on certain self-similar set in the next chapter, where we will approximate a self-similar set by a increasing sequence of finite sets and then construct a Laplacian on the self-similar set by taking a limit of Laplacians on the finite sets.

More precisely, we will define a Dirichlet form and a Laplacian on a finite set in §2.1. The key idea is that every Dirichlet form on a finite set can be associated with a electrical network consisting of resistors. From such a point of view, we will introduce an important notion of effective resistance. In §2.2, we will study a limit of a "compatible" sequence of Dirichlet forms on increasing finite sets. Roughly speaking, the word "compatible" means that Dirichlet forms appearing in the sequence induce the same effective resistance on the union of the increasing finite sets. In §2.3 and §2.4, we will present further properties of limits of compatible sequences of Dirichlet forms.

§2.1 Dirichlet forms and Laplacians on a finite set

In this section, we give fundamental notions of analysis on a finite set, namely, Dirichlet forms, Laplacians and effective resistance.

Notation. For a set V, we define $\ell(V) = \{f : f : V \to \mathbf{R}\}$. If V is a finite set, $\ell(V)$ is thought to be equipped with a standard inner product $(\cdot,\cdot)$ defined by $(u,v) = \sum_{p\in V} u(p)v(p)$ for any $u, v \in \ell(V)$.

First we give a definition of Dirichlet forms on a finite set V. In §A.4, one can find a definition of Dirichlet forms for general locally compact metric space.

Definition 2.1.1 (Dirichlet forms). Let V be a finite set. A symmetric bilinear form on $\ell(V)$, $\mathcal{E}$ is called a Dirichlet form on V if it satisfies

(DF1) $\mathcal{E}(u,u) \geq 0$ for any $u \in \ell(V)$,

(DF2) $\mathcal{E}(u,u) = 0$ if and only if u is a constant on V

and

(DF3) For any $u \in \ell(V)$, $\mathcal{E}(u,u) \geq \mathcal{E}(\bar{u},\bar{u})$, where $\bar{u}$ is defined by

$$\bar{u}(p) = \begin{cases} 1 & \text{if } u(p) \geq 1, \\ u(p) & \text{if } 0 < u(p) < 1, \\ 0 & \text{if } u(p) \leq 0. \end{cases}$$

We use $\mathcal{DF}(V)$ to denote the collection of Dirichlet forms on V. Also we define

$$\widetilde{\mathcal{DF}}(V) = \{\mathcal{E} : \mathcal{E} \text{ is a symmetric bilinear form on } \ell(V) \text{ with (DF1) and (DF2)}\}.$$

Condition (DF3) is called the Markov property. Obviously $\mathcal{DF}(V) \subset \widetilde{\mathcal{DF}}(V)$.

This definition is a special case of Definition A.4.2 when X is a finite set V and the measure μ is the discrete measure on V.

Notation. Let V be a finite set. The characteristic function χ_U of a subset $U \subseteq V$ is defined by $\chi_U(q) = \begin{cases} 1 & \text{if } q \in U, \\ 0 & \text{otherwise.} \end{cases}$ If $U = \{p\}$ for a point $p \in V$, we write χ_p instead of $\chi_{\{p\}}$. If $H : \ell(V) \to \ell(V)$ is a linear map, then we set $H_{pq} = (H\chi_q)(p)$ for $p, q \in V$. For $f \in \ell(V)$, $(Hf)(p) = \sum_{q \in V} H_{pq} f(q)$.

Definition 2.1.2 (Laplacians). A symmetric linear operator $H : \ell(V) \to \ell(V)$ is called a Laplacian on V if it satisfies

(L1) H is non-positive definite,

(L2) $Hu = 0$ if and only if u is a constant on V,

and

(L3) $H_{pq} \geq 0$ for all $p \neq q \in V$.

We use $\mathcal{L}(V)$ to denote the collection of Laplacians on V. Also

$$\widetilde{\mathcal{L}}(V) = \{H : H : \ell(V) \to \ell(V) \text{ is symmetric and linear with (L1) and (L2)}\}.$$

Obviously $\mathcal{L}(V) \subset \widetilde{\mathcal{L}}(V)$.

There is a natural correspondence between $\mathcal{DF}(V)$ and $\mathcal{L}(V)$. For a symmetric linear operator $H : \ell(V) \to \ell(V)$, we can define a symmetric quadratic form $\mathcal{E}_H(\cdot,\cdot)$ on $\ell(V)$ by $\mathcal{E}_H(u,v) = -(u, Hv)$ for $u, v \in \ell(V)$. If we write $\pi(H) = \mathcal{E}_H$, it is easy to see that π is a bijective mapping between symmetric liner operators and symmetric quadratic forms.

This correspondence between Dirichlet forms and non-negative symmetric operators is a special case of the correspondence described in Theorem A.4.4.

Proposition 2.1.3. *π is a bijective mapping between $\widetilde{\mathcal{L}}(V)$ and $\widetilde{\mathcal{DF}}(V)$. Moreover, $\pi(\mathcal{L}(V)) = \mathcal{DF}(V)$.*

Proof. It is routine to show $\pi(\widetilde{\mathcal{L}}(V)) = \widetilde{\mathcal{DF}}(V)$. To show $\pi(\mathcal{L}(V)) = \mathcal{DF}(V)$, first note that $\mathcal{E}_H(u,u) = \frac{1}{2}\sum_{p,q\in V} H_{pq}(u(p)-u(q))^2$. By this expression, it is easy to see that $\pi(\mathcal{L}(V)) \subseteq \mathcal{DF}(V)$. Now suppose $H \in \widetilde{\mathcal{L}}(V)\backslash\mathcal{L}(V)$. So there exist $p \neq q \in V$ with $H_{pq} < 0$. We can assume that $H_{pq} = -1$ without loss of generality. Set $u(p) = x, u(q) = y$ and $u(a) = z$ for all $a \in V\backslash\{p,q\}$. Then we have $\mathcal{E}_H(u,u) = \alpha(x-z)^2 + \beta(y-z)^2 - (x-y)^2$. As $\mathcal{E}_H$ is non-negative definite, α and β should be non-negative. If $x = 1$, $z = 0$ and $y < 0$, then $\mathcal{E}_H(u,u) = \alpha - 1 + 2y + (\beta-1)y^2$ and $\mathcal{E}_H(\bar{u},\bar{u}) = \alpha - 1$. If $|y|$ is small, we have $\mathcal{E}_H(u,u) < \mathcal{E}_H(\bar{u},\bar{u})$. Hence $\mathcal{E}_H \notin \mathcal{DF}(V)$. This shows that $\pi(H) \in \mathcal{DF}(V)$ if and only if $H \in \mathcal{L}(V)$. □

Example 2.1.4. Let V be a set with three elements, say, p_1, p_2, p_3. Set $H = \begin{pmatrix} -(1+\epsilon) & 1 & \epsilon \\ 1 & -2 & 1 \\ \epsilon & 1 & -(1+\epsilon) \end{pmatrix}$. Then $\mathcal{E}_H(u,u) = (x-y)^2 + (y-z)^2 + \epsilon(x-z)^2$, where $x = u(p_1), y = u(p_2)$ and $z = u(p_3)$. Letting $X = x-y$ and $Y = y-z$, we have

$$\begin{aligned}\mathcal{E}_H(u,u) &= X^2 + Y^2 + \epsilon(X+Y)^2 \\ &= (1+2\epsilon)(X^2+Y^2) - \epsilon(X-Y)^2\end{aligned}$$

So it is clear that if $\epsilon > -\frac{1}{2}$, then $H \in \widetilde{\mathcal{L}}(V)$ and if $\epsilon \geq 0$, then $H \in \mathcal{L}(V)$.

If V is a finite set and H is a Laplacian on V, the pair (V,H) is called a resistance network(an r-network, for short). In fact, we can relate an r-network to an actual electrical network as follows. For an r-network (V,H), we will attach a resistor of resistance $r_{pq} = {H_{pq}}^{-1}$ to the terminals p and q for $p,q \in V$. Also a plus-side of a battery is connected to every terminal p while its minus-side is grounded so that we can put any electrical potential on each terminal. For a given electric potential $v \in \ell(V)$, the current i_{pq} between p and q is given by $i_{pq} = H_{pq}(v(p)-v(q))$. So the total current $i(p)$ from a terminal p to the ground is obtained by $i(p) = (Hv)(p)$.

Let (V,H) be an r-network and let U be a proper subset of V. We next discuss what is the proper way of restricting H onto U from analytical point of view.

Lemma 2.1.5. *Let V be a finite set and let U be a proper subset of V. For $H \in \widetilde{\mathcal{L}}(V)$, we define $T_U : \ell(U) \to \ell(U), J_U : \ell(U) \to \ell(V\backslash U)$ and $X_U : \ell(V\backslash U) \to \ell(V\backslash U)$ by*

$$H = \begin{pmatrix} T_U & {}^tJ_U \\ J_U & X_U \end{pmatrix},$$

where ${}^tJ_U$ is the transpose matrix of J_U. (When no confusion may occur, we use T, J and X instead of T_U, J_U and X_U.) Then, $X = X_U$ is negative definite *and*

$$\mathcal{E}_H(u,u) = \mathcal{E}_X(u_1 + X^{-1}Ju_0, u_1 + X^{-1}Ju_0) + \mathcal{E}_{T-{}^tJX^{-1}J}(u_0,u_0), \qquad (2.1.1)$$

where $u_0 = u|_U$ *and* $u_1 = u|_{V\backslash U}$ *for* $u \in \ell(V)$.

Proof. For $v \in \ell(V\backslash U)$, define $\tilde{v}$ by $\tilde{v}|_U = 0$ and $\tilde{v}|_{V\backslash U} = v$. Then $\mathcal{E}_X(v,v) = \mathcal{E}_H(\tilde{v},\tilde{v}) \geq 0$. By (L2), we see that if $\mathcal{E}_X(v,v) = 0$ then $\tilde{v}$ should be a constant on V. This implies $v = 0$. Hence $\mathcal{E}_X$ is positive definite. By the definition of $\mathcal{E}_X$, X is negative definite. Now (2.1.1) can be obtained by an easy calculation. □

Theorem 2.1.6. *Assume the same situation as in Lemma 2.1.5. For given* $u \in \ell(U)$*, define* $h(u) \in \ell(V)$ *by* $h(u)|_U = u$ *and* $h(u)|_{V\backslash U} = -X^{-1}Ju$*. Then* $h(u)$ *is the unique element that attains* $\min_{v\in\ell(V),v|_U=u} \mathcal{E}_H(v,v)$*. Also define* $P_{V,U}(H) = T - {}^tJX^{-1}J$*. Then,* $P_{V,U} : \widetilde{\mathcal{L}}(V) \to \widetilde{\mathcal{L}}(U)$ *and*

$$\mathcal{E}_{P_{V,U}(H)}(u,u) = \mathcal{E}_H(h(u),h(u)) = \min_{v\in\ell(V),v|_U=u} \mathcal{E}_H(v,v) \qquad (2.1.2)$$

Moreover, if $H \in \mathcal{L}(V)$*, then* $P_{V,U}(H) \in \mathcal{L}(U)$*.*

Proof. By (2.1.1), $\min_{v\in\ell(V),v|_U=u} \mathcal{E}_H(v,v)$ is attained if and only if $v|_{V\backslash U} + X^{-1}Ju = 0$. Hence we have the first part of the theorem.

Next we show that $P_{V,U}(H) = T - {}^tJX^{-1}J \in \widetilde{\mathcal{L}}(V)$. By (2.1.1), we can verify (2.1.2). Hence, $\mathcal{E}_{P_{V,U}(H)}$ is non-negative definite. By (2.1.2), $\mathcal{E}_{P_{V,U}(H)}(u,u) = 0$ implies that $h(u)$ is a constant on V and therefore u is a constant on U. Thus we can show that $P_{V,U}(H) \in \widetilde{\mathcal{L}}(U)$.

Finally, if $H \in \mathcal{L}(V)$, we have $\mathcal{E}_{P_{V,U}(H)}(u,u) = \mathcal{E}_H(h(u),h(u)) \geq \mathcal{E}_H(\overline{h(u)},\overline{h(u)})$. As $\overline{h(u)}|_U = \bar{u}$, we obtain $\mathcal{E}_H(\overline{h(u)},\overline{h(u)}) \geq \mathcal{E}_{P_{V,U}(H)}(\bar{u},\bar{u})$. Hence $\mathcal{E}_{P_{V,U}(H)}$ has the Markov property. By Proposition 2.1.3, $P_{V,U}(H) \in \mathcal{L}(U)$. □

The linear operator $P_{V,U}(H)$ is thought of as the proper restriction of H onto U for the viewpoint of electrical circuits. In fact, V and $P_{V,U}$ give exactly the same effective resistance (which will be defined in Definition 2.1.9) on U. When no confusion may occur, we write $[H]_U$ in place of $P_{V,U}(H)$.

Remark. In general, $P_{V,U}$ is not injective. For example, set $V = \{p_1,p_2,p_3\}$ and $U = \{p_1,p_2\}$. If $H_\epsilon = \begin{pmatrix} -(1+\epsilon) & 1 & \epsilon \\ 1 & -1 & 0 \\ \epsilon & 0 & -\epsilon \end{pmatrix}$ for $\epsilon > 0$, then $H_\epsilon \in \mathcal{L}(V)$ and $[H_\epsilon]_U = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$

Note that $h(u)$ is the unique solution of $(Hv)|_{V\backslash U} = 0$ and $v|_U = u$. Therefore if we think U as a boundary of V, $h(u)$ may be called the harmonic function with a boundary value $u \in \ell(U)$. For a Laplacian H on V, we have the following maximum principle for harmonic functions.

Proposition 2.1.7 (Maximum Principle). *Let V be a finite set and let $H \in \mathcal{L}(V)$. Also let U be a subset of V. For $p \in V \backslash U$, set*

$$U_p = \{q \in U : \text{There exist } p_1, p_2, \cdots, p_m \in V \backslash U \text{ with } p_1 = p \text{ such that } H_{p_i p_{i+1}} > 0 \text{ for } i = 1, 2, \cdots, m-1 \text{ and } H_{p_m q} > 0.\}.$$

Then if $(Hu)|_{V \backslash U} = 0$,

$$\min_{q \in U_p} u(q) \leq u(p) \leq \max_{q \in U_p} u(q)$$

for any $p \in V \backslash U$. Moreover, $u(p) = \max_{q \in U_p} u(q)$ (or $u(p) = \min_{q \in U_p} u(q)$) if and only if u is constant on U_p.

Proof. For $p \in V$, set $N_p = \{q : H_{pq} > 0\}$. Also define

$$W_p = \{q \in V \backslash U : \text{There exist } p_1, p_2, \cdots, p_m \in V \backslash U \text{ with } p_1 = p \text{ and } p_m = q \text{ such that } H_{p_i p_{i+1}} > 0 \text{ for } i = 1, 2, \cdots, m-1.\}$$

and $V_p = W_p \cup U_p$. First assume $u(p_*) = \max_{q \in V_p} u(q)$ for $p_* \in W_p$. Then $N_{p_*} \subseteq V_p$ and $(Hu)(p_*) = \sum_{q \in N_{p_*}} H_{p_* q}(u(q) - u(p_*)) = 0$. Since $H_{p_* q} > 0$ and $u(q) - u(p_*) \leq 0$ for any $q \in N_{p_*}$, we have $u(q) = u(p_*)$ for all $q \in N_{p_*}$. Iterating this argument, we see that u is constant on V_p. Using the same discussion, it follows that if there exists $p_* \in W_p$ such that $u(p_*) = \min_{q \in V_p} u(q)$, then u is constant on V_p. Hence,

$$\min_{q \in U_p} u(q) = \min_{q \in V_p} u(q) \leq u(p) \leq \max_{q \in V_p} u(q) = \max_{q \in U_p} u(q).$$

The rest of the statement is now obvious. □

The following corollary of the maximum principle is called the Harnack inequality.

Corollary 2.1.8 (Harnack inequality). *Let V be a finite set and let $H \in \mathcal{L}(V)$. Also let U be a subset of V. Assume that $A \subseteq V \backslash U$ and that $V_p = V_q$ for any $p, q \in A$. Then there exists a positive constant c such that*

$$\max_{p \in A} u(p) \leq c \min_{p \in A} u(p)$$

for any non-negative $u \in \ell(V)$ with $(Hu)|_{V \backslash U} = 0$. The above inequality is called the Harnack inequality.

Proof. Let $V' = V_p$ for some $p \in A$. (Note that V' is independent of a choice of $p \in A$.) Set $\mathcal{A} = \{u : (Hu)_{V \backslash U} = 0, \min_{p \in V} u(p) \geq 0, \max_{p \in V'} u(p) = 1\}$. By Proposition2.1.7, we see that $\min_{p \in A} u(p) > 0$ for $u \in \mathcal{A}$. If $\mathcal{A}_0 = \{u|_{V'} : u \in \mathcal{A}\}$, then $\mathcal{A}_0$ is a compact subset of $\ell(V')$. Therefore, $c = \inf\{\min_{p \in A} u(p) : u \in \mathcal{A}_0\} > 0$. By the definition of $\mathcal{A}_0$, it follows that $c = \inf\{\min_{p \in A} u(p) : u \in \mathcal{A}\}$. This immediately implies the Harnack inequality. □

Next, we define effective resistances associated with a Laplacian or, equivalently, a Dirichlet form. From the viewpoint of electrical circuits, the effective resistance between two terminals is an actual resistance considering all the resistors in the circuit.

Definition 2.1.9 (effective resistance). Let V be a finite set and let $H \in \widetilde{\mathcal{L}}(V)$. For $p \neq q \in V$, we define

$$R_H(p,q) = \big(\min\{\mathcal{E}_H(u,u) : u \in \ell(V), u(p) = 1, u(q) = 0\}\big)^{-1} \qquad (2.1.3)$$

Also we define $R_H(p,p) = 0$ for all $p \in V$. $R_H(p,q)$ is called the effective resistance between p and q with respect to H.

By Theorem 2.1.6, if $U = \{p,q\}$, then it follows that

$$[H]_U = \frac{1}{R_H(p,q)} \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}. \qquad (2.1.4)$$

Definition 2.1.10. Let V_i be a finite set and let $H_i \in \widetilde{\mathcal{L}}(V_i)$ for $i = 1,2$. We write $(V_1, H_1) \leq (V_2, H_2)$ if and only if $V_1 \subseteq V_2$ and $P_{V_2,V_1}(H_2) = H_1$.

The next proposition is obvious by the above definitions.

Proposition 2.1.11. *Let V_i be a finite set and let $H_i \in \widetilde{\mathcal{L}}(V_i)$ for $i = 1,2$. If $(V_1, H_1) \leq (V_2, H_2)$, then $R_{H_1}(p,q) = R_{H_2}(p,q)$ for any $p, q \in V_1$.*

In fact, the converse of the above proposition is also true if both H_1 and H_2 satisfies (L3). This fact is a corollary of the following theorem, which says that a Laplacian is completely determined by associated effective resistances.

Theorem 2.1.12. *Let V be a finite set. Suppose $H_1, H_2 \in \mathcal{L}(V)$. Then $H_1 = H_2$ if and only if $R_{H_1}(p,q) = R_{H_2}(p,q)$ for any $p, q \in V$.*

Proof. We need to show the "if" part. We use an induction on $\#(V)$. When $\#(V) = 2$, the theorem follows immediately by (2.1.4). Now suppose the statement holds if $\#(V) < n$. Let $V = \{p_1, p_2, \cdots, p_n\}$. We write $h_{ij} = (H_1)_{p_i p_j}$ and $H_{ij} = (H_2)_{p_i p_j}$. Also let $V_i = V \backslash \{p_i\}$ and let

$$D_k^i = T_k^i - {}^t J_k^i (X_k^i)^{-1} J_k^i$$

for $k = 1,2$, where $T_k^i : \ell(V_i) \to \ell(V_i)$, $J_k^i : \ell(V_i) \to \ell(\{p_i\})$ and $X_k^i : \ell(\{p_i\}) \to \ell(\{p_i\})$ are defined by

$$H_k = \begin{pmatrix} T_k^i & {}^t J_k^i \\ J_k^i & X_k^i \end{pmatrix}.$$

As $(V_i, D_k^i) \leq (V, H_k)$, we have $R_{D_1^i}(p,q) = R_{D_2^i}(p,q)$ for all $p, q \in V_i$. By the induction hypothesis, $D_1^i = D_2^i$. Now define $D^i = D_1^i = D_2^i$ and $d_{kl}^i = (D^i)_{p_k p_l}$.

Calculating directly and then using the fact that $h_{ik} = h_{ki}$ and $H_{ik} = H_{ki}$, we obtain

$$d^i_{kl} = h_{kl} - h_{ik}h_{il}/h_{ii} = H_{kl} - H_{ik}H_{il}/H_{ii}.$$

In particular,

$$d^i_{kk} = h_{kk} - h^2_{ik}/h_{ii} = H_{kk} - H^2_{ik}/H_{ii}. \tag{2.1.5}$$

Exchanging k and i, we can show that $d^k_{ii}/d^i_{kk} = h_{ii}/h_{kk} = H_{ii}/H_{kk}$. Therefore, there exists $t > 0$ such that $H_{ii} = t\,h_{ii}$ for $i = 1, 2, \cdots, N$. Again by (2.1.5), we have

$$(h_{ik})^2 = h_{kk}h_{ii} - d^i_{kk}h_{ii}$$

and

$$(H_{ik})^2 = H_{kk}H_{ii} - d^i_{kk}H_{ii} = t^2h_{kk}h_{ii} - td^i_{kk}h_{ii}.$$

As $-h_{kk} = \sum_{i:i\neq k} h_{ik}$ and $-th_{kk} = -H_{kk} = \sum_{i:i\neq k} H_{ik}$, we have

$$-h_{kk} = \sum_{i:i\neq k} \sqrt{h_{kk}h_{ii} - d^i_{kk}h_{ii}} = \sum_{i:i\neq k} \sqrt{h_{kk}h_{ii} - d^i_{kk}h_{ii}/t}.$$

As a function of t, the right-hand side of the above equation is monotonically increasing. Hence the above equality holds only for $t = 1$. Therefore we obtain $H_1 = H_2$. □

Corollary 2.1.13. *Let V_i be a finite set and let $H_i \in \mathcal{L}(V_i)$ for $i = 1, 2$. Then $(V_1, H_1) \leq (V_2, H_2)$ if and only if $R_{H_1}(p, q) = R_{H_2}(p, q)$ for any $p, q \in V$.*

Remark. It is reasonable to expect that Theorem 2.1.12 remains true even if we only assume $H_1, H_2 \in \widetilde{\mathcal{L}}(V)$. However, the above proof cannot be extended to such a case, because it uses the fact that $H_{pq} \geq 0$. Unfortunately, we don't know whether such an extension is true or not.

One reason why effective resistance is important is that it becomes a metric on V if $H \in \mathcal{L}(V)$. This metric called the effective resistance metric will play a crucial roll in the theory of Laplacians and Dirichlet forms on (post critically finite) self-similar sets.

Theorem 2.1.14. *Let V be a finite set and let $H \in \mathcal{L}(V)$. Then $R_H(\cdot, \cdot)$ is a metric on V. This metric R_H is called the effective resistance metric on V associated with H.*

Remark. Not every metric on a finite set V corresponds to an effective resistance metric with respect to a Laplacian $H \in \mathcal{L}(V)$. See Exercise 2.1 and Exercise 2.2.

We need the following well-known formula about electrical network to show Theorem 2.1.14.

Lemma 2.1.15 (Δ-Y transform). *Let $U = \{p_1, p_2, p_3\}$ and let $V = \{p_0\} \cup U$. Set $R_{ij} = H_{p_ip_j}^{-1}$ for $H \in \mathcal{L}(U)$, where we assume that $H_{p_i,p_j} > 0$. Define*

$$R_1 = \frac{R_{12}R_{31}}{R_{12} + R_{23} + R_{31}}, R_2 = \frac{R_{23}R_{12}}{R_{12} + R_{23} + R_{31}}, R_3 = \frac{R_{31}R_{23}}{R_{12} + R_{23} + R_{31}}.$$

If $H' \in \mathcal{L}(V)$ is defined by

$$H'_{p_ip_j} = \begin{cases} R_j^{-1} & \textit{if } i = 0, \\ 0 & \textit{othewise}, \end{cases}$$

for $i < j$, then $[H']_U = H$.

A direct calculation shows this formula.

As we mentioned before, we can associate an actual electrical circuit to a Laplacian. In the above lemma, the circuit associated with $H \in \mathcal{L}(U)$ has three terminals $\{p_1, p_2, p_3\}$ and the terminals p_i and p_j are connected by a resistor of resistance R_{ij}. Let us call this circuit a Δ-circuit, which reflects the triangular shape of the circuit. At the same time,the circuit associated with $H' \in \mathcal{L}(V)$ consists of four terminals $\{p_0, p_1, p_2, p_3\}$ and each terminal p_i is only connected to p_0 by a resistor of resistance R_i for $i = 1, 2, 3$. p_0 is a kind of a focal point of the circuit. Let us call this circuit a Y-circuit because of it "upside-down Y" shape. The Δ-Y transform says that the Δ-circuit and the Y-circuit are equivalent to each other as electrical networks.

Proof of Theorem 2.1.14. By Definition 2.1.9 and (2.1.4), it follows that $R_H(p, q) \geq 0$ and that $R_H(p, q) = 0$ if and only if $p = q$. Next we should show the triangle inequality. We may assume that $\#(V) \geq 3$. For $U = \{p_1, p_2, p_3\} \subset V$, let $H' = [H]_U$. By Proposition 2.1.11, we have $R_{H'}(p_i, p_j) = R_H(p_i, p_j)$.

First assume that $H'_{p_mp_n} > 0$ for any $m \neq n$. Then the Δ-Y transform shows

$$R_{H'}(p_i, p_j) = \frac{R_{ij}(R_{ik} + R_{kj})}{R_{12} + R_{23} + R_{31}}, \tag{2.1.6}$$

where $R_{mn} = (H'_{p_mp_n})^{-1}$ and $\{i, j, k\} = \{1, 2, 3\}$. Hence we can easily see that

$$R_H(p_1, p_2) + R_H(p_2, p_3) \geq R_H(p_1, p_3).$$

Next, if one of $H'_{p_mp_n} = 0$, say $H'_{p_1p_3} = 0$, then $R_{H'}(p_1, p_2) = R_{12}$, $R_{H'}(p_2, p_3) = R_{23}$ and $R_{H'}(p_1, p_3) = R_{12} + R_{23}$, where $R_{ij} = (H'_{p_ip_j})^{-1}$. So we can verify the triangle inequality. □

$R_H(\cdot, \cdot)$ is not a metric on V for general $H \in \widetilde{\mathcal{L}}(V)$. In fact, if $\#(V) > 3$, there exists $H \notin \mathcal{L}(V)$ such that $R_H(\cdot, \cdot)$ is not a metric on V. (See Exercise 2.4 and Exercise 2.5.) As we will see, however, $\sqrt{R_H(\cdot, \cdot)}$ always becomes a metric on V for all $H \in \widetilde{\mathcal{L}}(V)$.

The following is an alternative expression of the effective resistance.

Proposition 2.1.16. *Let V be a finite set and let $H \in \widetilde{\mathcal{L}}(V)$. Then for any $p, q \in V$,*

$$R_H(p,q) = \max\{\frac{|u(p)-u(q)|^2}{\mathcal{E}_H(u,u)} : u \in \ell(V), \mathcal{E}_H(u,u) \neq 0)\} \tag{2.1.7}$$

Proof. Note that $\frac{|u(p)-u(q)|^2}{\mathcal{E}_H(u,u)} = \frac{|v(p)-v(q)|^2}{\mathcal{E}_H(v,v)}$ if $v = \alpha u + \beta$ for any $\alpha, \beta \in \mathbf{R}$ with $\alpha \neq 0$. For given $u \in \ell(V)$ with $u(p) \neq u(q)$, there exist α and β such that $v(p) = 1$ and $v(q) = 0$ where $v = \alpha u + \beta$. Hence the right-hand side of (2.1.7) equals

$$\max\{\frac{1}{\mathcal{E}_H(v,v)} : v \in \ell(V), v(p) = 1, v(q) = 0\}.$$

Now by (2.1.3), we can verify (2.1.7). □

Applying (2.1.7), we can obtain an inequality between $|u(p)-u(q)|$, $R_H(p,q)$ and $\mathcal{E}_H(u,u)$.

Corollary 2.1.17. *Let V be a finite set and let $H \in \widetilde{\mathcal{L}}(V)$. For any $p, q \in V$ and any $u \in \ell(V)$,*

$$|u(p)-u(q)|^2 \leq R_H(p,q)\mathcal{E}_H(u,u) \tag{2.1.8}$$

This estimate will plays an important role when we will discuss the limit of a sequence of r-networks in the following sections.

As another application of Proposition 2.1.16, we can easily show that $\sqrt{R_H(\cdot,\cdot)}$ is a metric on V.

Theorem 2.1.18. *Let V be a finite set and let $H \in \widetilde{\mathcal{L}}(V)$. Set $R_H^{1/2}(p,q) = \sqrt{R_H(p,q)}$. Then $R_H^{1/2}(\cdot,\cdot)$ is a metric on V.*

Proof. We only need to show the triangle inequality. By (2.1.7), we see that

$$R_H^{1/2}(p,q) = \max\{\frac{|u(p)-u(q)|}{\sqrt{\mathcal{E}_H(u,u)}} : u \in \ell(V), \mathcal{E}_H(u,u) \neq 0\}.$$

This immediately imply the triangle inequality for $R_H^{1/2}(\cdot,\cdot)$. □

§2.2 Sequence of discrete Laplacians

In this section, we will discuss the limit of r-networks on a increasing sequence of finite sets that satisfied certain compatible condition, namely ;

Definition 2.2.1. Let V_m be a finite set and let $H_m \in \widetilde{\mathcal{L}}(V)$ for each $m \geq 0$. $\{(V_m, H_m)\}_{m \geq 0}$ is called a *compatible sequence* if $(V_m, H_m) \leq (V_{m+1}, H_{m+1})$ for all $m \geq 0$. For a compatible sequence $\mathcal{S} = \{(V_m, H_m)\}_{m \geq 0}$, set $V_* = \cup_{m \geq 0} V_m$ and define

$$\mathcal{F}(\mathcal{S}) = \{u : u \in \ell(V_*), \lim_{m \to \infty} \mathcal{E}_{H_m}(u|_{V_m}, u|_{V_m}) < +\infty\} \tag{2.2.1}$$

$$\mathcal{E}_{\mathcal{S}}(u, v) = \lim_{m \to \infty} \mathcal{E}_{H_m}(u|_{V_m}, v|_{V_m}), \tag{2.2.2}$$

for $u, v \in \mathcal{F}(\mathcal{S})$. Also, for $p, q \in V_*$, define the effective resistance associated with $\mathcal{S}$ by

$$R_{\mathcal{S}}(p, q) = R_{H_m}(p, q), \tag{2.2.3}$$

where m is chosen so that $p, q \in V_m$.

In the next chapter, we will approximate a self-similar set by a sequence of increasing finite sets. Then we will construct Dirichlet forms and Laplacians on the self-similar set by taking a limit of a compatible sequence of r-networks.

Throughout this section, $\mathcal{S} = \{(V_m, H_m)\}_{m \geq 0}$ is assumed to be a compatible sequence.

Let us regard V_m as a boundary of V_*. Then for any $u \in \ell(V_m)$, we consider a minimizing problem of $\mathcal{E}_{\mathcal{S}}(\cdot, \cdot)$ under the fixed boundary value u as follows.

Lemma 2.2.2. *There exists a linear map $h_m : \ell(V_m) \to \mathcal{F}(\mathcal{S})$ such that $h_m(u)|_{V_m} = u$ and*

$$\mathcal{E}_{H_m}(u, u) = \mathcal{E}_{\mathcal{S}}(h_m(u), h_m(u)) = \min_{v \in \mathcal{F}(\mathcal{S}), v|_{V_m} = u} \mathcal{E}_{\mathcal{S}}(v, v) \tag{2.2.4}$$

Moreover if $v \in \mathcal{F}(\mathcal{S})$ with $v|_{V_m} = u$ attains the above minimum then $v = h_m(u)$.

Proof. As $[H_n]_{V_m} = H_m$ for $n > m$, we can apply Theorem 2.1.6 with $V = V_n, U = V_m$ and $H = H_n$. Set $h_{n,m} = h$ where h is the linear map $\ell(U) \to \ell(V)$ defined in Theorem 2.1.6. Then define $h_m(u)|_{V_n} = h_{n,m}(u)$. For any $n > m$, this definition is compatible and $h_m(u) \in \ell(V_*)$ is well-defined. By (2.1.2), we have

$$\mathcal{E}_{H_m}(u, u) = \mathcal{E}_{H_n}(h_m(u)|_{V_n}, h_m(u)|_{V_n})$$

for all $n > m$. Therefore $h_m(u) \in \mathcal{F}(\mathcal{S})$. Also (2.1.2) implies (2.2.4) immediately. □

Let us fix m. Then $h_m(u)$ is also characterized by the unique solution of

$$\begin{cases} (H_n v_n)|_{V_n \setminus V_m} & = 0 \quad \text{for all } n > m, \\ \qquad v|_{V_m} & = u, \end{cases}$$

where $v \in \ell(V_*)$ and $v_n = v|_{V_n}$. So $h_m(u)$ may be thought of as a harmonic function with boundary values $u \in V_m$. If $H_m \in \mathcal{L}(V_m)$ for all $m \geq 0$, we can show the following maximum principal for harmonic functions.

We will sometimes think $\ell(V_m)$ as a subset of $\mathcal{F}(\mathcal{S})$ by identifying $\ell(V_m)$ with $h_m(\ell(V_m))$ through the injective map h_m. By this identification, one can write $\mathcal{E}_{H_m}(u, u) = \mathcal{E}_{\mathcal{S}}(u, u)$ for any $u \in \ell(V_m) \subset \mathcal{F}(\mathcal{S})$.

Lemma 2.2.3 (Maximum principle). *Assume $H_m \in \mathcal{L}(V_m)$ for all $m \geq 0$. If $v \in \ell(V_*)$ satisfies $(H_n v_n)|_{V_n \setminus V_m} = 0$ for all $n > m$, where $v_n = v|_{V_n}$, then*

$$\min_{q \in V_m} v(q) \leq v(p) \leq \max_{q \in V_m} v(q)$$

for any $p \in V_*$.

Proof. This follows immediately by the maximum principle for harmonic functions on a finite set, Proposition 2.1.7. □

Next we discuss the effective resistance $\mathcal{R}_{\mathcal{S}}(\cdot,\cdot)$. As in the case for finite sets, $\sqrt{R_{\mathcal{S}}(\cdot,\cdot)}$ becomes a metric on V_*.

Proposition 2.2.4. *If $R_{\mathcal{S}}^{1/2}(\cdot,\cdot) = \sqrt{R_{\mathcal{S}}(\cdot,\cdot)}$, then $R_{\mathcal{S}}^{1/2}$ is a metric on V_*. Moreover if $H_m \in \mathcal{L}(V_m)$ for all $m \geq 0$, then $R_{\mathcal{S}}$ is a metric on V_*.*

Proof. This is an easy corollary of Theorem 2.1.14 and Theorem 2.1.18 along with Proposition 2.1.11 □

The following lemma follows immediately from its counterpart, Proposition 2.1.16.

Lemma 2.2.5. *For any $p, q \in V_*$,*

$$\begin{aligned} R_{\mathcal{S}}(p,q) &= \big(\min\{\mathcal{E}_{\mathcal{S}}(u,u) : u \in \mathcal{F}(\mathcal{S}), u(p)=1, u(q)=0\}\big)^{-1} \\ &= \max\{\frac{|u(p)-u(q)|^2}{\mathcal{E}_{\mathcal{S}}(u,u)} : u \in \mathcal{F}(\mathcal{S}), \mathcal{E}_{\mathcal{S}}(u,u) > 0\} \end{aligned} \tag{2.2.5}$$

This lemma implies that

$$|u(p)-u(q)|^2 \leq R_{\mathcal{S}}(p,q)\mathcal{E}_{\mathcal{S}}(u,u) \tag{2.2.6}$$

for any $u \in \mathcal{F}(\mathcal{S})$ and $p, q \in V_*$. By (2.2.6), we can see that $\mathcal{F}(\mathcal{S}) \subset C(V_*, R_{\mathcal{S}}^{1/2})$. For a metric space (X,d), $C(X,d)$ is the collection of real-valued functions on X that are uniformly continuous on (X,d) and bounded on every bounded subset of (X,d).

Next we present important results of the limit of compatible sequence. In the following chapter, the results will be applied in constructing Dirichlet forms and Laplacians on a self-similar set.

Theorem 2.2.6. (1) $\mathcal{F}(\mathcal{S}) \subset C(V_*, R_{\mathcal{S}}^{1/2})$

(2) *$\mathcal{E}_{\mathcal{S}}$ is a non-negative symmetric form on $\mathcal{F}(\mathcal{S})$. Moreover $\mathcal{E}_{\mathcal{S}}(u,u) = 0$ if and only if u is a constant on V_*.*

(3) *Define an equivalence relation $\sim$ on $\mathcal{F}(\mathcal{S})$ by letting $u \sim v$ if and only if $u - v$ is a constant on V_*. Then $\mathcal{E}_{\mathcal{S}}$ is naturally defined positive definite symmetric form on $\mathcal{F}(\mathcal{S})/\sim$ and $(\mathcal{F}(\mathcal{S})/\sim, \mathcal{E}_{\mathcal{S}})$ is a Hilbert space.*

(4) *Assume that $H_m \in \mathcal{L}(V)$ for all $m \geq 0$. If $\bar{u}$ is defined as in* (DF3) *for any $u \in \mathcal{F}(\mathcal{S})$, then $\bar{u} \in \mathcal{F}(\mathcal{S})$ and $\mathcal{E}_{\mathcal{S}}(\bar{u},\bar{u}) \leq \mathcal{E}_{\mathcal{S}}(u,u)$.*

Proof. Every statement but (3) follows easily from the results and discussions in this and the previous section. We will use $\mathcal{E}$ and $\mathcal{F}$ in place of $\mathcal{E}_{\mathcal{S}}$ and $\mathcal{F}(\mathcal{S})$ respectively. To show (3), first note that $\mathcal{E}(u,u) = \mathcal{E}(v,v)$ if $u \sim v$. Hence $\mathcal{E}$ is a well-defined positive definite symmetric form on $\mathcal{F}/\!\sim$. Choose any $p \in V_*$ and set $\mathcal{F}_p = \{u : u \in \mathcal{F}, u(p) = 0\}$. Then $(\mathcal{F}/\!\sim, \mathcal{E})$ is naturally isomorphic to $(\mathcal{F}_p, \mathcal{E})$. Hence it suffices to show that $(\mathcal{F}_p, \mathcal{E})$ is a Hilbert space. Now let $\{v_n\}_{n\geq 0}$ be a Cauchy sequence in $(\mathcal{F}_p, \mathcal{E})$ and let $v_n^m = h_m(v_n|_{V_m})$. Then by Lemma 2.2.2

$$\mathcal{E}(v_k^m - v_l^m, v_k^m - v_l^m) \leq \mathcal{E}(v_k - v_l, v_k - v_l).$$

Note that, $p \in V_m$ for sufficiently large m. Hence $\mathcal{E}$ becomes an inner product on $\mathcal{F}_p \cap \ell(V_m)$, where $\ell(V_m)$ is identified with $h_m(\ell(V_m))$. So there exists $v^m \in \mathcal{F}_p \cap \ell(V_m)$ such that $v_n^m \to v^m$ as $n \to \infty$. As $v^{m+1}|_{V_m} = v^m$, there exists $v \in \ell(V_*)$ such that $v|_{V_m} = v^m$.

On the other hand, let $C = \sup_{n\geq 0} \mathcal{E}(v_n, v_n)$. Then we have $\mathcal{E}(v^m, v^m) \leq \sup_{n,m} \mathcal{E}(v_n^m, v_n^m) = C$. Hence $v \in \mathcal{F}$.

Now, we fix $\epsilon > 0$. Then, we can choose n so that $\mathcal{E}(v_n - v_k, v_n - v_k) < \epsilon$ for all $k > n$. Also, we can choose m so that

$$|\mathcal{E}(v_n - v, v_n - v) - \mathcal{E}(v_n^m - v^m, v_n^m - v^m)| < \epsilon.$$

Furthermore, we can choose k so that $k > n$ and

$$|\mathcal{E}(v_n^m - v_k^m, v_n^m - v_k^m) - \mathcal{E}(v_n^m - v^m, v_n^m - v^m)| < \epsilon.$$

As $\mathcal{E}(v_n^m - v_k^m, v_n^m - v_k^m) \leq \mathcal{E}(v_n - v_k, v_n - v_k) < \epsilon$, we have $\mathcal{E}(v_n - v, v_n - v) < 3\epsilon$. Thus we have completed the proof of (3). □

Finally we show two examples. The first one is related to one of the most basic examples in probability.

Example 2.2.7 (Simple random walk on Z). Let $V_m = \{-m, -m+1, \cdots, 0, \cdots, m-1, m\}$ and let $H_m \in \mathcal{L}(V_m)$ be defined by $(H_m)_{ij} = 1$ if $|i-j| = 1$, $(H_m)_{ij} = 0$ if $|i-j| > 1$. Then $\mathcal{S} = \{(V_m, H_m)\}_{m\geq 1}$ is a compatible sequence. We can easily see that $V_* = \mathbf{Z}$, $R_S(i,j) = |i-j|$ and

$$\mathcal{E}_{\mathcal{S}}(u,v) = \sum_{i\in\mathbb{Z}} (u(i+1) - u(i))(v(i+1) - v(i)).$$

Also we can see that $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}) \cap L^2(\mathbf{Z}, \mu))$ becomes a regular Dirichlet form on $L^2(\mathbf{Z}, \mu)$ for every Borel measure μ on $\mathbf{Z}$ that satisfies $0 < \mu(\{i\}) < \infty$ for all $i \in \mathbf{Z}$. (See Definition A.4.2 for the definition of regular Dirichlet form.)

Define a linear operator Δ_μ on $L^2(\mathbf{Z}, \mu)$ by

$$(\Delta_\mu u)(i) = \mu(i)^{-1}(u(i+1) + u(i-1) - 2u(i)).$$

Then Δ_μ is a non-positive self-adjoint operator on $L^2(\mathbf{Z}, \mu)$. Also $\mathrm{Dom}(\Delta_\mu) \subset \mathcal{F}(\mathcal{S}) \cap L^2(\mathbf{Z}, \mu)$ and

$$\mathcal{E}_{\mathcal{S}}(u,v) = -\int_{\mathbb{Z}} u \Delta_\mu v d\mu$$

for any $u, v \in \text{Dom}(\Delta_\mu)$. From this fact, Δ_μ is identified as the self-adjoin operator associated with the closed form $(\mathcal{E}_\mathcal{S}, \mathcal{F}(\mathcal{S}) \cap L^2(\mathbf{Z}, \mu))$ on $L^2(\mathbf{Z}, \mu)$. (See §A.2 about a closed form and an associated self-adjoint operator.)

Now if $\nu(i) = 1$ for all $i \in \mathbf{Z}$, then Δ_ν is the self-adjoint operator associated with the simple random walk on $\mathbf{Z}$ in the following sense. Let $u_0 \in \text{Dom}(\Delta_\nu)$ and think about the following evolution equation with discrete time $n = 0, 1, 2, \cdots$;

$$u_{n+1} - u_n = \Delta_\nu u/2.$$

One can easily see that $u_n = (I + \Delta_\nu/2)^n u_0$ for any n. For $i \in \mathbf{Z}$, if $u_0(i) = 1$ and $u_0(k) = 0$ for any $k \neq i$, then $u_n(j)$ is the transition probability from i at time 0 to j at time n under the simple random walk on $\mathbf{Z}$.

Next example is an extreme case where $R_\mathcal{S}$ becomes a trivial metric on V_*.

Example 2.2.8 (Discrete topology). Let $V_m = \{1, 2, \cdots, m\}$ and let $H_m \in \mathcal{L}(V_m)$ be defined by $(H_m)_{ij} = 2/m$ for $i \neq j$. Then $\mathcal{S} = \{(V_m, H_m)\}_{m \geq 2}$ is a compatible sequence. We can easily see that $V_* = \mathbf{N}$ and $R_\mathcal{S}(i, j) = 1$ for $i \neq j$. This metric $R_\mathcal{S}$ induces the discrete topology on $\mathbf{N}$. As $\chi_i \in \mathcal{F}(\mathcal{S})$ for all $i \in \mathbf{N}$, $(\mathcal{E}_\mathcal{S}, \mathcal{F}(\mathcal{S}))$ is a regular Dirichlet form on $L^2(\mathbf{N}, \mu)$ for every Borel measure μ on $\mathbf{N}$ that satisfies $0 < \mu(\{i\}) < \infty$ for all $i \in \mathbf{N}$. See Definition A.4.2 for the definition of a regular Dirichlet form. In particular, let $\mu(\{i\}) = 1$ for all $i \in \mathbf{N}$, then $L^2(\mathbf{N}, \mu) = \ell^2(\mathbf{N})$. We can see that $\ell^2(\mathbf{N}) \cap \mathcal{F}(\mathcal{S}) = \ell^2(\mathbf{N})$ and, for all $u, v \in \ell^2(\mathbf{N})$,

$$\mathcal{E}_S(u, v) = 2 \int_{\mathbb{N}} uv d\mu.$$

§2.3 Resistance Form and Resistance Metric

In the previous section, we constructed a quadratic form $(\mathcal{E}_\mathcal{S}, \mathcal{F}(\mathcal{S}))$ and a metric $R_\mathcal{S}$ from a compatible sequence of r-networks $\mathcal{S} = \{(V_m, H_m)\}_{m \geq 0}$. In this section, we will give characterizations of the form $(\mathcal{E}_\mathcal{S}, \mathcal{F}(\mathcal{S}))$ and the metric $R_\mathcal{S}$ and show that there is an one-to-one correspondence between such forms and metrics.

First we give a characterization of quadratic forms.

Definition 2.3.1 (Resistance form). Let X be a set. A pair $(\mathcal{E}, \mathcal{F})$ is called a resistance form on X if it satisfies the following conditions (RF1) through (RF5).

(RF1) $\mathcal{F}$ is a linear subspace of $\ell(X)$ containing constants and $\mathcal{E}$ is a non-negative symmetric quadratic form on $\mathcal{F}$. $\mathcal{E}(u, u) = 0$ if and only if u is a constant on X.

(RF2) Let $\sim$ be an equivalent relation on $\mathcal{F}$ defined by $u \sim v$ if and only if $u - v$ is a constant on X. Then $(\mathcal{F}/\!\sim, \mathcal{E})$ is a Hilbert space.

(RF3) For any finite subset $V \subset X$ and for any $v \in \ell(V)$, there exists $u \in \mathcal{F}$ such that $u|_V = v$.

(RF4) For any $p, q \in X$,

$$\sup\{\frac{|u(p)-u(q)|^2}{\mathcal{E}(u,u)} : u \in \mathcal{F}, \mathcal{E}(u,u) > 0\}$$

is finite. The above supremum is denoted by $M(p,q)$.

(RF5) If $u \in \mathcal{F}$, then $\bar{u} \in \mathcal{F}$ and $\mathcal{E}(\bar{u},\bar{u}) \leq \mathcal{E}(u,u)$, where $\bar{u}$ is defined in the same way as (DF3) in Definition 2.1.1.

We use $\mathcal{RF}(X)$ to denote the collection of resistance forms on X. Also we define

$$\widetilde{\mathcal{RF}}(X) = \{(\mathcal{E},\mathcal{F}) : (\mathcal{E},\mathcal{F}) \text{ satisfies the condtions (RF1) through (RF4).}\}$$

The condition (RF5) is called the Markov property.

Let V be a finite set. Then $(\mathcal{E}, \ell(V)) \in \widetilde{\mathcal{RF}}(V)$ (or $(\mathcal{E}, \ell(V)) \in \mathcal{RF}(V)$) if and only if $\mathcal{E} \in \widetilde{\mathcal{DF}}(V)$ (or $\mathcal{E} \in \mathcal{DF}(V)$ respectively.) Also immediately from Theorem 2.2.6, $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ belongs to $\widetilde{\mathcal{RF}}(V_*)$ for any compatible sequence $\mathcal{S} = \{(V_m, H_m)\}_{m\geq 0}$. Moreover, if $H_m \in \mathcal{L}(V_m)$ for all m, then $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ becomes a resistance form on V_*.

Next we consider a characterization of metrics.

Definition 2.3.2 (Resistance Metric). Let X be a set. A function $R : X \times X \to \mathbf{R}_+$ is called a resistance metric on X if and only if, for any finite subset $V \subset X$, there exists $H_V \in \mathcal{L}(V)$ such that $R|_{V\times V} = R_{H_V}$, where R_{H_V} is the effective resistance with respect to H_V. The collection of resistance metrics on X is denoted by $\mathcal{RM}(X)$. Also we define

$$\widetilde{\mathcal{RM}}(X) = \{R : X \times X \to \mathbf{R}_+ : \text{For any finite subset } V \subset X, \text{ there exists } H_V \in \widetilde{\mathcal{L}}(V) \text{ with } R|_{V\times V} = R_{H_V} \text{ and } H_{V_1} = [H_{V_2}]_{V_1} \text{ if } V_1 \subseteq V_2\}$$

Remark. Recall that $[H_{V_2}]_{V_1} = P_{V_2,V_1}(H_{V_2})$ by definition. Notice that by Corollary 2.1.13, the condition $H_{V_1} = [H_{V_2}]_{V_1}$ is satisfied for a resistance metric R. If we could extend Theorem 2.1.12 to $\widetilde{\mathcal{L}}(V)$, which is quite likely, then we could remove the assumption $H_{V_1} = [H_{V_2}]_{V_1}$ from the definition of $\widetilde{\mathcal{RM}}(X)$.

Since R_{H_V} is a metric on V, a resistance metric R is a distance on X. Also, for $R \in \widetilde{\mathcal{RM}}(X)$, $\sqrt{R(\cdot,\cdot)}$ is a distance on X.

Let V be a finite set. Then $R \in \widetilde{\mathcal{RM}}(V)$ (or $R \in \mathcal{RM}(V)$) if and only if $R = R_H$ for some $H \in \widetilde{\mathcal{L}}(V)$ (or $H \in \mathcal{L}(V)$ respectively). Also it is natural to expect that $R_{\mathcal{S}}$ is a resistance metric. More precisely, we have the following proposition.

Proposition 2.3.3. *If $\mathcal{S} = \{(V_m, H_m)\}_{m\geq 0}$ is a compatible sequence, then $R_{\mathcal{S}} \in \mathcal{RM}(V_*)$. In particular, if $H_m \in \mathcal{L}(V)$ for all m, then $R_{\mathcal{S}}$ is a resistance metric.*

Proof. Let V be a finite subset of V_*. Then $V \subseteq V_m$ for sufficiently large m. If $H_V = [H_m]_V$, then $R_{H_V} = R_{\mathcal{S}}|_{V\times V}$. The rest of the conditions are obvious. □

There is a natural one-to-one correspondence between resistance forms and resistance metrics. First we will construct a resistance metric from a resistance form.

Theorem 2.3.4. *If $(\mathcal{E}, \mathcal{F}) \in \widetilde{\mathcal{RF}}(X)$, then*

$$\min\{\mathcal{E}(u,u) : u \in \mathcal{F}, u(p) = 1, u(q) = 0\}$$

exists for any $p, q \in X$ with $p \neq q$. If we define $R(p,q)^{-1}$ to be equal to the minimum value, then $R \in \mathcal{RM}(X)$ and

$$R(p,q) = \max\{\frac{|u(p) - u(q)|^2}{\mathcal{E}(u,u)} : u \in \mathcal{F}, \mathcal{E}(u,u) > 0\}. \tag{2.3.1}$$

Moreover, if $(\mathcal{E}, \mathcal{F}) \in \mathcal{RF}(X)$, then $R \in \mathcal{RM}(X)$.

To prove the above theorem, we need the following lemma.

Lemma 2.3.5. *If $(\mathcal{E}, \mathcal{F}) \in \widetilde{\mathcal{RF}}(X)$ and V is a finite subset of X, then there exists a linear map $h_V : \ell(V) \to \mathcal{F}$ such that $h_V(u)|_V = u$ and*

$$\mathcal{E}(h_V(u), h_V(u)) = \min_{v\in\mathcal{F}, v|_V = u} \mathcal{E}(v,v). \tag{2.3.2}$$

Furthermore $h_V(u)$ is the unique element that attains the above minimum. Also set $\mathcal{E}^V(u_1, u_2) = \mathcal{E}(h_V(u_1), h_V(u_2))$. Then $\mathcal{E}^V \in \widetilde{\mathcal{DF}}(V)$. Moreover if $(\mathcal{E}, \mathcal{F})$ is a resistance form, then $\mathcal{E}^V \in \mathcal{DF}(V)$.

Proof. For $p \in V$, let $\mathcal{F}^p = \{u : u \in \mathcal{F}, u|_{V\backslash\{p\}} = 0\}$. Then by (RF.3), $\mathcal{F}^p$ is not trivial. By (RF.2), $(\mathcal{F}^p, \mathcal{E})$ is a Hilbert space. Define $\Phi_p\ \mathcal{F}^p \to \mathbf{R}$ by $\Phi_p(u) = u(p)$. Then by (RF.4) we have, for $q \in V\backslash\{p\}$, $|\Phi_p(u)|^2 \leq M(p,q)\mathcal{E}(u,u)$ for all $u \in \mathcal{F}^p$. Hence Φ_p is a continuous linear functional on $(\mathcal{F}^p, \mathcal{E})$. Therefore there exists $g_p \in \mathcal{F}^p$ such that for all $u \in \mathcal{F}^p$, $\mathcal{E}(g_p, u) = \Phi_p(u) = u(p)$. As $\mathcal{E}(g_p, g_p) = g_p(p) > 0$, we can define $\psi_p^V = g_p/g_p(p)$.

Now for any $u \in \ell(V)$, define $h_V(u) = \sum_{p\in V} u(p)\psi_p^V$. If $v \in \mathcal{F}$ with $v|_V = u$, set $\tilde{v} = v - h_V(u)$. Then

$$\mathcal{E}(v,v) = \mathcal{E}(\tilde{v} + h_V(u), \tilde{v} + h_V(u)) = \mathcal{E}(\tilde{v}, \tilde{v}) + 2\mathcal{E}(\tilde{v}, h_V(u)) + \mathcal{E}(h_V(u), h_V(u)).$$

As $\mathcal{E}(\tilde{v}, h_V(u)) = \sum_{p\in V} u(p)\tilde{v}(p)/g_p(p) = 0$, we have

$$\mathcal{E}(v,v) = \mathcal{E}(\tilde{v}, \tilde{v}) + \mathcal{E}(h_V(u), h_V(u)) \geq \mathcal{E}(h_V(u), h_V(u)).$$

Equality holds only when $\tilde{v}$ is constant on X and so $\tilde{v} \equiv 0$ on X. It is easy to see that $\mathcal{E}^V \in \widetilde{\mathcal{DF}}(V)$. Also the Markov property (RF5) of $(\mathcal{E}, \mathcal{F})$ implies the Markov property (DF3) of $\mathcal{E}^V$. Thus we have completed the proof. □

Proof of Theorem 2.3.4. By Lemma 2.3.5,

$$\min\{\mathcal{E}(u,u) : u \in \mathcal{F}, u(p) = 1, u(q) = 0\}$$

exists for any $p, q \in X$ with $p \neq q$. Now define $H_V \in \widetilde{\mathcal{L}}(V)$ by $\mathcal{E}^V = \mathcal{E}_{H_V}$. If $V_1 \subset V_2$, then

$$\begin{aligned}\mathcal{E}_{[H_{V_2}]_{V_1}}(u,u) &= \min_{v \in \ell(V_2), v|_{V_1} = u} \mathcal{E}(h_{V_2}(v), h_{V_2}(v)) \\ &= \min_{v' \in \mathcal{F}, v'|_{V_1} = u} \mathcal{E}(v', v') = \mathcal{E}_{H_{V_1}}(u,u).\end{aligned}$$

Hence $[H_{V_2}]_{V_1} = H_{V_1}$. This fact also implies $R_{H_V} = R|_{V \times V}$. Therefore $R \in \widetilde{\mathcal{RM}}(X)$. If $(\mathcal{E}, \mathcal{F}) \in \mathcal{RF}(X)$, then $H_V \in \mathcal{DF}(V)$ and hence $R \in \mathcal{RM}(X)$. The same argument as the proof of Proposition 2.1.16 implies (2.3.1). $\square$

Theorem 2.3.4 says that each $(\mathcal{E}, \mathcal{F}) \in \widetilde{\mathcal{RF}}(X)$ is associated with $R \in \widetilde{\mathcal{RM}}(X)$. So we can define a map $FM_X : \widetilde{\mathcal{RF}}(X) \to \widetilde{\mathcal{RM}}(X)$, which is called the "form to metric" map, by $R = FM_X((\mathcal{E}, \mathcal{F}))$. This form to metric map is, in fact, bijective. Namely, we can construct the inverse of FM_X.

Theorem 2.3.6. *For $R \in \widetilde{\mathcal{RM}}(X)$, there exists a unique $(\mathcal{E}, \mathcal{F}) \in \widetilde{\mathcal{RF}}(X)$ that satisfies* (2.3.1)*. Moreover if $R \in \mathcal{RM}(X)$, then $(\mathcal{E}, \mathcal{F}) \in \mathcal{RF}(X)$.*

Assuming the above theorem, we can define the "metric to form" map $MF_X : \widetilde{\mathcal{RM}}(X) \to \widetilde{\mathcal{RF}}(X)$. It is easy to see that MF_X is the inverse of FM_X.

We will only present the proof of a special case of Theorem 2.3.6, namely Theorem 2.3.7. Theorem 2.3.6 can proven by using routine and tedious discussions about limiting procedure from the special case.

If $R \in \widetilde{\mathcal{RM}}(X)$, $R^{1/2}(\cdot,\cdot) = \sqrt{R(\cdot,\cdot)}$ is a metric on X. Assume that the metric space $(X, R^{1/2})$ is separable. Equivalently, there exists a family of finite subsets $\{V_m\}_{m \geq 0}$ of X that satisfies $V_m \subset V_{m+1}$ for $m \geq 0$ and $V_* = \cup_{m \geq 0} V_m$ is dense in X. Set $H_m = H_{V_m}$. Then $H_m \in \widetilde{\mathcal{L}}(V_m)$ and $[H_{m+1}]_{V_m} = H_m$ by definition. Hence $(V_m, H_m) \leq (V_{m+1}, H_{m+1})$ and so $\mathcal{S} = \{(V_m, H_m)\}_{m \geq 0}$ is a compatible sequence. We know that $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})) \in \widetilde{\mathcal{RF}}(V_*)$. Also it is obvious that $R = R_{\mathcal{S}}$ on V_*. Now as $\mathcal{F}(\mathcal{S}) \in C(V_*, R_{\mathcal{S}}^{1/2})$, $u \in \mathcal{F}(\mathcal{S})$ has a natural extension to a function in $C(X, R^{1/2})$. We will think $\mathcal{F}(\mathcal{S})$ as a subset of $C(X, R^{1/2})$ in this way. Then it is easy to see that $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ satisfies (RF1) and (RF2). This $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ is the candidate for $(\mathcal{E}, \mathcal{F})$ in Theorem 2.3.6. The problem is to show (RF3) and (2.3.1) for any $p, q \in X$. (We already know that (2.3.1) holds for $p, q \in V_*$ by Lemma 2.2.5.) We do this in the next theorem.

Theorem 2.3.7. *For $R \in \widetilde{\mathcal{RM}}(X)$, assume that $(X, R^{1/2})$ is separable. Let $\{V_m\}_{m \geq 0}$ be a family of finite subsets of X such that $V_m \subset V_{m+1}$ for any $m \geq 0$ and that $V_* = \cup_{m \geq 0} V_m$ is dense in X. Set $\mathcal{S} = \{(V_m, H_m)\}_{m \geq 0}$ where $H_m = H_{V_m}$. Then $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})) \in \widetilde{\mathcal{RF}}(X)$ and*

$$R(p,q) = \max\{\frac{|u(p) - u(q)|^2}{\mathcal{E}_{\mathcal{S}}(u,u)} : u \in \mathcal{F}(\mathcal{S}), \mathcal{E}_{\mathcal{S}}(u,u) > 0\} \tag{2.3.3}$$

for all $p, q \in X$. Moreover, $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ is independent of the choice of $\{V_m\}_{m\geq 0}$. Also if $R \in \mathcal{RM}(X)$, then $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})) \in \mathcal{RF}(X)$.

Before proving the theorem, we need two lemmas.

Lemma 2.3.8. *Let $(\mathcal{E}, \mathcal{F}) \in \widetilde{\mathcal{RF}}(X)$ and let $\{V_m\}$ be a sequence of finite subsets of X such that $V_m \subset V_{m+1}$ for $m \geq 0$ and that $V_* = \cup_{m\geq 0} V_m$ is dense in $(X, R^{1/2})$, where $R = FM_X((\mathcal{E}, \mathcal{F}))$. If $\mathcal{S} = \{(V_m, H_m)\}_{m\geq 0}$ where $H_m = H_{V_m}$, then $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})) = (\mathcal{E}, \mathcal{F})$.*

Remark. In this lemma, again we think $\mathcal{F}(\mathcal{S})$ as a subset of $C(X, R^{1/2})$ because $R = R_{\mathcal{S}}$ on V_* and $\mathcal{F}(\mathcal{S}) \subset C(V_*, R_{\mathcal{S}}^{1/2})$.

Proof. First we show that $\mathcal{F}(\mathcal{S}) \subseteq \mathcal{F}$ and $\mathcal{E}_{\mathcal{S}}(u, u) = \mathcal{E}(u, u)$ for $u \in \mathcal{F}(\mathcal{S})$. Let $u \in \mathcal{F}(\mathcal{S})$. Set $u_m = h_{V_m}(u|_{V_m})$, where h_{V_m} is defined in Lemma 2.3.5. As $\mathcal{E}_{H_m}(u|_{V_m}, u|_{V_m}) = \mathcal{E}(u_m, u_m)$, we obtain $\mathcal{E}(u_m, u_m) \leq \mathcal{E}_(u_{m+1}, u_{m+1}) \leq \mathcal{E}_{\mathcal{S}}(u, u)$. Now without loss of generality we may assume that $u(p) = 0$ for some $p \in V_0$. (We can just replace u by $u - u(p)$.) Note that $(u_m - u_n)|_{V_n} = 0$ for $m \geq n$. Then recalling the definition of h_V in the proof of Lemma 2.3.5, it follows that

$$\mathcal{E}(u_m - u_n, u_n) = \sum_{p\in V_n} (u_m(p) - u_n(p))u(p)/g_p(p) = 0$$

for $m \geq n$. Hence $\mathcal{E}(u_m - u_n, u_m - u_n) = \mathcal{E}(u_m, u_m) - \mathcal{E}(u_n, u_n) \to 0$ as $m, n \to \infty$. Therefore $\{u_m\}_{m\geq 0}$ is a Cauchy sequence in $(\mathcal{F}_p, \mathcal{E})$. As $(\mathcal{F}_p, \mathcal{E})$ is complete by (RF2), there exists $u_* \in \mathcal{F}_p$ such that $\mathcal{E}(u_* - u_m, u_* - u_m) \to 0$ as $m \to \infty$. So $\mathcal{E}(u_*, u_*) = \lim_{m\to\infty} \mathcal{E}(u_m, u_m) = \mathcal{E}_{\mathcal{S}}(u, u)$. For $q \in V_*$, we have

$$|u_*(q) - u_m(q)|^2 \leq R(p, q)\mathcal{E}(u_* - u_m, u_* - u_m).$$

Letting $m \to \infty$, we obtain that $u|_{V_*} = u_*|_{V_*}$. As u and u_* is continuous on X with respect to $R^{1/2}$, we can see that $u = u_*$. Thus we have shown that $u \in \mathcal{F}$ and $\mathcal{E}(u, u) = \mathcal{E}_{\mathcal{S}}(u, u)$.

Secondly, for $u \in \mathcal{F}$, define u_m exactly same as before. Then $\mathcal{E}(u_m, u_m) = \min_{v\in\mathcal{F}, v|_{V_m}=u|_{V_m}} \mathcal{E}(v, v) \leq \mathcal{E}(u, u)$. Hence $u \in \mathcal{F}(\mathcal{S})$. Now use the discussion of the latter half of this proof, we can see that $\mathcal{E}_{\mathcal{S}}(u, u) = \mathcal{E}(u, u)$. □

Lemma 2.3.9. *Let $(\mathcal{E}, \mathcal{F}) \in \widetilde{\mathcal{RF}}(Y)$. Let $(\overline{Y}, \overline{R}^{1/2})$ be the completion of $(Y, R^{1/2})$, where $R = FM_Y((\mathcal{E}, \mathcal{F}))$. Then for any $p, q \in \overline{Y}$,*

$$\overline{R}(p, q) = \max\{\frac{|u(p) - u(q)|^2}{\mathcal{E}(u, u)} : u \in \mathcal{F}, \mathcal{E}(u, u) > 0\}.$$

Proof. First we will show that

$$\overline{R}(p, q) = \sup\{\frac{|u(p) - u(q)|^2}{\mathcal{E}(u, u)} : u \in \mathcal{F}, \mathcal{E}(u, u) > 0\}. \tag{2.3.4}$$

We will denote the right-hand side of (2.3.4) by $M(p,q)$. Choose $\{p_n\}, \{q_n\} \subset Y$ so that $p_n \to p$ and $q_n \to q$ as $n \to \infty$. Note that,

$$R(p_n, q_n) = \max\{\frac{|u(p_n) - u(q_n)|^2}{\mathcal{E}(u,u)} : u \in \mathcal{F}, \mathcal{E}(u,u) > 0\}. \tag{2.3.5}$$

Hence, we have $|u(p_n) - u(q_n)|^2 \leq R(p_n, q_n)\mathcal{E}(u,u)$ for any $u \in \mathcal{F}$. Letting $n \to \infty$, we obtain $|u(p) - u(q)|^2 \leq \overline{R}(p,q)\mathcal{E}(u,u)$. Hence $M(p,q) \leq \overline{R}(p,q)$.

Suppose $M(p,q) < \overline{R}(p,q)$. Then we can choose ϵ so that for all $u \in \mathcal{F}$,

$$|u(p) - u(q)| < (\sqrt{\overline{R}(p,q)} - 5\epsilon)\sqrt{\mathcal{E}(u,u)}.$$

On the other hand, since $R(p_n, q_n) \to \overline{R}(p,q)$ as $n \to \infty$, using (2.3.5), there exists $\{u_n\}$ such that $\mathcal{E}(u_n, u_n) = 1$ and $|u_n(p_n) - u_n(q_n)|^2 \to \overline{R}(p,q)$ as $n \to \infty$. For sufficiently large n, we have

$$|u_n(p_n) - u_n(q_n)| > (\sqrt{\overline{R}(p,q)} - \epsilon)$$

and $R(p_n, p_m), R(q_n, q_m) < \epsilon^2$ for all $m > n$. Furthermore we can choose m so that $m > n$ and

$$|u_n(p_m) - u_n(p)| < \epsilon \quad \text{and} \quad |u_n(q_m) - u_n(q)| < \epsilon.$$

Now we have

$$\begin{aligned} |u_n(p_n) - u_n(q_n)| \leq\ & |u_n(p_n) - u_n(p_m)| + |u_n(p_m) - u_n(p)| + |u_n(p) - u_n(q)| + \\ & |u_n(q) - u_n(q_m)| + |u_n(q_m) - u_n(q_n)| \\ \leq\ & |u_n(p_n) - u_n(p_m)| + |u_n(q_n) - u_n(q_m)| + (\sqrt{\overline{R}(p,q)} - 3\epsilon). \end{aligned}$$

Hence $|u_n(p_n) - u_n(p_m)| \geq \epsilon$ or $|u_n(q_n) - u_n(q_m)| \geq \epsilon$. This contradicts the fact that $R(p_n, p_m), R(q_n, q_m) < \epsilon^2$. Therefore we have shown (2.3.4).

Now using the same argument as in the proof of Lemma 2.3.5, it follows that there exists $\psi \in \mathcal{F}$ such that $\psi(p) = 1$, $\psi(q) = 0$ and ψ attains the supremum in (2.3.4). □

Proof of Theorem 2.3.7. As we mentioned before, $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ (recall that we think $\mathcal{F}(\mathcal{S})$ as a subset of $C(X, R^{1/2})$) satisfies (RF1) and (RF2). To show (RF3), set $V'_m = V_m \cup V$ for a finite set $V \subset X$. Let $H'_m = H_{V'_m}$ and let $\mathcal{S}' = \{(V'_m, H'_m)\}$. Then for any $u \in \ell(V)$, there exists $v \in \mathcal{F}(\mathcal{S}')$ such that $v|_V = u$. As $(V_m, H_m) \leq (V'_m, H'_m)$, $\mathcal{E}_{H_m}(v|_{V_m}, v|_{V_m}) \leq \mathcal{E}_{H'_m}(v|_{V'_m}, v|_{V'_m}) \leq \mathcal{E}_{\mathcal{S}'}(v,v)$. Hence $\lim_{m\to\infty} \mathcal{E}_{H_m}(v|_{V_m}, v|_{V_m}) \leq \mathcal{E}_{\mathcal{S}'}(v,v)$. Therefore $v \in \mathcal{F}(\mathcal{S})$. This shows (RF3).

Next, applying Lemma 2.3.9 for the case that $Y = V_*$, we obtain (2.3.3) because $X \subset \overline{Y}$. This implies (RF4). Thus we have shown $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})) \in \widetilde{\mathcal{RF}}(X)$. Furthermore, (2.3.3) also implies $R = FM_X((\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})))$.

Let $\{U_m\}$ be a sequence of finite subsets of X that satisfies the same condition as $\{V_m\}$ and let $\mathcal{S}_1$ be the compatible sequence associated with U_m.

Then applying Lemma 2.3.8, we can see that $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})) = (\mathcal{E}_{\mathcal{S}_1}, \mathcal{F}(\mathcal{S}_1))$. Hence $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ is independent of the choice of $\{V_m\}$.

Finally if $R \in \mathcal{RM}(X)$, then $H_{V_m} \in \mathcal{L}(V_m)$. Hence $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ has the Markov property. □

Using the discussions in this section, we can show another important fact about resistance forms and resistance metrics. If $\mathcal{S} = \{(V_m, H_m)\}_{m \geq 0}$ is a compatible sequence, then $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})) \in \widetilde{\mathcal{RF}}(V_*)$ and $R_{\mathcal{S}} \in \widetilde{\mathcal{RM}}(V_*)$. The space V_* is merely a countable set. So if we would construct analytical objects like Laplacians or Dirichlet forms from $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$, we would end up with an analysis on a countable set. That is hardly what we want! One way of overcoming this difficulty is to consider the completion of V_* with respect to the metric $R_{\mathcal{S}}^{1/2}$. Let $(\Omega_{\mathcal{S}}, R_{\mathcal{S}}^{1/2})$ be the completion of $(V_*, R_{\mathcal{S}}^{1/2})$. Then $(\Omega_{\mathcal{S}}, R_{\mathcal{S}}^{1/2})$ could be an interesting uncountable infinite set. As we mentioned before, $\mathcal{F}(\mathcal{S})$ can be naturally thought of as a subset of $C(\Omega_{\mathcal{S}}, R_{\mathcal{S}}^{1/2})$. Hence $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ can be considered as a quadratic form on $(\Omega_{\mathcal{S}}, R_{\mathcal{S}}^{1/2})$. There is, however, a little delicate question about this completion procedure. Is the extended $R_{\mathcal{S}}$ in $\widetilde{\mathcal{RM}}(\Omega_{\mathcal{S}})$? Equivalently, do we have $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})) \in \widetilde{\mathcal{RF}}(X)$? This is not an trivial problem. In fact, this is not true in general. See Exercise 2.7 for a counter example. Fortunately, if we assume the Markov property, i.e. $H_m \in \mathcal{L}(V_m)$ for all $m \geq 0$, then it follows that $R_{\mathcal{S}}^{1/2} \in \mathcal{RM}(\Omega_{\mathcal{S}})$ and $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S})) \in \mathcal{RF}(\Omega_{\mathcal{S}})$ by virtue of the next theorem.

Theorem 2.3.10. *Let $(\mathcal{E}, \mathcal{F}) \in \mathcal{RF}(X)$. If $(\overline{X}, R)$ is the completion of (X, R), where $R = FM_X((\mathcal{E}, \mathcal{F}))$, then $(\mathcal{E}, \mathcal{F}) \in \mathcal{RF}(\overline{X})$ and $R \in \mathcal{RM}(\overline{X})$.*

Proof. (RF1), (RF2) and (RF5) follow immediately. Also (RF4) is an obvious consequence of Lemma 2.3.9. Instead of (RF3), we may show the following (RF3*).

(RF3*) For each finite subset $V \subset \overline{X}$ and for each $p \in V$, there exists $u \in \mathcal{F}$ such that $u|_V = \chi_p$, where χ_p is the characteristic function of the one point set $\{p\}$.

We use an induction on $\#(V)$ to prove (RF3*). If $\#(V) = 2$, say $V = \{p, q\}$, then by Lemma 2.3.9, there exists $u \in \mathcal{F}$ such that $u(p) \neq u(q)$. If we set $f = (u - u(q))/(u(p) - u(q))$, we have $f|_V = \chi_p$.

Next suppose (RF3*) holds for $\#(V) < n$. Let $V = \{p_1, p_2, \cdots, p_n\}$, then by the induction hypothesis, there exists $u \in \mathcal{F}$ such that $u(p_1) = 1$ and $u(p_i) = 0$ for $i \geq 3$.

Case 1 If $u(p_2) < u(p_1)$, then for some $\alpha, \beta \in \mathbf{R}$, $v = \alpha u + \beta$ satisfies $v(p_1) = 1$ and $v(p_j) \leq 0$ for $j \geq 2$. Define $\bar{v}$ as in (DF3) of Definition 2.1.1. Then by the Markov property (RF5), we have $\bar{v} \in \mathcal{F}$. Obviously $\bar{v}|_V = \chi_{p_1}$.

Case 2 If $u(p_2) = u(p_1)$, then choose $f \in \mathcal{F}$ that satisfies $f(p_1) > f(p_2)$ and $|f(p_i)| < 1/2$ for all $i = 1, 2, \cdots, n$. For some $\alpha, \beta \in \mathbf{R}$, $v = \alpha(u + f) + \beta$ has the same properties as v in Case 1.

Case 3 If $u(p_1) < u(p_2)$, then using the same discussion as in Case 1, we can find $v \in \mathcal{F}$ that satisfies $v|_V = \chi_{p_2}$. Thus if $f = u - u(p_2)v$, then $f|_V = \chi_{p_1}$.

Thus we have shown that (RF3*) holds for $\#(V) = n$. $\square$

§2.4 Dirichlet forms and Laplacians on limits of networks

In the last section, we have studied relations between a compatible sequence of r-networks, a resistance form and a resistance metric. In this section, we will take a first step to establish an "analysis" on limits of networks. In particular, we are interested in constructing a counterpart of the Laplacian defined as a differential operator in the classical calculus. By the results in the last section, it is reasonable to start from a compatible sequence of r-networks $\mathcal{S} = \{(V_m, H_m)\}$. (We do not concern how to obtain a compatible sequence of r-networks in this section.) Then we obtain a resistance form $(\mathcal{E}, \mathcal{F})$ and a resistance metric R by taking a limit of $\mathcal{S}$. Naturally, the resistance form $(\mathcal{E}, \mathcal{F})$ and the resistance metric R are important elements in our "analysis". However, those are not enough. We need to introduce an integration, namely, a measure on the space. The following is a general result concerning a resistance form, a resistance metric and a measure.

Theorem 2.4.1. *Let $R \in \mathcal{RM}(X)$ and suppose that $(X, R^{1/2})$ is separable. Set $(\mathcal{E}, \mathcal{F}) = MF_X(R)$. Also let μ be a σ-finite Borel measure on $(X, R^{1/2})$. Define*

$$\mathcal{E}_1(u, v) = \mathcal{E}(u, v) + \int_X u(x)v(x)\mu(dx)$$

for $u, v \in L^2(X, \mu) \cap \mathcal{F}$. Then $(L^2(X, \mu) \cap \mathcal{F}, \mathcal{E}_1)$ is a Hilbert space. Moreover, if $\mu(X) < \infty$ and $\int_X R(p, p_*)\mu(dp) < \infty$ for some $p_* \in X$, then the identity map from $L^2(X, \mu) \cap \mathcal{F}$ with $\mathcal{E}_1$-norm to $L^2(X, \mu)$ with L^2-norm is a compact operator.

Proof. Let $\{u_n\}_{n \geq 0}$ be a Cauchy sequence in $(L^2(X, \mu) \cap \mathcal{F}, \mathcal{E}_1)$ and let $v_n = u_n - u_n(p)$ for $p \in X$. Then by (RF.2), there exists $v \in \mathcal{F}_p$ such that $\mathcal{E}(v_n - v, v_n - v) \to 0$ as $n \to \infty$. By (RF4), we have

$$|v_n(q) - v(q)|^2 \leq R(p, q)\mathcal{E}(v_n - v, v_n - v). \tag{2.4.1}$$

Since μ is σ-finite, there exists $\{K_m\}_{m \geq 0}$ such that $K_m \subset X$ is bounded, $0 < \mu(K_m) < \infty$ and $\cup_{m \geq 0} K_m = X$. By (2.4.1), we see that $v_n \to v$ as $n \to \infty$ in $L^2(K_m, \mu|_{K_m})$. Also $\{u_n|_{K_m}\}_{n \geq 0}$ is a Cauchy sequence in $L^2(K_m, \mu|_{K_m})$. As $u_n(p) = (u_n - v_n)|_{K_m}$, there exists $c \in \mathbf{R}$ such that $u_n(p) \to c$ as $n \to \infty$. If we let $u = v + c$, then $\mathcal{E}(u - u_n, u - u_n) \to 0$ as $n \to \infty$. Also, $u_n|_{K_m} \to u|_{K_m}$ as $n \to \infty$ in $L^2(K_m, \mu|_{K_m})$.

On the other hand, $\{u_n\}_{n \geq 0}$ is a Cauchy sequence in $L^2(X, \mu)$ and so there exists $u^* \in L^2(X, \mu)$ such that $u_n \to u^*$ as $n \to \infty$ in $L^2(X, \mu)$. As $u^*|_{K_m} =$

$u|_{K_m}$ in $L^2(K_m, \mu|_{K_m})$, $u = u^*$ in $L^2(X, \mu)$. Hence $u_n \to u$ as $n \to \infty$ in $(L^2(X, \mu) \cap \mathcal{F}, \mathcal{E}_1)$.

Now suppose $\mu(X) < \infty$. Let $\mathcal{U}$ be a bounded subset of $(L^2(X, \mu) \cap \mathcal{F}, \mathcal{E}_1)$. If $C = \sup_{u \in \mathcal{U}} \mathcal{E}_1(u, u)$, then

$$|u(p) - u(q)|^2 \leq C\, R(p, q) \tag{2.4.2}$$

for all $u \in \mathcal{U}$ and all $p, q \in X$. Let V be a countable dense subset of X. Note that $\{u(p)\}_{u \in \mathcal{U}}$ is bounded for any $p \in V$ by (2.4.2). Hence by standard diagonal construction, we can find $v \in \ell(V)$ and $\{v_n\} \subset \mathcal{U}$ satisfying $v_n(p) \to v(p)$ as $n \to \infty$ for all $p \in V$. Using (2.4.2), we see that v satisfies (2.4.2) for $p, q \in V$. Therefore v extends naturally to a function $v \in C(X, R^{1/2})$ and it satisfies (2.4.2) for all $p, q \in X$ as well. For any $p \in X$, choose $\{p_n\} \subset V$ so that $p_n \to p$ as $n \to \infty$. Then

$$\begin{aligned}|v_k(p) - v(p)| &\leq |v_k(p) - v_k(p_n)| + |v_k(p_n) - v(p_n)| + |v(p_n) - v(p)| \\ &\leq 2\sqrt{C\, R(p, p_n)} + |v_k(p_n) - v(p_n)|.\end{aligned}$$

Hence we see that $v_n(p) \to v(p)$ as $n \to \infty$ for all $p \in X$. By (RF4),

$$|(v_k(p) - v_l(p)) - (v_k(p_*) - v_l(p_*))|^2 \leq \mathcal{E}(v_k - v_l, v_k - v_l) R(p, p_*).$$

As $\mathcal{E}(v_n, v_n) \leq C$, the above inequality implies

$$|v_k(p) - v_l(p)| \leq \sqrt{4C\, R(p, p_*)} + |v_k(p_*) - v_l(p_*)|.$$

Letting $l \to \infty$, we have $|v_k(p) - v(p)|^2 \leq 4C\, R(p, p_*) + 1$ for large k. Since $\int_X R(p, p_*) \mu(dp) < \infty$, it follows that $v_k \to v$ in $L^2(X, \mu)$ as $k \to \infty$ by Lebesgue's dominate convergence theorem. □

Now, we have collected enough facts to use an abstract theory in functional analysis. In fact, by Theorem 2.4.1, we can apply the well-developed theory of closed forms and self-adjoint operators, which is introduced in Appendix §A.2.

Theorem 2.4.2. *Let $R \in \mathcal{RM}(X)$ and suppose that $(X, R^{1/2})$ is separable. Set $(\mathcal{E}, \mathcal{F}) = MF_X(R)$. Also let μ be a σ-finite Borel measure on $(X, R^{1/2})$. Also assume that $L^2(X, \mu) \cap \mathcal{F}$ is dense in $L^2(X, \mu)$ with respect to the L^2-norm. Then there exists a non-negative self-adjoint operator H on $L^2(K, \mu)$ such that $\mathrm{Dom}(H^{1/2}) = \mathcal{F}$ and $\mathcal{E}(u, v) = (H^{1/2}u, H^{1/2}v)$ for all $u, v \in \mathcal{F}$. Moreover if $\mu(X) < \infty$ and $\int_X R(p, p_*)\mu(dp) < \infty$ for some $p_* \in X$, then H has compact resolvent.*

Proof. Set $\mathcal{H} = L^2(X, \mu)$, $Q(\cdot, \cdot) = \mathcal{E}$ and $\mathrm{Dom}(Q) = \mathcal{F}$. Then Theorem 2.4.1 along with Theorem A.2.6 immediately imply the required results. □

Assume that $R \in \mathcal{RM}(X)$. Also in addition to the assumptions of Theorem 2.4.2, we assume that X is a locally compact metric space. Then we see that $(\mathcal{E}, \mathcal{F} \cap L^2(X, \mu))$ is a Dirichlet form on $L^2(X, \mu)$. Moreover if $\mathcal{F} \cap L^2(X, \mu) \cap C_0(X)$ is dense in $C_0(X)$ with respect to the supremum norm, then

$(\mathcal{E}, \mathcal{F} \cap L^2(X,\mu))$ is a regular Dirichlet forms. (See §A.3 for the definition of Dirichlet forms and $C_0(X)$.) In fact, if $(\mathcal{E}, \mathcal{F})$ comes from a regular harmonic structure, which is defined in §3.1, we can verify all the conditions above and get a Dirichlet forms and a Laplacian immediately from the theorems in this section. See the next chapter for details.

Form an abstract point of view, the self-adjoint operator $-H$ should be our Laplacian. However, this abstract construction is too general to study detailed information on our Laplacian. For example, it is quite difficult to get concrete expressions of harmonic functions and Green's function only from this abstract definition. So, we also need to construct a Laplacian on a self-similar set in a classical way, namely, as a direct limit of discrete Laplacians. See Chapter 3, in particular §3.7.

Example 2.4.3. Let K be any closed subset of $\mathbf{R}$. We can always find an increasing sequence of finite sets $\{V_m\}_{m\geq 0}$ that satisfies $V_m \subseteq V_{m+1}$ and $\overline{\cup_{m\geq 0} V_m} = K$. If $V_m = \{p_{m,i}\}_{i=1}^{n_m}$ and $p_{m,i} < p_{m,i+1}$ for all i, then we define $H_m \in \mathcal{L}(V_m)$ by

$$(H_m)_{p_{m,i}p_{m,j}} = \begin{cases} |p_{m,i} - p_{m,j}|^{-1} & \text{if } |i-j| = 1, \\ 0 & \text{otherwise,} \end{cases}$$

for $i \neq j$. Then $\{(V_m, H_m)\}_{m\geq 0}$ becomes a compatible sequence. Also if R is the effective resistance defined on $\cup_{m\geq 0} V_m$, then R coincides with the restriction of the Euclidean metric. Let $(\mathcal{E}, \mathcal{F})$ be the corresponding resistance form and let μ be a σ-finite Borel regular measure on K. First note that $f|_K$ belongs to $\mathcal{F}$ for any piecewise linear function f on $\mathbf{R}$ with $\mathrm{supp}(f)$ compact. By this fact, it follows that $\mathcal{F} \cap L^2(K,\mu)$ is dense in $L^2(K,\mu)$ with respect to the L^2-norm. Set $\mathcal{F}_1 = \mathcal{F} \cap L^2(K,\mu)$. Then $(\mathcal{E}, \mathcal{F}_1)$ becomes a local regular Dirichlet form on $L^2(K,\mu)$.

This example contains many interesting cases. The most obvious one is the case where $K = \mathbf{R}$. In this case, $\mathcal{F}_1$ coincides with $H^1(\mathbf{R})$, which is the completion of

$$\{u \in C^1(\mathbf{R}) : \int_{\mathbb{R}} u'(x)^2 dx < \infty, \mathrm{supp}(f) \text{ is compact.}\}$$

with respect to the H^1-norm $||\cdot||_1$ defined by

$$||u||_1 = \sqrt{\int_{\mathbb{R}} (u(x)^2 + u'(x)^2) dx}.$$

Also $\mathcal{E}(u,v) = \int_{\mathbb{R}} u'(x)v'(x)dx$. If μ is the Lebesgue measure on $\mathbf{R}$, then the non-negative self-adjoint operator H coincides with the standard $-\Delta = -d^2/dx^2$.

One of other interesting cases is the Cantor set. Let K be the Cantor set defined in Example 1.2.6. Let μ be a self-similar measure on K. (See §1.4 for the definition of self-similar measures.) Then $(\mathcal{E}, \mathcal{F})$ becomes a local regular Dirichlet form on $L^2(K,\mu)$. By using Theorem 2.4.2, we obtain a non-negative self-adjoint operator H. Set $\Delta_\mu = -H$. Then Δ_μ is though of as a Laplacian on the Cantor set K. In [32, 33], Fujita studied the spectrum of Δ_μ and the asymptotic behavior of the associated (generalized) diffusion process on the Cantor set.

Exercise

Exercise 2.1. Show that every metric on V coincides with an effective resistance metric associated with a Laplacian $H \in \mathcal{L}(V)$ if $\#(V) = 3$.

Exercise 2.2. Let $V = \{p_1, p_2, p_3, p_4\}$ and let d be a metric on V defined by, for

$$d(p_i, p_j) = \begin{cases} 1 & \text{if } (i,j) \neq (1,2) \text{ and } i \neq j, \\ x & \text{if } (i,j) = (1,2), \\ 0 & \text{if } i = j, \end{cases}$$

for some x with $0 < x \leq 2$. Show that there exists a Laplacian $H \in \mathcal{L}(V)$ such that $R_H = d$ if and only if $x \leq 3/2$.

Exercise 2.3. Verify that the Δ-Y transform remains true even if $H \in \widetilde{\mathcal{L}}(V)$.

Exercise 2.4. Show that if $\#(V) = 3$, $H \in \mathcal{L}(V)$ if and only if $R_H(\cdot,\cdot)$ is a metric on V.

Exercise 2.5. Let $V = \{p_1, p_2, p_3, p_4\}$. For $i \neq j$, set

$$H_{p_ip_j} = \begin{cases} 1 & \text{if } (i,j) \neq (1,4), \\ -\epsilon & \text{if } (i,j) = (1,4), \end{cases}$$

where $\epsilon > 0$. Show that if ϵ is sufficiently small, $H \in \widetilde{\mathcal{L}}(V)$ and $R_H(\cdot,\cdot)$ becomes a metric on V.

Exercise 2.6. Let $V = \{p_1, p_2, p_3\}$. Define

$$H_m = \begin{pmatrix} -(1+m) & 1+2m & -m \\ 1+2m & -2(1+2m) & 1+2m \\ -m & 1+2m & -(1+m) \end{pmatrix}.$$

Show that $R_{H_m}(p_i, p_j)$ converges as $m \to \infty$. Also show that there exists no $H \in \widetilde{\mathcal{L}}(V)$ such that $R = R_H$, where $R(p_i, p_j) = \lim_{m\to\infty} R_{H_m}(p_i, p_j)$.

Exercise 2.7. Let $X = \{a, b\} \cup \{p_m\}_{m\geq 1}$. Define $R(a,b) = 2, R(a, p_m) = R(b, p_m) = \frac{1+m}{1+2m}$ and $R(p_j, p_k) = \frac{|k-j|}{(1+2k)(1+2j)}$.

(1) Show that $R \in \mathcal{RM}(X)$.

(2) Let $(\overline{X}, R^{1/2})$ be the completion of $(X, R^{1/2})$. Show that $R \notin \mathcal{RM}(\overline{X})$. (Hint: See Exercise 2.6.)

Chapter 3

Construction of Laplacians on P.C.F. Self Similar Structures

In this chapter, we will construct analysis associated with Laplacians on connected post critically finite self-similar structure. Precisely, in this chapter, $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$ is a post critically finite (p. c. f. for short) self-similar structure and K is assumed to be connected. (Also in this chapter, we always set $S = \{1, 2, \cdots, N\}$.) Recall that a condition for K being connected was given in §1.6.

The key idea of constructing a Laplacian (or a Dirichlet form) on K is finding a "self-similar" compatible sequence of r-networks on $\{V_m\}_{m \geq 0}$, where $V_m = V_m(\mathcal{L})$ defined in Lemma 1.3.10. Note that $\{V_m\}_{m \geq 0}$ is a monotone increasing sequence of finite sets. We will formulate such a self-similar compatible sequence in §3.1. Once we get such a sequence, we can use the general theory in the last chapter and construct a resistance form $(\mathcal{E}, \mathcal{F})$ and a resistance metric R on V_*, where $V_* = \cup_{m \geq 0} V_m$

If the closure of V_* with respect to the metric R were always identified with K, then we could apply Theorem 2.4.2 and see that $(\mathcal{E}, \mathcal{F})$ is a regular local Dirichlet form on $L^2(K, \mu)$ for any self-similar measure μ on K. Consequently we could immediately obtain a Laplacian associated with the Dirichlet form $(\mathcal{E}, \mathcal{F})$ on $L^2(K, \mu)$. Unfortunately, as we will see in §3.3, the closure of V_* with respect to R is merely a proper subset of K for certain case. In spite of this difficulty, we will show a condition on a probability measure μ which is sufficient for $(\mathcal{E}, \mathcal{F})$ to be a regular local Dirichlet form on $L^2(K, \mu)$ in §3.4

As is mentioned in §2.4, there is an abstract way of constructing Laplacian from a Dirichlet form $(\mathcal{E}, \mathcal{F})$ on $L^2(K, \mu)$. (See §A.2 for details.) However, we will develop our analysis on a p. c. f. self-similar set K in a classical and explicit way like the ordinary Laplacian d^2/dx^2 on the unit interval. In §3.5, Green's function will be given in a constructive manner. In the following sections, we will study counterparts of classical analysis on Euclidean spaces, for example, Green's operator in §3.6 and Gauss-Green's formula in Theorem 3.7.8. Finally we will define a Laplacian as a scaling limit of discrete Laplacians on V_m in §3.7.

Throughout this chapter, d is a metric on K which gives the original topology of K as a compact metric space. Also we write $C(K) = C(K, d)$. Since (K, d) is compact, $C(K)$ is the collection of all real-valued continuous functions on K.

§3.1 Harmonic structures

In this section, we start constructing Dirichlet forms and Laplacians on K. As is mentioned above, the basic idea is finding a "self-similar" compatible sequence of r-networks on $\{V_m\}_{m\geq 0}$ and taking a limit of it. (Recall that $V_m \subseteq V_{m+1}$ by Lemma 1.3.10.)

For any initial $D \in \mathcal{L}(V_0)$, we can construct a sequence of self-similar Laplacians $H_m \in \mathcal{L}(V_m)$ as follows.

Definition 3.1.1. If $D \in \mathcal{L}(V_0)$ and $\mathbf{r} = (r_1, r_2, \cdots, r_N)$, where $r_i > 0$ for $i \in S$, we define $\mathcal{E}^{(m)} \in \mathcal{DF}(V_m)$ by

$$\mathcal{E}^{(m)}(u, v) = \sum_{w \in W_m} \frac{1}{r_w} \mathcal{E}_D(u \circ F_w, v \circ F_w)$$

for $u, v \in \ell(V_m)$, where $r_w = r_{w_1} \cdots r_{w_m}$ for $w = w_1 w_2 \cdots w_m \in W_m$. Also $H_m \in \mathcal{L}(V_m)$ is characterized by $\mathcal{E}^{(m)} = \mathcal{E}_{H_m}$.

It is easy to see that

$$\mathcal{E}^{(m+1)}(u, v) = \sum_{i=1}^{N} \frac{1}{r_i} \mathcal{E}^{(m)}(u \circ F_i, v \circ F_i) \tag{3.1.1}$$

for $u, v \in \ell(V_m)$. Also $H_m = \sum_{w \in W_m} \frac{1}{r_w} {}^t R_w D R_w$, where $R_w : \ell(V_m) \to \ell(V_0)$ is defined by $R_w f = f \circ F_w$ for $w \in W_m$. We write $\mathcal{E}_m = \mathcal{E}^{(m)}$ hereafter.

Considering (3.1.1), we may think (V_m, H_m) as a self-similar sequence of r-networks. If it is also a compatible sequence, then it is possible to construct a Laplacian on K using the theory in the previous chapter.

Definition 3.1.2 (Harmonic structures). $(D, \mathbf{r})$ is called a harmonic structure if and only if $\{(V_m, H_m)\}_{m\geq 0}$ becomes a compatible sequence of r-networks. Also a harmonic structure $(D, \mathbf{r})$ is said to be regular if $0 < r_i < 1$ for all $i \in S$.

Once we get a harmonic structure, we can use the general framework in Chapter 2 (in particular, Theorem 2.2.6, Theorem 2.4.1 and Theorem 2.4.2) to construct a quadratic form $(\mathcal{E}, \mathcal{F})$ on $V* = \cup_{m\geq 0} V_m$ and an associated non-negative self-adjoint operator H on $L^2(\Omega, \mu)$, where Ω is completion of V_* under the resistance metric associated with $(\mathcal{E}, \mathcal{F})$ and μ is a given σ-finite Borel regular measure. This H should be our Laplacian. It seems easy but there remains a "slight" problem : the topology on V_* given by the resistance metric may be different from the original topology of K. In such a case, Ω does not coincides with K. In fact, we will see in the next section that $\Omega = K$ if and only if the harmonic structure is regular.

Another important problem is whether there exists any harmonic structure on a p. c. f. self-similar structure. By virtue of the self-similar construction of H_m, we can simplify the condition for harmonic structures as follows.

Proposition 3.1.3. $(D, \mathbf{r})$ *is a harmonic structure if and only if* $(V_0, D) \leq (V_1, H_1)$.

Proof. Assume that $(V_{m-1}, H_{m-1}) \leq (V_m, H_m)$. Then, for any $u \in \ell(V_m)$, we have $\mathcal{E}_{m-1}(u \circ F_i, u \circ F_i) = \min\{\mathcal{E}_m(v \circ F_i, v \circ F_i) : v \in \ell(V_{m+1}), v|_{V_m} = u\}$. Hence by (3.1.1), we have $\mathcal{E}_m(u, u) = \min\{\mathcal{E}_{m+1}(v, v) : v \in \ell(V_{m+1}), v|_{V_m} = u\}$. Therefore $(V_m, H_m) \leq (V_{m+1}, H_{m+1})$. So by induction, if $(V_0, D) \leq (V_1, H_1)$, then $(V_m, H_m) \leq (V_{m+1}, H_{m+1})$ for any $m \geq 0$. The converse is obvious. □

For given $\mathbf{r} = (r_1, r_2, \cdots, r_N)$, define $\mathcal{R}_{\mathbf{r}} : \mathcal{L}(V_0) \to \mathcal{L}(V_0)$ by $\mathcal{R}_{\mathbf{r}}(D) = [H_1]_{V_0}$, where $H_1 \in \mathcal{L}(V_1)$ is given by Definition 3.1.1. $\mathcal{R}_{\mathbf{r}}$ is called a renormalization operator on $\mathcal{L}(V_0)$. By the above proposition, D is a harmonic structure if and only if D is a fixed point of $\mathcal{R}_{\mathbf{r}}$. Also it is easy to see that $\mathcal{R}_{\lambda\mathbf{r}}(\alpha D) = (\lambda)^{-1}\alpha\mathcal{R}_{\mathbf{r}}(D)$ for any $\alpha, \lambda > 0$. Hence if D is an eigenvector of $\mathcal{R}_{\mathbf{r}}$, i. e. $\mathcal{R}_{\mathbf{r}}(D) = \lambda D$, then D is a fixed point of $\mathcal{R}_{\lambda\mathbf{r}}$. So, the existence problem of harmonic structures is reduced to a fixed point problem (or eigenvalue problem) for the non-linear homogeneous map $\mathcal{R}_r$. In general, this problem is not easy and we have not fully understand the situation yet. For example, it is not known whether there exists at least one harmonic structure on a p.c.f. self-similar set. The only one general result on existence of a harmonic structure is the theory of nested fractals by Linstrøm [94]. The nested fractals are highly symmetric self-similar structures. (See §3.8 for the definition.) We will present slightly extended version of Lindstrøm's result on existence of a harmonic structure on nested fractals in §3.8.

Example 3.1.4 (Interval). Set $F_1(x) = x/2$ and $F_2(x) = x/2 + 1/2$. Then $\mathcal{L} = ([0, 1], \{1, 2\}, \{F_1, F_2\})$ becomes a p. c. f. self-similar structure. We see that $V_m = \{i/2^m\}_{i=1,2,\cdots,2^m}$. Let us define $D \in \mathcal{L}(V_0)$ by

$$D = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Then $(D, \mathbf{r})$ becomes a harmonic structure on $\mathcal{L}$ if $\mathbf{r} = (r_1, r_2)$ satisfies that $r_1 + r_2 = 1$ and $0 < r_i < 1$ for $i = 1, 2$. Also it is easy to see that those are all the harmonic structures on $\mathcal{L}$.

Example 3.1.5 (Sierpinski gasket). Recall Example 1.2.8 and 1.3.13. The Sierpinski gasket is a p. c. f self-similar set with $V_0 = \{p_1, p_2, p_3\}$. Define

$D \in \mathcal{L}(V_0)$ by

$$D = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}.$$

Also set $\mathbf{r} = (3/5, 3/5, 3/5)$. Then we see that $(D, \mathbf{r})$ is a harmonic structure on the Sierpinski gasket. $(D, \mathbf{r})$ is called the standard harmonic structure on the Sierpinski gasket. There are other harmonic structures on the Sierpinski gasket if we loosen the symmetry. See Exercise 3.1.

Example 3.1.6 (Hata's tree-like set). Recall Example 1.2.9 and 1.3.14. Let $\mathcal{L}$ be the self-similar structure associated with Hata's tree-like set. Then $V_0 = \{c, 0, 1\}$ as in the previous example. Define $D \in \mathcal{L}(V_0)$ by

$$D = \begin{pmatrix} -h & h & 0 \\ h & -(h+1) & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

and define $\mathbf{r} = (r, 1 - r^2)$ for $r \in (0, 1)$. If $rh = 1$, then $(D, \mathbf{r})$ becomes a regular harmonic structure on $\mathcal{L}$.

So far we presented examples of regular harmonic structure. Of course, there are many example of non-regular harmonic structure.

Example 3.1.7. Set $F_1(z) = z/2, F_2(z) = z/2 + 1/2$ and $F_3(z) = \sqrt{-1}z/3 + 1/2$. Let K be the self-similar set with respect to $\{F_1, F_2, F_3\}$ and let $\mathcal{L} = (K, \{1, 2, 3\}, \{F_1, F_2, F_3\})$. Then $\mathcal{L}$ is a p. c. f. self-similar structure. In fact, $\mathcal{C}_{\mathcal{L}} = \{1\dot{2}, 2\dot{1}, 3\dot{1}\}, \mathcal{P}_{\mathcal{L}} = \{\dot{1}, \dot{2}\}$ and $V_0 = 0, 1$. If $D = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ and $\mathbf{r} = (r, 1 - r, s)$ for $r \in (0, 1)$ and $s > 0$, then $(D, \mathbf{r})$ is a harmonic structure on $\mathcal{L}$. Obviously, $(D, \mathbf{r})$ is not regular for $s \geq 1$.

See Exercise 3.2 for a more natural example of non-regular harmonic structures.

Proposition 3.1.8. *For $w \in W_*$, let $\dot{w}$ denote the periodic sequence in Σ defined by $\dot{w} = www\cdots$. Let $(D, \mathbf{r})$ be a harmonic structure and let $\dot{w} \in \mathcal{P}$ for $w \in W_*$. Then $r_w < 1$. In particular, there exists $i \in S$ such that $r_i < 1$.*

Remark. If $(K, S, \{F_i\}_{i \in S})$ is a p.c.f. self-similar structure,then the post critical set $\mathcal{P}$ consists of eventually periodic points : for any $\omega \in \mathcal{P}$, there exist $w \in W_*$ and $m \geq 0$ such that $\sigma^m \omega = \dot{w}$.

Corollary 3.1.9. *Let $(D, \mathbf{r})$ be a harmonic structure. If $r_1 = \cdots = r_N$, then $r_i < 1$ for any $i \in S$. In particular, $(D, \mathbf{r})$ is a regular harmonic structure.*

To prove Proposition 3.1.8, we need the following lemma.

Lemma 3.1.10. *Let V be a finite set and let $H \in \mathcal{L}(V)$. Suppose $U \subseteq V$ and $p \in U$. If there exists $q_* \in V \backslash U$ such that $H_{pq_*} \neq 0$, then $-h_{pp} < -H_{pp}$, where $(h_{kl})_{k,l \in U} = [H]_U$.*

Proof. H can be expressed as $\begin{pmatrix} T & {}^tJ \\ J & X \end{pmatrix}$, where $T : \ell(U) \to \ell(U), J : \ell(U) \to \ell(V\backslash U)$ and $X : \ell(V\backslash U) \to \ell(V\backslash U)$. Then $[H]_U = T - {}^tJX^{-1}J$. Now let $\psi_p = -X^{-1}J\chi_p^U$, where $\chi_p^U(x) = 1$ if $x = p$ and $\chi_p^U(x) = 0$ if $x \neq p$ on U. It follows that

$$h_{pp} = H_{pp} + \sum_{q \in V\backslash U} H_{pq}\psi_p(q).$$

As $H_{pq_*} \neq 0$, the maximum principle (Proposition 2.1.7) implies that $\psi_p(q_*) > 0$. Therefore $\sum_{q \in V\backslash U} H_{pq}\psi_p(q) > 0$. □

Proof of Proposition 3.1.8. First we assume that

$$\#(F_i(V_0) \cap V_0) \leq 1 \text{ for all } i \in S. \tag{3.1.2}$$

As $H_1 = \sum_{i=1}^N \frac{1}{r_i} {}^tR_iDR_i$, we have $(H_1)_{pp} = \sum_{(q,i):q \in V_0, F_i(q)=p} \frac{1}{r_i} D_{qq}$. Set $p = \pi(\dot{w})$, where $w = w_1w_2 \cdots w_m \in W_*$ and $\dot{w} \in \mathcal{P}$. As $\mathcal{P}$ is a finite set, $\pi^{-1}(\pi(\dot{w})) = \{\dot{w}\}$. Hence, $\{(q,i) : q \in V_0, F_i(q) = p\} = \{(\pi(\sigma\dot{w}), w_1)\}$. Therefore, $(H_1)_{pp} = \frac{1}{r_{w_1}} D_{qq}$, where $q = \pi(\sigma\dot{w})$. So, let $p_i = \pi(\sigma^{i-1}\dot{w})$ for $i = 1, 2, \cdots, m+1$, then $(H_1)_{p_ip_i} = \frac{1}{r_{w_i}} D_{p_{i+1}p_{i+1}}$, where we set $w_{m+1} = w_1$. Now by (3.1.2), we can apply Lemma 3.1.10 and obtain $-D_{p_{i+1}p_{i+1}} < -(H_1)_{p_{i+1}p_{i+1}}$. So we have $\prod_{i=1}^m -(H_1)_{p_ip_i} < {r_w}^{-1} \prod_{i=1}^m -(H_1)_{p_{i+1}p_{i+1}}$. Hence we have $r_w < 1$.

If (3.1.2) is not satisfied, we will replace the original self-similar structure $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$ by $\mathcal{L}_m = (K, W_m, \{F_w\}_{w \in W_m})$. Then by Proposition 1.3.11, we can see that $\mathcal{P}_{\mathcal{L}} = \mathcal{P}_{\mathcal{L}_m}$. Also, it is easy to see that $(D, \mathbf{r}_m)$, where $\mathbf{r}_m = (r_w)_{w \in W_m}$, becomes a harmonic structure on $\mathcal{L}_m$. For sufficiently large m, $\mathcal{L}_m$ satisfies (3.1.2) and hence we can apply the above argument to the harmonic structure $(D, \mathbf{r}_m)$. Therefore $(r_w)^m < 1$. Thus we obtain that $r_w < 1$. □

Exercise 3.1. Let $\mathcal{L}$ be the harmonic structure associated with the Sierpinski gasket. (See Example 3.1.5.) Set

$$D = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -(1+h) & h \\ 1 & h & -(1+h) \end{pmatrix}.$$

and $\mathbf{r} = (s, st, st)$, where h, s and t are positive real numbers. Show that if we fix $h > 0$, there exist unique s and t such that $(D, \mathbf{r})$ becomes a harmonic structure on the Sierpinski gasket. Also prove that $(D, \mathbf{r})$ is a regular harmonic structure.

Hint: let $R = 1/h$. Then calculate the effective resistances for D and H_1 by using the Δ-Y transform (Lemma 2.1.15). Then apply Corollary 2.1.13. You will find that the condition for $(D, \mathbf{r})$ being a harmonic structure is

$$s(1 + \frac{(R+t)^2}{2(tR+t+R)}) = s(t + \frac{t(R+t)}{tR+t+R}) = 1.$$

Exercise 3.2 (modified Sierpinski gasket). Let $\{p_1, p_2, p_3\}$ be the vertices of a regular triangle in the complex plane $\mathbf{C}$. Set $p_4 = (p_2+p_3)/2, p_5 = (p_3+p_1)/2$ and $p_6 = (p_1+p_2)/2$. Choose real numbers α and β so that $2\alpha + \beta = 1$ and $\alpha > \beta > 0$. We define $F_i(z) = \alpha(z - p_i) + p_i$ for $i = 1, 2, 3$ and $F_i(z) = \beta(z - p_i) + p_i$ for $i = 4, 5, 6$. Let K be the self-similar set with respect to $\{F_i\}_{i \in S}$, where $S = \{1, 2, \cdots, 6\}$.

(1) Prove that the self-similar structure $\mathcal{L}$ associated with K is post critically finite with $V_0 = \{p_1, p_2, p_3\}$.

(2) Define $D \in \mathcal{L}(V_0)$ by

$$D = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}.$$

Let $\mathbf{r} = (r, r, r, rs, rs, rs)$ where $r, s > 0$. Show that if we fix s, then there exists an unique r such that $(D, \mathbf{r})$ becomes a harmonic structure on $\mathcal{L}$. Is this harmonic structure regular?

Hint : use Δ-Y transform and calculate effective resistances as in Exercise 3.1.

§3.2 Harmonic functions

Let $(D, \mathbf{r})$ be a harmonic structure on $\mathcal{L} = (K, S, \{F_i\}_{i \in S})$, where $S = \{1, 2, \cdots, N\}$. Then $\{(V_m, H_m)\}_{m \geq 0}$ is a compatible sequence of r-networks. So we can construct $(\mathcal{E}, \mathcal{F})$ as in (2.2.1) and (2.2.2). By Theorem 2.2.6, $(\mathcal{E}, \mathcal{F}) \in \mathcal{RF}(V_*)$, where $V_* = \cup_{m \geq 0} V_m$. In this section, we consider harmonic functions associated with $(\mathcal{E}, \mathcal{F})$. Arguments in the last chapter, in particular Lemma 2.2.2, imply the following result.

Proposition 3.2.1. *For any $\rho \in \ell(V_0)$, there exists a unique $u \in \mathcal{F}$ such that $u|_{V_0} = \rho$ and $\mathcal{E}(u, u) = \min\{\mathcal{E}(v, v) : v \in \mathcal{F}, v|_{V_0} = \rho\}$. Furthermore, u is characterized by the unique solution of*

$$\begin{cases} (H_m v)|_{V_m \setminus V_0} & = 0 \quad \text{for all } m \geq 1, \\ \qquad v|_{V_0} & = \rho. \end{cases}$$

The function u obtained in the above theorem is called a harmonic function with boundary value ρ. Let R be the resistance metric on V_* associated with $(\mathcal{E}, \mathcal{F})$. Then by (2.2.6), $u \in C(V_*, R)$.

Note that V_* is a countable subset of K and the topology of (V_*, R) may be different from that of V_* with the relative topology form the original metric on K. We will see, however, that a harmonic function has a unique extension to a continuous function on K.

Now recall that d is a metric on K which is compatible with the original topology of K.

Theorem 3.2.2. *Let u be a harmonic function. Then there exists a unique $\tilde{u} \in C(K)$ such that $u|_{V_*} = \tilde{u}|_{V_*}$.*

Remark. As is shown in §3.3, the closure of (V_*, R) equals K with the original topology if and only if $(D, \mathbf{r})$ is regular harmonic structure. In such a case, the above theorem is obvious.

Proof. Let u be a harmonic function with boundary value ρ. Set $H_1 = \begin{pmatrix} T & {}^tJ \\ J & X \end{pmatrix}$, where $T : \ell(V_0) \to \ell(V_0), J : \ell(V_0) \to \ell(V_1 \backslash V_0)$ and $X : \ell(V_1 \backslash V_0) \to \ell(V_1 \backslash V_0)$. Then it follows that

$$(u \circ F_i)|_{V_0} = R_i(u|_{V_1}) = R_i \begin{pmatrix} \rho \\ -X^{-1} J \rho \end{pmatrix}.$$

As F_w is bijective mapping between V_0 and $F_w(V_0)$ for $w \in W_*$, we will identify $\ell(V_0)$ and $\ell(F_w(V_0))$ through F_w. Define a linear map $A_i : \ell(V_0) \to \ell(F_i(V_0)) \cong \ell(V_0)$ by

$$A_i \rho = R_i \begin{pmatrix} \rho \\ -X^{-1} J \rho \end{pmatrix}, \tag{3.2.1}$$

then

$$u|_{F_w(V_0)} = A_{w_m} A_{w_{m-1}} \cdots A_{w_1} \rho \tag{3.2.2}$$

for $w = w_1 w_2 \cdots w_m \in W_*$ and

$$(A_i)_{pq} \geq 0 \quad \text{for any } p, q \in V_0 \quad \text{and} \quad A_i \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{3.2.3}$$

First we prove the theorem assuming (3.1.2).

Claim 1 Set $v(f) = \max_{p,q \in V_0} |f(p) - f(q)|$ for $f \in \ell(V_0)$. Then $v(A_i f) < v(f)$ if $v(f) \neq 0$.

Proof of Claim 1: If $H_1 g|_{V_1 \backslash V_0} = 0$ and $g|_{V_0} = f$, then $A_i f = g|_{F_i(V_0)}$. Applying the maximum principle (Proposition 2.1.7) and taking (3.1.2) into account, we can see that $\max_{q \in F_i(V_0)} g(q) - \min_{q \in F_i(V_0)} g(q) < v(f)$. Hence $v(A_i f) < v(f)$.

Claim 2 There exists c_i such that $0 < c_i < 1$ and $v(A_i f) \leq c_i v(f)$ for any $f \in \ell(V_0)$.

Proof of Claim 2: Define $Q : \ell(V_0) \to \ell(V_0)$ by

$$(Qf)(p) = f(p) - \#(V_0)^{-1} \sum_{q \in V_0} f(q),$$

then $v(f) = v(Qf)$ and $v(A_i f) = v(A_i Q f)$ for any $f \in \ell(V_0)$. Hence

$$\sup\{\frac{v(A_i f)}{v(f)} : f \in \ell(V_0), v(f) \neq 0\} = \sup\{\frac{v(A_i Q f)}{v(Qf)} : f \in \ell(V_0), v(f) \neq 0\}$$
$$= \sup\{\frac{v(A_i f)}{v(f)} : f \in \ell(V_0), \sum_{q \in V_0} f(q) = 0, v(f) = 1\}.$$

As $\{f \in \ell(V_0) : \sum_{q \in V_0} f(q) = 0, v(f) = 1\}$ is compact, the above supremum is less than 1.

Now by Claim 2 and the maximum principle, we can see that $v_w(u) = \sup\{|f(p) - f(q)| : p, q \in K_w \cap V_*\} \leq c^m v(\rho)$ for any $w \in W_*$, where $c = \max_{i \in S} c_i$. Hence, if $\{p_i\}_{i \geq 1}$ is a Cauchy sequence with respect to a metric on K which is compatible with the original topology of K, then $\{u(p_i)\}_{i \geq 1}$ is convergent as $i \to \infty$. Using this limit, we can extend u to a continuous function $\tilde{u}$ on K.

Next if (3.1.2) is not satisfied, we can exchange the harmonic structure as in the proof of Proposition 3.1.8. Then again we can use the result under a new self-similar structure $\mathcal{L}_m$ and a harmonic structure $(D, \mathbf{r}_m)$. Note that harmonic functions remain same after we replace the harmonic structure. □

Hereafter, we identify u with its extension $\tilde{u}$ and think of a harmonic function as a continuous function on K.

Example 3.2.3 (Sierpinski gasket). Let us calculate the probabilistic matrices $\{A_i\}_{i \in S}$ for the standard harmonic structure on the Sierpinski gasket given in Example 3.1.5. Recall that $V_0 = \{p_1, p_2, p_3\}$. Set $q_1 = (p_2 + p_3)/2, q_2 = (p_3 + p_1)/2$ and $q_3 = (p_1 + p_2)$. Then $V_1 = \{p_i, q_i\}_{i \in S}$, where $S = \{1, 2, 3\}$. Now Let $f(p_1) = a, f(p_2) = b$ and $f(p_3) = c$ and solve the linear equation $(H_1 f)(q_i) = 0$ for $i \in S$. Then we get $f(q_1) = (2b + 2c + a)/5, f(q_2) = (2c + 2a + b)/5$ and $f(q_3) = (2a + 2b + c)/5$. By this result, we see that

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \end{pmatrix} A_2 = \begin{pmatrix} \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \\ 0 & 1 & 0 \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \end{pmatrix} A_3 = \begin{pmatrix} \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \\ 0 & 0 & 1 \end{pmatrix}.$$

It is easy to see that the eigenvalues of A_i are $1, \frac{3}{5}, \frac{1}{5}$. Note that the second eigenvalue $\frac{3}{5}$ is equal to r_i. In fact, this is not a coincidence. (Recall that $\mathbf{r} = (3/5, 3/5, 3/5)$.) In §A.1, we will show a general result on the second eigenvalue of A_i.

See Exercise 3.3 for more examples.

The probability matrices $\{A_i\}_{i\in S}$ determine the harmonic functions through (3.2.2). The behavior of a harmonic function around a point $\pi(\omega)$ for $\omega \in \Sigma(S)$ is given by the asymptotic behavior of (3.2.2) for $m \to \infty$. This is the problem of random iterations of matrices and, in general, it is very difficult. Even in the above example, we don't know how to calculate the behavior of $A_{\omega_m}A_{\omega_{m-1}}\cdots A_{\omega_1}$ as $m \to \infty$ unless the sequence ω is (eventually) periodic. Kusuoka used $\{A_i\}_{i\in S}$ to construct Dirichlet forms on finitely ramified self-similar sets in Kusuoka [85] and got some result about almost sure behavior of the random iteration of $\{A_i\}_{i\in S}$.

An important property of harmonic functions is the Harnack inequality, which follows from the discrete version, Corollary 2.1.8.

Proposition 3.2.4 (the Harnack inequality). *If X is a compact subset of K that is contained in a connected component of $K\backslash V_0$, then there exists a constant $c > 0$ such that $max_{x\in X}u(x) \leq c\min_{x\in X} u(x)$ for any non-negative harmonic function u on K.*

Proof. Set $X_m = \cup_{w\in W_m:K_w\cap X\neq\emptyset}K_w$. Then we can choose m so that $X_m \cap V_0 = \emptyset$. Now set $V = V_m$, $U = V_0$, $H = H_m$ and $A = X_m \cap V_m$. Applying the Harnack inequality (Corollary 2.1.8), we see that there exists $c > 0$ such that $\max_{p\in A} u(p) \leq c\min_{p\in A} u(p)$ for any non-negative harmonic function u on K. Using the maximum principle, $\max_{x\in X} u(x) \leq \max_{p\in A} u(p)$ and $\min_{p\in A} u(p) \leq \max_{x\in X} u(x)$. Hence we have shown the required inequality. □

As an corollary of Theorem 3.2.2, we can define piecewise harmonic functions as follows.

Corollary 3.2.5. *For $\rho \in \ell(V_m)$, there exists a unique continuous function u on K such that $u|_{V_m} = \rho$ and $\mathcal{E}(u|_{V_*}, u|_{V_*}) = \min\{\mathcal{E}(v,v) : v \in \mathcal{F}, v|_{V_m} = \rho\}$.*

u in the above corollary is called an m-harmonic function with boundary value ρ. Another characterization of m-harmonic functions is that u is an m-harmonic function if and only if $u \circ F_w$ is a harmonic function for any $w \in W_m$. For $p \in V_m$, define ψ_p^m to be the m-harmonic function with boundary value $\chi_p^{V_m}$. Then any m-harmonic function u is a linear combination of $\{\psi_p^m\}$. In fact, $u = \sum_{p\in V_m} u(p)\psi_p^m$. Note that if $u_m = \sum_{p\in V_m} u(p)\psi_p^m$ for $u \in \ell(V_*)$, then $\mathcal{E}_m(u|_{V_m}, u|_{V_m}) = \mathcal{E}(u_m, u_m)$.

In the rest of this section, we will give an expansion of $u \in \ell(V_*)$ in a piecewise harmonic basis $\{\psi_p\}_{p\in V_*}$, where $\psi_p = \psi_p^m$ if $p \in V_m\backslash V_{m-1}$.

Lemma 3.2.6. *Let u be an m-harmonic function. Then $\mathcal{E}(u, f) = 0$ for $f \in \mathcal{F}$ if $f|_{V_m} = 0$.*

Proof. For $n > m$, we have $(H_n u)(p) = 0$ if $p \in V_n\backslash V_m$ and $f(p) = 0$ if $p \in V_m$. Hence $\mathcal{E}_n(u, f) = -\sum_{p\in V_n} f(p)(H_n u)(p) = 0$. □

Lemma 3.2.7. *For $u \in \mathcal{F}$, $\mathcal{E}(u - u_m, u - u_m) \to 0$ as $m \to \infty$.*