BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
INSTITUTE OF MATHEMATICS
DEPARTMENT OF STOCHASTICS

THESIS

Chaos game representation based on DNA sequences

ANNA ORSZÁG

Supervisor:

Balázs Bárány

Assistant Professor, Department of Stochastics

2016

Contents

Chapter 1

Introduction

The motivation for this BSc thesis came from three articles, the work of *H. J. Jeffrey* [1], *D. Koslicki* and *D. J. Thompson* [2], and *M. F. Barnsley* and *K. Leśniak* [3].

Jeffrey presented a method for representing DNA sequences, called *Chaos Game Representation (CGR)*, see precisely in Chapter 3. A nucleotide sequence is composed of the four acid bases: adenine (A), cytozine (C), guanine (G) and thymine (T) – or in certain cases uracil (U). Take a square where the four vertices are labelled 'A', 'C', 'G' and 'T'. The first point is placed halfway between the center and the vertex corresponding to the first nucleotide in the sequence. The next point is plotted halfway between the previous point and the vertex corresponding to the second nucleotide, etc. So our functions are the following:

$$f_A\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}, \quad f_T\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x + 1 \\ y \end{pmatrix},$$

$$f_G\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x + 1 \\ y + 1 \end{pmatrix}, \quad f_C\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x + 1 \\ y + 1 \end{pmatrix}.$$

We can see an example for the CGR's few steps in Fig. 1.1. Following this algorithm, we get a picture with a complex structure depending on the sequence. The chaos game representation of $73\,356$ bases of HUMHBB (Human Beta Globin Region on Chromosome 11) is in Fig. 1.2.
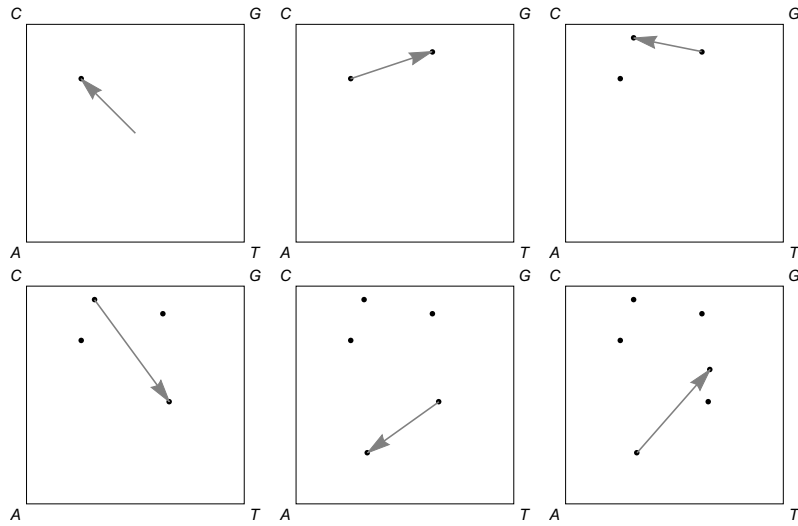
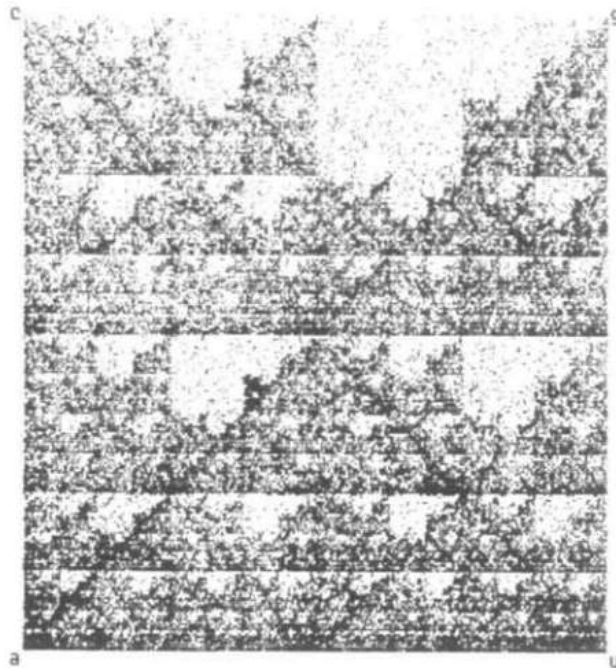Figure 1.1: CGRs of {C,G,C,T,A,G}. Made with Wolfram Mathematica

Figure 1.2: CGR of Human Beta Globin Region on Chromosome 11 (73 357 bases). Source: [1]

Barnsley and Leśniak investigated approximations of the attractor of an iterated function system with probabilities (see later in Section 2.5) with the use of stochastic process [3]. They showed that the sequences generated with the chaos game, converges to the attractor of the iterated function system (3.2).

We tried to represent it on DNA sequences, like Koslicki and Thompson in [2]. They used topological pressure to approach the nucleotide triplets' distribution, which is a special topological entropy, but this is out of our interest, we tried a more simple method. We considered the nucleotide sequences as Markov chains. We counted statistical probabilities. We can see the homogeneity of Mus musculus' (mouse) nucleotide sequence in Fig. 1.3.
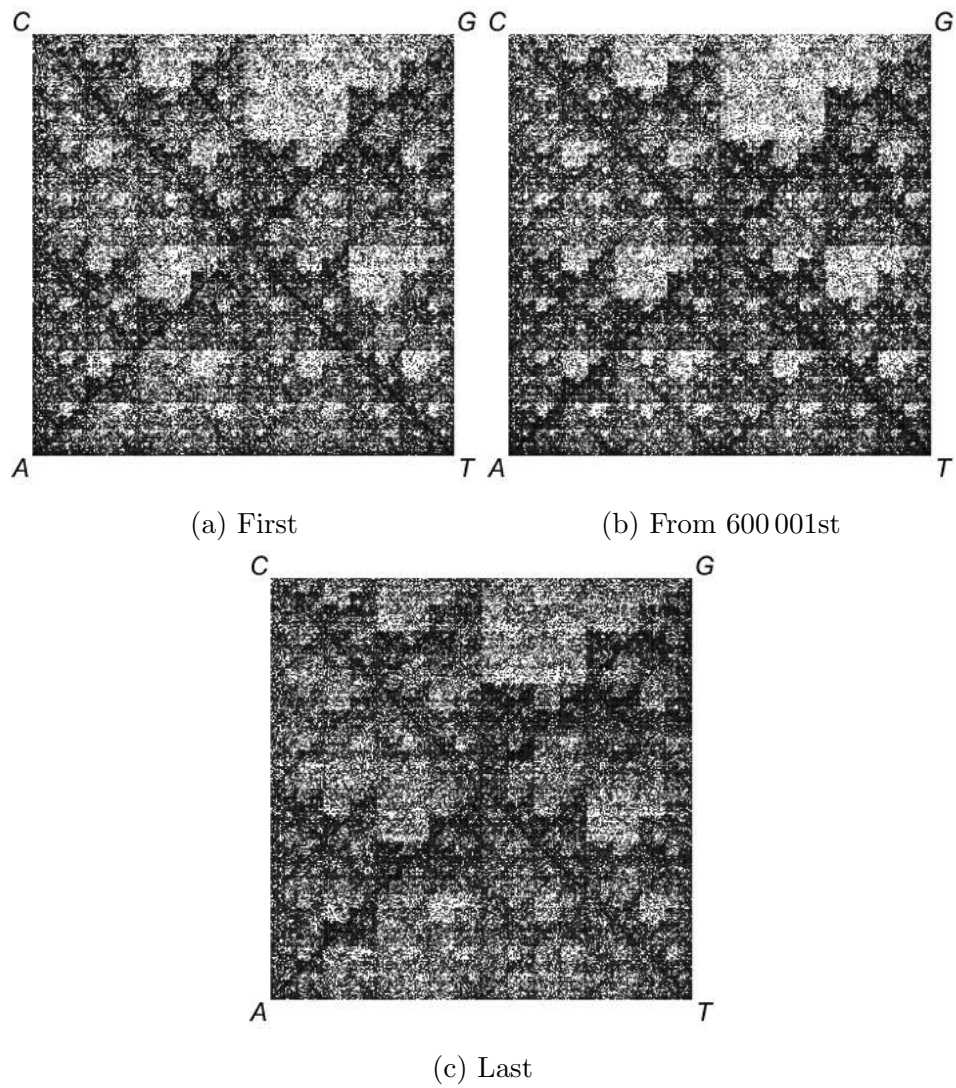
(a) First

(b) From 600 001st

(c) Last

Figure 1.3: CGR of Mus musculus (200 000 bases from different parts). Made with Wolfram Mathematica

Chapter 2

Theoretical background

The purpose of this section is to introduce the theory of the iterated function systems and Hausdorff metric, for which we generally follow the book [4].

2.1 Definition of IFS (Iterated Function System)

Definition 2.1. *Let* $(\mathbb{X}, d)$ *be a complete metric space. We call a function* $f : \mathbb{X} \to \mathbb{X}$ Lipschitz continuous *if there exists a real constant c such that* $d(f(x), f(y)) \le c\, d(x, y)$ *for all* $x, y \in \mathbb{X}$. *We denote the set of the Lipschitz continuous functions on* $\mathbb{X}$ *by* $\mathrm{Lip}(\mathbb{X})$.

We will use a special type of Lipschitz continuity, called contraction.

Definition 2.2. *Let* $(\mathbb{X}, d)$ *be a complete metric space. We call a function* $f : \mathbb{X} \to \mathbb{X}$ contraction *if there exists some* $c \in [0, 1)$ *with* $d(f(x), f(y)) \le c\, d(x, y)$ *for all* $x, y \in \mathbb{X}$. *The* contraction factor *for the contraction f is the smallest such constant c. We denote the set of these functions by* $\mathrm{Lip}_1(\mathbb{X})$.

Definition 2.3 (IFS)**.** *Let* $\mathbb{X}$ *be a metric space. Then an* iterated function system *on* $\mathbb{X}$ *is a finite collection of mappings* $f_i : \mathbb{X} \to \mathbb{X}$, $i = 0, 1, \ldots, N$, *which are usually contractive. If* c_i *is the conractivity of* f_i*, then the* contractivity *of the IFS is* $c := \max_i c_i$.

For example, consider the Cantor set C. It is a set of real numbers from $[0, 1]$, whose ternary expansion in base three does not contain the digit 1. The Cantor set's IFS $\{f_0, f_1\}$ on $\mathbb{X} = \mathbb{R}$, where $f_0(x) = x/3$, $f_1(x) = x/3 + 1/3$ and the contraction factor of the IFS is $1/3$ since $c_0 = c_1 = 1/3$.

The importance of an IFS lies in the fact that it formally encodes the idea of self-similarity. Let us just define what we mean by *similarity transform*. It is a mapping $f : \mathbb{X} \to \mathbb{X}$ with the property that $d(f(x), f(y)) = r\,d(x, y)$ for all x, y, where $r > 0$ is the *similarity ratio*. For example, each part of the Cantor set is clearly similar to the whole set with similiraty ratio $r = \frac{1}{3}$.

2.2 Space of compact subsets and the Hausdorff metric

Definition 2.4. *Let be*

$$\mathbb{H}(\mathbb{X}) = \{A \subseteq \mathbb{X} : A \neq \emptyset \text{ and } A \text{ is compact}\}.$$

In the following we define a metric on this space.

Definition 2.5 (Hausdorff distance)**.** *For $A, B \in \mathbb{H}(\mathbb{X})$, let us define*

$$d_{\mathbb{H}}(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\}.$$

Another way to define the Hausdorff metric is through the idea of an ε-dilation of a set.

Definition 2.6. *If A denotes a subset of a metric space $\mathbb{X}$ and $\varepsilon > 0$, then the ε-dilation of A is the set*

$$A_{\varepsilon} = \{x : d(x, a) < \varepsilon \text{ for some } a \in A\}.$$

Clearly $A \subseteq A_{\varepsilon}$ for any $\varepsilon > 0$.

Notice that

$$d(a, B) := \inf_{b \in B} d(a, b)$$

is the distance from the point a to the compact set B. Then, we can see that

$$d(A, B) := \sup_{a \in A} \inf_{b \in B} d(a, b)$$

is the least ε for which $A \subseteq B_\varepsilon$. Thus it is like a one-sided distance from B to A. The Hausdorff distance is a maximum of these one-sided distances, $d(A, B)$ and $d(B, A)$. Thus we proved the following proposition.

Proposition 2.7. *If $A, B \in \mathbb{H}(\mathbb{X})$, then the Hausdorff distance is*

$$d_{\mathbb{H}}(A, B) = \inf\{\varepsilon > 0 : A \subseteq B_\varepsilon \text{ and } B \subseteq A_\varepsilon\}.$$

Now we show that the Hausdorff distance is a metric.

Theorem 2.8. *If $(\mathbb{X}, d)$ is a metric, then so is the Hausdorff metric based on d.*

Proof. To verify this theorem, we have to prove the symmetry, the non-negativity and the triangle inequality.

(i) It is clear from the definition that $d_{\mathbb{H}}(A, B) = d_{\mathbb{H}}(B, A)$.

(ii) We can see that $d_{\mathbb{H}}(A, B) \geq 0$ for any $A, B \in \mathbb{H}(\mathbb{X})$. It is obvious that $d_{\mathbb{H}}(A, A) = 0$. On the other hand, suppose that $d_{\mathbb{H}}(A, B) = 0$, then we have $\inf_b d(a, b) = 0$ for each $a \in A$. It means that $a \in B$ because B is compact. Thus $A \subseteq B$. By similar arguments we get that $B \subseteq A$, thus $A = B$.

(iii) Let $A, B, C \in \mathbb{H}(\mathbb{X})$. We can see that

$$d(a, B) \leq \inf_{b \in B} (d(a, c) + d(c, b)) = d(a, c) + \inf_{b \in B} d(c, b) = d(a, c) + d(c, B)$$

for any $a \in A$ and $c \in C$. We can see that

$$d(a, B) \leq \inf_{c \in C} (d(a, c) + d(c, B)).$$

8

From this follows that

$$d(a, B) \leq \inf_{c \in C} d(a, c) + \sup_{c \in C} d(c, B) = d(a, C) + d(C, B).$$

Thus

$$d(A, B) = \sup_{a \in A} \inf_{b \in B} d(a, b) \leq d(A, C) + d(C, B).$$

In a similar manner we can see that $d(B, A) \leq d(B, C) + d(C, A)$, thus we get that $d_{\mathbb{H}}(A, B) \leq d_{\mathbb{H}}(A, C) + d_{\mathbb{H}}(C, B)$.

$\square$

Theorem 2.9. *Let $(\mathbb{X}, d)$ be complete. Then $(\mathbb{H}(\mathbb{X}), d_{\mathbb{H}})$ is also complete.*

Proof. We begin the proof by suggest that $A_n \in \mathbb{H}(\mathbb{X})$ is a Cauchy sequence, and define the set A by

$$A = \bigcap_{m=1}^{\infty} \overline{\bigcup_{n \geq m} A_n},$$

where we use the bar to denote closure. We wish to show that A_n converges, and for this we have to prove that $d_{\mathbb{H}}(A_n, A) \to 0$ and $A \in \mathbb{H}(\mathbb{X})$.

We show first that $A \in \mathbb{H}(\mathbb{X})$. We know that $\mathbb{X}$ is a complete metric space, so being compact is equivalent to being closed and totally bounded. Define

$$B_m = \overline{\bigcup_{n \geq m} A_n}.$$

We will show that each B_m is compact and nonempty, wich would mean that so is A. It is clear that each B_m is closed and $B_{m+1} \subseteq B_m$, so it will be sufficient to show that B_1 is totally bounded. Let $\varepsilon > 0$ be given. Since A_n is Cauchy, there is some m such that $d_{\mathbb{H}}(A_m, A_n) < \varepsilon/2$ for any $n > m$. It means that $A_n \subseteq (A_m)_{\varepsilon/2}$, where $(A_m)_{\varepsilon/2}$ is the $\varepsilon/2$-dilation of A_m. From this follows that $B_m \subseteq (A_m)_{\varepsilon/2}$, thus since A_m is totally bounded so is B_m. It is easy to see that

$$B_1 = \overline{(A_1 \cup A_2 \cup \cdots \cup A_{m-1} \cup B_m)},$$

so it is the closure of a finite union of totally bounded sets. This implies that B_1 is totally bounded, thus is compact.

Now we show that $d_{\mathbb{H}}(A_n, A) \to 0$. Let $\varepsilon > 0$ be given, then there is some m such that $d_{\mathbb{H}}(A_m, A_n) < \varepsilon/2$ for any $n > m$. This means that $A_n \subseteq (B_m)_{\varepsilon/2}$. Moreover, since $B_m \subseteq (A_m)_{\varepsilon/2}$ and $A_m \subseteq (A_n)_{\varepsilon/2}$, we have $B_m \subseteq (A_n)_\varepsilon$. From this follows that $A \subseteq (A_n)_\varepsilon$. To show the reverse containing, let $x \in A_n$. Then we have a sequence $x_k \in A_k$ with $d(x_k, x) < \varepsilon/2$ for all $k > n$. It means that $x_k \in B_n$, which is compact, so x_k has a cluster point $y \in B_n$ and $d(y, x) \leq \varepsilon/2$. The tail of x_k is in B_l for all $l > n$, so $y \in A$, which means that $A_n \subseteq A_\varepsilon$. Thus $d_{\mathbb{H}}(A, A_n) < \varepsilon$. $\qquad\square$

Any mapping $f : \mathbb{X} \to \mathbb{X}$ naturally induces another mapping $\hat{f} : \mathbb{H}(\mathbb{X}) \to \mathbb{H}(\mathbb{X})$, where $\hat{f}(A) = \{f(a) : a \in A\}$. Although formally these two functions are different, often we will not distinguish them. An IFS $\{f_0, f_1, \ldots, f_N\}$ induces a mapping on $\mathbb{H}(\mathbb{X})$:

$$F(S) = \bigcup_{i=0}^{N} \hat{f}_i(S).$$

Remark 2.10. *If $\{f_i\}$ is an IFS on $\mathbb{X}$ with contractivity $c < 1$, then the induced mapping F on $\mathbb{H}(\mathbb{X})$ is also contractive with contractivity c.*

Proof. Firstly, it is easy to see that if $f : \mathbb{X} \to \mathbb{X}$ is a contraction with contractivity $c < 1$, then $\hat{f} : \mathbb{H}(\mathbb{X}) \to \mathbb{H}(\mathbb{X})$ is also a contraction with contractivity c, because the one-sided distance satisfies the inequality $d(f(x), f(y)) \leq c\, d(x, y)$, so have $d(\hat{f}(A), \hat{f}(B)) \leq c\, d(A, B)$.

Now we will prove that

$$d_{\mathbb{H}}(A_1 \cup A_2, B_1 \cup B_2) \leq \max\{d_{\mathbb{H}}(A_1, B_1), d_{\mathbb{H}}(A_2, B_2)\}.$$

We can see that for any $C \in \mathbb{H}(\mathbb{X})$ we have

$$
\begin{aligned}
d(A_1 \cup A_2, C) &= \sup_{a \in A_1 \cup A_2} d(a, C) = \\
&= \max\{\sup_{a \in A_1} d(a, C), \sup_{a \in A_2} d(a, C)\} = \\
&= \max\{d(A_1, C), d(A_2, C)\}
\end{aligned}
$$

and we have

$$d(a, B_1 \cup B_2) = \inf_{b \in B_1 \cup B_2} d(a, b) =$$
$$= \min \{ \inf_{b \in B_1} d(a, b), \inf_{b \in B_2} d(a, b) \} =$$
$$= \min \{ d(a, B_1), d(a, B_2) \}.$$

From these follows, that

$$d_{\mathbb{H}}(A_1 \cup A_2, B_1 \cup B_2) =$$
$$= \max \{ \min \{ d(A_1, B_1), d(A_1, B_2) \}, \min \{ d(A_2, B_1, d(A_2, B_2) \},$$
$$\min \{ d(B_1, A_1), d(B_1, A_2) \}, \min \{ d(B_2, A_1), d(B_2, A_2) \} \} \leq$$
$$\leq \max \{ d(A_1, B_1), d(B_1, A_1), d(A_2, B_2), d(B_2, A_2) \} =$$
$$= \max \{ d_{\mathbb{H}}(A_1, B_1), d_{\mathbb{H}}(A_2, B_2) \}.$$

From this property of the Hausdorff distance follows the statement. $\qquad \square$

Theorem 2.11 (Contraction mapping theorem)**.** *Assume that* $(\mathbb{X}, d)$ *is a complete metric space and* $f : \mathbb{X} \to \mathbb{X}$ *is a contraction with contraction factor* $c < 1$. *Then there exists a unique* $\bar{x} \in \mathbb{X}$ *having the property that* $f(\bar{x}) = \bar{x}$. *Moreover, the sequence defined by* $x_{n+1} = f(x_n)$ *converges to* $\bar{x}$ *with the estimate*

$$d(x_n, \bar{x}) \leq c^n \, d(x_0, \bar{x}),$$

for any x_0.

The special strength of this theorem (which is also called *Banach's fixed-point theorem*) lies in the fact that that it not only guarantees a unique fixed point but also gives an effective construction to approximate this fixed point to any degree of accuracy – we just iterate the function sufficiently many times.

Corollary 2.12. *Every IFS induced mapping has a unique fixed point* A *in* $\mathbb{H}(\mathbb{X})$, *which satisfies the equation*

$$A = \bigcup_{i=0}^{N} f_i(A). \tag{2.1}$$

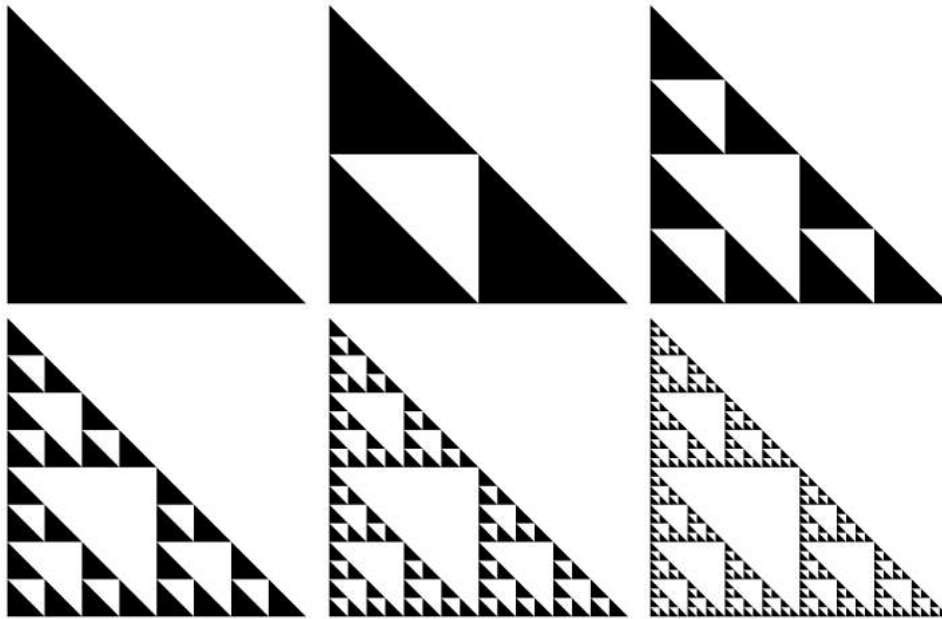*This fixed point is called the* attractor *of the IFS.*

Figure 2.1: The Sierpinski triangle's iteration's first few steps. Made with Wolfram Mathematica

The *Sierpinski triangle* is composed of three similar copies of itself, with similarity ratio $r = 1/2$. The three mappings are the following:

$$f_0\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x \\ y+1 \end{pmatrix}, \ f_1\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}, \ f_2\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x+1 \\ y \end{pmatrix}.$$

In Fig. 2.1 we can see the first few steps of the Sierpinski triangle's iteration. The initial set S_0 is a simple triangle, and it seems clear that the limit should be the Sierpinski triangle. That is the fixed point of the mapping F, so that F is a contraction on $\mathbb{H}(\mathbb{X})$.

2.3 Code space and the address map

There is a recursive "addressing scheme" which labels all the points of the attractor of an IFS. Let A denote the attractor of the IFS $\{f_0, f_1, \ldots, f_N\}$.

We can see that A satisfies the self-similarity identity

$$A = f_0(A) \cup f_1(A) \cup \cdots \cup f_N(A) = \bigcup_{i=0}^{N} f_i(A).$$

If we apply this decomposing again, we get the equation

$$A = f_0 \left(\bigcup_{i=0}^{N} f_i(A) \right) \cup f_1 \left(\bigcup_{i=0}^{N} f_i(A) \right) \cup \cdots \cup f_N \left(\bigcup_{i=0}^{N} f_i(A) \right).$$

Continuing, the nth decomposition is

$$A = \bigcup_{i_0=0}^{N} \bigcup_{i_1=0}^{N} \cdots \bigcup_{i_n=0}^{N} f_{i_0} \circ f_{i_1} \circ \cdots \circ f_{i_n}(A).$$

Definition 2.13 (Code space). *The* code space *of the IFS* $\{f_0, \ldots f_N\}$ *is the set*

$$\Sigma = \{0, 1, \ldots, N\}^{\mathbb{N}} = \{\bar{\imath} = \{i_0, i_1, \ldots\} : i_n \in \{0, 1, \ldots, N\}, n \in \mathbb{N}\}.$$

We define a metric on Σ:

$$d(\bar{\imath}, \bar{\jmath}) = \sum_{n=0}^{\infty} \frac{|i_n - j_n|}{(N+1)^n}$$

Definition 2.14 (Address map). *Let us fix* $x \in \mathbb{X}$. *The* address map *is* $\pi : \Sigma \to \mathbb{X}$ *where*

$$\pi(\bar{\imath}) = \lim_{n \to \infty} f_{i_0} \circ f_{i_1} \circ \cdots \circ f_{i_{n-1}} \circ f_{i_n}(x).$$

Let us introduce some notations. Sometimes we denote the space of sequences of length n by $\Sigma^n = \{0, 1, \ldots, N\}^n$,

$$A_{i_0 i_1 \ldots i_n} := f_{i_0} \circ f_{i_1} \circ \cdots \circ f_{i_n}(A),$$

and the truncation of $\bar{\imath}$ to the first $n+1$ terms by i^n. Thus we get the following notation:

$$A_{i^n} = f_{i^n}(A) = f_{i_0} \circ f_{i_1} \circ \cdots \circ f_{i_n}(A).$$
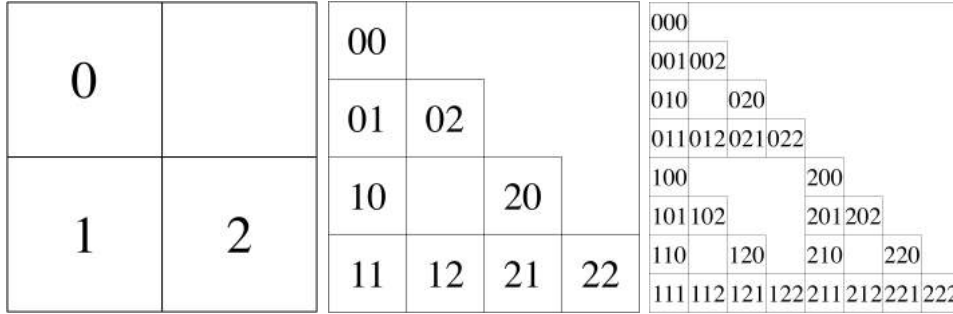
Figure 2.2: The Sierpinski triangle with addresses

If we have a point $x \in A$, there must be an $i_0 \in \{0, 1, \ldots, N\}$ such that $x \in A_{i_0}$. After this, there must be an i_1 such that $x \in A_{i_0 i_1}$, and continuing, we get a sequence i_n for each n such that $x \in A_{i^n}$. The longer sequence we know, the more precisely we know the location of the point x. This "addressing" of the points for the Sierpinski triangle is illustrated in Fig. 2.2.

The address map is independent from the chosen point. Let x, y be two different points, and let c be the contraction factor of the IFS. Then, for $\bar{\imath} \in \Sigma$,

$$d(f_{i_0} \circ \cdots \circ f_{i_n}(x), f_{i_0} \circ \cdots \circ f_{i_n}(y)) \leq c^{n+1} d(x, y).$$

It is clear that as $n \to \infty$, the distance goes to zero, thus the address map is independent of the chosen point x. Notice that:

$$A_{i_0 i_1 \ldots i_{n-1} i_n} = f_{i_0} \circ f_{i_1} \circ \cdots \circ f_{i_{n-1}} \circ f_{i_n}(A) \subseteq f_{i_0} \circ f_{i_1} \circ \cdots \circ f_{i_{n-1}}(A) = A_{i_0 i_1 \ldots i_{n-1}}.$$

Since $\lim_{n \to \infty} f_{i_0} \circ f_{i_1} \circ \cdots \circ f_{i_n}(x)$ is some single point in $\mathbb{R}^2$, it implies that

$$\bigcap_{n=0}^{\infty} A_{i_0 i_1 \ldots i_n} = \{z\},$$

where $z \in \mathbb{R}^2$.

Now we show that the range of π is contained in the attractor A of the IFS. We saw that π is independent of the chosen point x, so let x be such that $x \in A$. But then, because of the self-tiling equation (2.1), $f_{i_k}(x) \in A$

14

for each $i_k = 0, 1, \ldots, N$. This means that the limit will also be a point in A, thus $\pi(\Sigma) \subseteq A$.

Next we show that the address map is continuous and surjective. Fix an $\varepsilon > 0$, and let $\bar{\imath}, \bar{\jmath} \in \Sigma$. The contraction factor of the IFS is $c = \max_i c_i$. Then we know that there is some n such that $c^{n+1} \, diam(A) < \varepsilon$. Now we have to find a suitable δ. Let be δ such that $0 < \delta < \frac{N-1}{N(N+1)^{n+1}}$. Thus if $d(\bar{\imath}, \bar{\jmath}) < \delta$, then they have to agree at least in the first n terms. It means that $\pi(\bar{\imath}), \pi(\bar{\jmath}) \in A_{i^n}$. Since $diam(A_{i^n}) \leq c^{n+1} < \varepsilon$, we have $d(\pi(\bar{\imath}), \pi(\bar{\jmath})) < \varepsilon$, so π is continuous.

Note that the address map is a surjection. Let $x \in A$ be given. Then we know that there is a sequence $\bar{\imath} = i_0 i_1 \ldots$ such that $x \in A_{i_n}$ for each n. Let $y = \pi(\bar{\imath})$. Then $y \in A_{i^n}$ for each n because of the construction of the address map. It means that $d(x, y) \leq diam(A_{i^n}) \to 0$, thus $x = y$, so the address map is surjective.

2.4 Markov measures

For the better understanding, let us define a probability measure on the code space. In this part we will refer to the book [5], more exactly to Chapter 0 and Chapter 1.

Let X be a set and $\mathcal{B} \subseteq \mathcal{P}(X)$. $\mathcal{B}$ is a *σ-algebra*, if

(i) $X \in \mathcal{B}$;

(ii) $B \in \mathcal{B} \Rightarrow X \backslash B \in \mathcal{B}$;

(iii) $B_n \in \mathcal{B} \quad (n \geq 1) \Rightarrow \bigcup_{n=0}^{\infty} B_n \in \mathcal{B}$.

We call $(X, \mathcal{B})$ a *measurable space*. Let $\mu : \mathcal{B} \to \mathbb{R}^+$ be a function. It is a *finite measure* on $(X, \mathcal{B})$, if $\mu(\emptyset) = 0$ and $\mu(\bigcup B_n) = \sum \mu(B_n)$ where $(n = 1, 2, \ldots)$, $B_n \in \mathcal{B}$ for each n and they are pairwise disjoint. This $(X, \mathcal{B}, \mu)$ triple is called *probability space* if $\mu(X) = 1$.

Let $(X_i, \mathcal{B}_i, \mu_i)$ be a probability space for each $i \in \mathbb{N}$. Let $X = \prod_{i=0}^{\infty} X_i$ and the σ-algebra $\mathcal{B}$ be the product of the σ-algebras $\mathcal{B}_i$. Let μ be such that

$\mu(A_0 \times \cdots \times A_n) = \prod_{j=0}^{n} \mu_j(A_j)$ where $A_j \in \mathcal{B}_j$ for each j. Then we call the probability space $(X, \mathcal{B}, \mu)$ the *direct product* of the spaces $(X_i, \mathcal{B}_i, \mu_i)$.

Theorem 2.15. *Let $S = \{0, 1, \ldots, N\}$ and $(\Sigma, \mathcal{B}) = \prod_{k=0}^{\infty} (S, 2^S)$ is a measurable space. There is given a non-negative real number $p_n(i_0, \ldots, i_n)$ for each $n \in \mathbb{N}$ and $i_0, \ldots, i_n \in S$, which satisfies the following conditions:*

$$\sum_{i_0 \in S} p_0(i_0) = 1$$

and

$$p_n(i_0, \ldots, i_n) = \sum_{i_{n+1} \in S} p_{n+1}(i_0, \ldots, i_{n+1}).$$

Then there exists a unique probability measure μ on $(\Sigma, \mathcal{B})$ with $\mu([i_0, \ldots, i_m]) = p_m(i_0, \ldots, i_m)$ for all $i_k \in S$, $k \le m$.

Proof. It is a special case of the Kolmogorov consistency theorem, which guarantees that a suitably "consistent" collection of finite-dimensional distributions will define a stochastic process.:

Let X be a set, and I an index set. Suppose that there exists a probability measure $\mu_{t_0, \ldots, t_n}$ on X^{n+1} for each $t_0, \ldots, t_n \in I$, which satisfy two conditions:

   (i) $\mu_{t_0, \ldots, t_n}(A_0 \times \cdots \times A_n) = \mu_{t_{\pi(0)}, \ldots, t_{\pi(n)}}(A_{\pi(0)} \times \cdots \times A_{\pi(n)})$;

   (ii) $\mu_{t_0, \ldots, t_{n+m}}(A_0 \times \cdots \times A_n \times \underbrace{X \times \cdots \times X}_{m}) = \mu_{t_0, \ldots, t_n}(A_0 \times \cdots \times A_n)$

for any $A_i \in \mathcal{B}(X)$. Then there exists a unique probability space and a stochastic process $(X_t)_{t \ge 0}$ on it, where $(X_t)_{t \ge 0}$ is uniquely defined by the finite dimensional distributions given by the probability measures above.

   With $X = S$, $I = \mathbb{N}$ and $\mu_{t_0, \ldots, t_n} = p_n$ we get the theorem. $\qquad \square$

In the following we will show some measure-preserving transformations.

Definition 2.16 (Measure-preserving transformation). *Let $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ be two probability spaces, and $T : X_1 \to X_2$ is a transformation.*

   (i) T is measurable *if $B_2 \in \mathcal{B}_2 \Rightarrow T^{-1}B_2 \in \mathcal{B}_1$.*

(ii) T *is* measure-preserving *if T is measurable and $\mu_1(T^{-1}B_2) = \mu_2(B_2)$ for each $B_2 \in \mathcal{B}_2$.*

(iii) T *is an* invertible measure-preserving transformation *if T is measure-preserving, bijective, and T^{-1} is also measure-preserving.*

Let $S = \{0, 1, \ldots, N\}$ and $(X, \mathcal{B}) = \prod_{k=0}^{\infty} (S, 2^S)$. Define the shift transformation $T : X \to X$ by $T((i_0, i_1, i_2, \ldots)) = (i_1, i_2, \ldots)$.

Let S be a finite set $\{0, 1, \ldots, N\}$, $\mathcal{B}$ the σ-algebra generated by 2^S, and $\Sigma = S^{\mathbb{N}}$ the code space. Let $(\Sigma, \mathcal{B}, \mu)$ be a probability space, which is the direct product of $(S, 2^S, \nu)$. We call $A \subset \Sigma$ a *measurable rectangle* if $A = A_0 \times \cdots \times A_n = \{\bar{\imath} \in \Sigma \mid i_k \in A_k \ \forall k \le n\}$. Now let us define a measure on Σ, given by a probability vector $(p_0, \ldots, p_N)$ where $p_j = \nu(\{j\})$:

$$\mu([i_0, i_1, \ldots, i_n]) = \prod_{k=0}^{n} p_{i_k}$$

where $[i_0, i_1, \ldots, i_n] := \{\bar{\jmath} \in \Sigma : j_0 = i_0, \ldots, j_n = i_n\}$. This μ is called the $(p_0, \ldots, p_N)$-*product measure*.

An example of the measure-preserving transformations provided by the Markov chains is the $(\boldsymbol{p}, P)$-*Markov shift*. Here $\boldsymbol{p} = (p_0, p_1, \ldots, p_N)$ is a probability vector, and $P = (P_{ij})_{i,j \in S}$ is a stochastic matrix $(P_{ij} \ge 0, \sum_{j=0}^{N} P_{ij} = 1)$. The matrix is such that $\sum_{i=0}^{N} p_i P_{ij} = p_j$ and the probability of a sequence is $p(i_0, \ldots, i_n) = p_{i_0} P_{i_0 i_1} \cdots P_{i_{n-1} i_n}$.

In the similar manner we can define the *n-step Markov shift*, but here the probability vector is $\boldsymbol{p} = (p_{0 \cdots 0}, p_{0 \cdots 1}, \ldots, p_{N \cdots N})$ and the stochastic matrix is $P = |S|^n \times |S|^n$. The probability of a sequence is $p_n(i_0, \ldots, i_m) = p_{i_0 \cdots i_n} P_{i_0 \cdots i_n, i_1 \cdots i_{n+1}} P_{i_1 \cdots i_{n+1}, i_2 \cdots i_{n+2}} \cdots P_{i_{m-n-1} \cdots i_{m-1}, i_{m-n} \cdots i_m}$.

Definition 2.17. *Let $(X, \mathcal{B}, \mu)$ be a probability space, and $T : X \to X$ a measure-preserving transformation. We call T* ergodic *if every $B \in \mathcal{B}$ with $T^{-1}B = B$ satisfies $\mu(B) = 0$ or $\mu(B) = 1$.*

Theorem 2.18. *Let $(X, \mathcal{B}, \mu)$ be a probability space. If $T : X \to X$ is a measure-preserving transformation then the following statements are equivalent:*

(i) T is ergodic.

(ii) For every $A, B \in \mathcal{B}$ with $\mu(A), \mu(B) > 0$ there is an $n > 0$ with $\mu(T^{-n}A \cap B) > 0$.

Proof. $(i) \Rightarrow (ii)$. Assume $A, B \in \mathcal{B}$ with $\mu(A), \mu(B) > 0$ is such that $\mu(T^{-n}A \cap B) = 0$, $\forall n > 0$. Then we have $\mu((\bigcup_{n=0}^{\infty} T^{-n}A) \cap B) = 0$. Let be $I = \bigcup_n T^{-n}A$, it is clear that $T^{-1}I \subseteq I$. Let be $J = \bigcap_{k=0}^{\infty} T^{-k}I$, then J is T-invariant, because

$$T^{-1}J = \bigcap_{k=1}^{\infty} T^{-k}I \supseteq J = \bigcap_{k=0}^{\infty} T^{-k}I = J$$

and

$$T^{-1}J = \bigcap_{k=0}^{\infty} T^{-k-1}I \subseteq \bigcap_{k=0}^{\infty} T^{-k}I = J.$$

Thus $\mu(J) = 0$ or 1. Since $\mu(I \cap B) = 0$, we have $\mu(T^{-n}I \cap B) = 0$ for any n, so that $\mu(J \cap B) = 0$. Then $\mu(J) = 0$, and it means that

$$\exists k \quad \mu(\bigcup_{n=k}^{\infty} T^{-n}A) < \varepsilon,$$

which implies

$$\mu(\bigcup_{n=0}^{\infty} T^{-n}A) < \varepsilon$$

because of the T-invariance. From this follows that $\mu(\bigcup T^{-n}A) = 0$, thus $\mu(A) = 0$, which contradicts the assumptions.

$(ii) \Rightarrow (i)$. Assume $B \in \mathcal{B}$ with $T^{-1}B = B$ is such that $0 < \mu(B) < 1$. Then $0 = \mu(B \cap (X \backslash B)) = \mu(T^{-n}B \cap (X \backslash B))$ for all $n \geq 1$, which contradicts the assumptions. $\qquad \square$

Theorem 2.19. *Let $(X, \mathcal{B}, \mu)$ be a probability space. If $T : X \to X$ is a measure-preserving transformation then the following statements are equivalent:*

(i) T is ergodic.

(ii) For every measurable f with $(f \circ T)(x) = f(x)$ a.e., f is constant a.e.

(iii) For every $f \in L^2$ with $(f \circ T)(x) = f(x)$ a.e., f is constant a.e.

Proof. $(i) \Rightarrow (ii)$. Suppose that T is ergodic, f is measurable and $(f \circ T)(x) = f(x)$ a.e. Let be

$$S(k, n) = \left\{ x : \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \right\} \quad \forall k \in \mathbb{Z}, n > 0.$$

Then we have

$$T^{-1}S(k, n) \triangle S(k, n) \subset \{x : (f \circ T)(x) \neq f(x)\}.$$

Thus $\mu(T^{-1}S(k, n) \triangle S(k, n)) = 0$ so that $\mu(S(k, n)) = 0$ or 1. We have $\bigcup_{k \in \mathbb{Z}} S(k, n) = X$ which is a disjoint union, so $\forall n \ \exists! k_n$ with $\mu(S(k_n, n)) = 1$. From this follows that $\mu(\bigcap_{n=1}^{\infty} S(k_n, n)) = 1$ and f is constant on it so that f is constant a.e.

$(ii) \Rightarrow (iii)$ is trivial, because if f is measurable, then $f \in L^2$.

$(iii) \Rightarrow (i)$. Let be $B \in \mathcal{B}$ such that $T^{-1}B = B$. Since $\chi_B \in L^2$ and $(\chi_B \circ T)(x) = \chi_B(x) \ \forall x \in X$, by the assumption we have that χ_B is constant a.e. From this follows that $\chi_B = 0$ or 1 a.e., and thus $\mu(B) = \int \chi_B \, d\mu = 0$ or 1. $\qquad \square$

Theorem 2.20. *The $(p_0, \ldots, p_N)$-shift is ergodic.*

Proof. Suppose that $E \in \mathcal{B}$ is such that $T^{-1}E = E$. We will show that then $\mu(E) = 0$ or $\mu(E) = 1$.

Denote the algebra of all finite unions of measurable rectangles by $\mathcal{A}$, let be $\varepsilon > 0$ given and $A \in \mathcal{A}$ such that $\mu(E \triangle A) < \varepsilon$. We can see that

$$|\mu(E) - \mu(A)| = |\mu(E \setminus A) + \mu(E \cap A) - \mu(A \setminus E) - \mu(A \cap E)|$$
$$< \mu(E \setminus A) + \mu(A \setminus E) < \varepsilon.$$

Let n be so large that $B = T^{-n}A$ depends upon different coordinates from A. Then $\mu(B \cap A) = \mu(B)\mu(A) = \mu(A)^2$ since μ is a product measure, and we have

$$\mu(E \triangle B) = \mu(T^{-n}E \triangle T^{-n}A) = \mu(E \triangle A) < \varepsilon.$$

We can see that $\mu(E\triangle(A\cap B)) < 2\varepsilon$ because we know that $E\triangle(A\cap B) \subset (E\triangle A)\cup(E\triangle B)$. Thus

$$
\begin{aligned}
|\mu(E) - \mu(A\cap B)| &= |\mu(E\setminus(A\cap B)) + \mu(E\cap(A\cap B)) \\
&\quad -\mu((A\cap B)\setminus E) - \mu((A\cap B)\cap E)| < \\
&< \mu(E\setminus(A\cap B)) + \mu(A\setminus(A\cap B)) < 2\varepsilon.
\end{aligned}
$$

and

$$
\begin{aligned}
|\mu(E) - \mu(E)^2| &\leq |\mu(E) - \mu(A\cap B)| + |\mu(A\cap B) - \mu(E)^2| < \\
&< 2\varepsilon + |\mu(A)^2 - \mu(E)^2| \leq \\
&\leq 2\varepsilon + \mu(A)|\mu(A) - \mu(E)| + \mu(E)|\mu(A) - \mu(E)| < \\
&< 4\varepsilon.
\end{aligned}
$$

From this follows that $\mu(E) = \mu(E)^2$, which implies that $\mu(E) = 0$ or 1. $\square$

Theorem 2.21. *Let T be the $(\boldsymbol{p}, P)$ Markov shift. Then T is ergodic iff P is irreducible (i.e. $\forall i, j\ \exists n > 0$ with $p_{i,j}^{(n)} > 0$ where $p_{i,j}^{(n)}$ is the (i,j)-entry of the matrix P^n).*

For the proof see [5, p.42].

Theorem 2.22 (Birkhoff)**.** *Let $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ be measure-preserving and $f \in L^1(\mu)$. Then $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$ converges a.e. to a function $f^* \in L^1(\mu)$, $f^* \circ T = f^*$ and if $\mu(X) < \infty$, then $\int f^*\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$.*

Theorem 2.23 (L^p Ergodic Theorem of Von Neumann)**.** *Let $(X, \mathcal{B}, \mu)$ be a probability space, $T : X \to X$ a measure-preserving transformation and $1 \leq p < \infty$. If $f \in L^p(\mu)$ then there exists $f^* \in L^p(\mu)$ such that $f^* \circ T = f^*$ a.e. and*

$$
\left\| \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) - f^*(x) \right\|_p \to 0.
$$

For the proof see [5, p.36]

Definition 2.24. *Let $(X, \mathcal{B}, \mu)$ be a probability space and $T : X \to X$ a measure-preserving transformation. Then T is* strong-mixing *if $\forall A, B \in \mathcal{B}$*

$$\lim_{n \to \infty} \mu(T^{-n} A \cap B) = \mu(A)\mu(B).$$

Remark 2.25. *Every strong-mixing transformation is ergodic, because if $\{a_n\}$ is a sequence of real numbers then $\lim_{n \to \infty} a_n = 0$ implies*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n-1} a_i = 0.$$

Theorem 2.26. *The $(p_0, \dots, p_N)$-shift is strong-mixing.*

Proof. Let $A, B \in \mathcal{B}$. There exist C_1, C_2 finite unions of rectangles such that $\mu(A \triangle C_1) < \varepsilon$ and $\mu(B \triangle C_2) < \varepsilon$. Thus it is enough to show that $\mu(T^{-n} C_1 \cap C_2) \to \mu(C_1)\mu(C_2)$. Let be $\bar{\imath} \in C_1$ and $\bar{\jmath} \in C_2$. We can see that

$$\mu(T^{-n}[i_0, i_1, \dots, i_k] \cap [j_0, j_1, \dots j_l]) = p_{i_0} \cdots p_{i_k} \cdot 1 \cdots 1 \cdot p_{j_0} \cdots p_{j_l}$$

if $n > l + 1$. Thus if $n \to \infty$, then $\mu(T^{-n} C_1 \cap C_2) \to \mu(C_1)\mu(C_2)$. $\qquad \square$

Theorem 2.27. *Let T be the $(\boldsymbol{p}, P)$ Markov shift. Then the following statements are equivalent:*

  (i) T is strong-mixing.

  (ii) P is irreducible and aperiodic.

  (iii) $P_{ij}^{(n)} \to p_j$ for all i, j.

For the proof see [5, p.51].

2.5   IFS with probabilities

In this section, we define an IFS whose attractor will no longer be a set, but an analytic object, a Borel probability measure. This will be the basic structure used in the chaos game. We will refer again to the book [4].

Definition 2.28 (IFS with probabilities)**.** *A finite collection of contractions* $\{f_0, \ldots, f_N\}$ *on the complete metric space* $\mathbb{X}$ *along with a finite collection of probabilities* $\{p_0, \ldots, p_N\}$ *(where* $p_i \geq 0$ *and* $\sum_i p_i = 1$*) is called an* IFS with probabilities*.*

For each IFSP there is an *invariant measure* (also called fractal measure or self-similar measure). This measure can be given as a fixed point of a contraction called the *Markov operator*.

Definition 2.29 (IFSP Markov operator)**.** *Let* $\mathbb{X}$ *be a metric space. Denote the space of all Borel probability measures on* $\mathbb{X}$ *by* $\mathcal{P}(\mathbb{X})$, *and let* $\{f_i, p_i\}$ *be an* N*-map IFSP. The* Markov operator $\mathbb{M} : \mathcal{P}(\mathbb{X}) \to \mathcal{P}(\mathbb{X})$ *associated with this IFSP is*

$$(\mathbb{M}\mu)(B) = \sum_{i=0}^{N} p_i \mu(f_i^{-1}(B))$$

for each Borel set $B \subseteq \mathbb{X}$*.*

Notice that $\mathbb{M}\mu$ is a combination of "smaller" and "distorted" copies of μ, as $\mathbb{M}$ uses the maps f_i to "contract" and "rotate".

We are looking for the measures such that $\mathbb{M}\mu = \mu$. Let $\mathcal{M}(\mathbb{X})$ denote the space of all measures on $\mathbb{X}$. We can see that $\mathbb{M}$ is linear on it, thus if $\mathbb{M}$ is contractive on the whole $\mathcal{M}(\mathbb{X})$, then the only possible fixed point would be the zero measure. To get a nontrivial solution, we will restrict our attention to a subspace, the probability measures.

The Markov operator maps a probability measure to another one, so it is a well-defined map. Now we need to define a metric on $\mathcal{P}(\mathbb{X})$ for $\mathbb{M}$ to be a contraction. This metric will be the Monge-Kantorovich metric. First we define a subspace of the probability measures.

Definition 2.30. *Let* $a \in \mathbb{X}$*. Let us define a subspace of the probability measure* $\mathcal{P}(\mathbb{X})$*:*

$$\mathcal{P}_1(\mathbb{X}) = \left\{ \mu \in \mathcal{P}(\mathbb{X}) : \int_{\mathbb{X}} d(a, x) \, \mathrm{d}\mu(x) < \infty \right\}.$$

We should remark that from the triangle inequality and the fact that μ is a probability measure, we get that if the integral above is finite for some given a, then it is also finite for any other fixed b.

Definition 2.31 (Monge-Kantorovich metric). *Let $\mathbb{X}$ be compact and let $\mu, \nu \in \mathcal{P}_1(\mathbb{X})$. The* Monge-Kantorovich metric *is given as*

$$d_{MK}(\mu, \nu) = \sup \left\{ \left| \int_{\mathbb{X}} g \, d\mu - \int_{\mathbb{X}} g \, d\nu \right| : g \in \mathrm{Lip}_1(\mathbb{X}) \right\}.$$

It is obvious that $d_{\mathrm{MK}}(\mu, \nu) = d_{\mathrm{MK}}(\nu, \mu)$ and $d_{\mathrm{MK}}(\mu, \nu) = 0$ if $\mu = \nu$. Let be $d_{\mathrm{MK}}(\mu, \nu) = 0$. It means that $\int_{\mathbb{X}} g \, d\mu = \int_{\mathbb{X}} g \, d\nu \ \forall g \in \mathrm{Lip}_1(\mathbb{X})$, so that it is true for Lipschitz functions with any Lipschitz constant. Then, since they are dense in the Banach space of continuous functions, there exists a subsequence g_n of g, that $g_n \to \chi_A$, thus $\mu(A) = \nu(A)$, for every A open set. It implies that $\mu = \nu$ on the σ-algebra generated by open sets.

The triangle inequality is also clear:

$$d_{\mathrm{MK}}(\mu_1, \mu_2) = \sup \left\{ \left| \int_{\mathbb{X}} g \, d\mu_1 - \int_{\mathbb{X}} g \, d\mu_2 \right| \right\} \leq$$

$$\leq \sup \left\{ \left| \int_{\mathbb{X}} g \, d\mu_1 - \int_{\mathbb{X}} g \, d\nu \right| + \left| \int_{\mathbb{X}} g \, d\nu - \int_{\mathbb{X}} g \, d\mu_2 \right| \right\} \leq$$

$$\leq \sup \left\{ \left| \int_{\mathbb{X}} g \, d\mu_1 - \int_{\mathbb{X}} g \, d\nu \right| \right\} + \sup \left\{ \left| \int_{\mathbb{X}} g \, d\nu - \int_{\mathbb{X}} g \, d\mu_2 \right| \right\} =$$

$$= d_{\mathrm{MK}}(\mu_1, \nu) + d_{\mathrm{MK}}(\nu, \mu_2).$$

Thus d_{MK} is metric.

Now let prove the completeness. Suppose that $(\mu_n)_{n \geq 1}$ is a Cauchy sequence in d_{MK}. There exists a subsequence $(\mu_{n_k})_{k \geq 1}$ and a μ such that

$$\int f \, d\mu_{n_k} \to \int f \, d\mu \text{ for all continuous } f.$$

Let be $\varepsilon > 0$. Since $(\mu_n)_{n \geq 1}$ is Cauchy in d_{MK}, there exists an $N(\varepsilon)$ such that for all $n, n_k \geq N(\varepsilon)$ we have

$$\left| \int_{\mathbb{X}} f \, \mathrm{d}\mu_{n_k} - \int_{\mathbb{X}} f \, \mathrm{d}\mu_n \right| < \varepsilon \text{ for all } f \in \mathrm{Lip}_1(\mathbb{X}).$$

If $k \to \infty$, we get

$$\left| \int_{\mathbb{X}} f \, \mathrm{d}\mu - \int_{\mathbb{X}} f \, \mathrm{d}\mu_n \right| \leq 2\varepsilon.$$

It means that $d_{\mathrm{MK}}(\mu, \mu_n) \to 0$.

Theorem 2.32 (Completeness of space of probability measures)**.** *If $\mathbb{X}$ is a complete, separable metric space then $\mathcal{P}_1(\mathbb{X})$ is a complete space under the Monge-Kantorovich metric. Moreover, if $\mathbb{X}$ is compact, then $\mathcal{P}(\mathbb{X}) = \mathcal{P}_1(\mathbb{X})$ and both are compact under the Monge-Kantorovich metric.*

For the proof see [4, p.53].

Theorem 2.33. *Let $\{f_i, p_i\}$ be an IFSP and $c := \max_i c_i$ where c_i is the contraction factor of f_i. The IFSP Markov operator satisfies the following inequality:*

$$d_{MK}(\mathbb{M}\mu_1, \mathbb{M}\mu_2) \leq \left(\sum_i p_i c_i \right) d_{MK}(\mu_1, \mu_2).$$

From this follows that if $\sum_i p_i c_i < 1$, then $\mathbb{M}$ is a contraction on $(\mathcal{P}_1(\mathbb{X}), d_{MK})$ and there exists a unique Borel probability measure $\mu \in \mathcal{P}_1(\mathbb{X})$ such that

$$\mathbb{M}(\mu) = \mu$$

or

$$\mu(B) = \sum_i p_i \mu(f_i^{-1}(B)) \text{ for every } B \subseteq \mathbb{X} \text{ Borel set.}$$

Proof. Let $g \in \mathrm{Lip}_1(\mathbb{X}, \mathbb{R})$. Then

$$\left| \int_{\mathbb{X}} g(x) \, \mathrm{d}(\mathbb{M}\mu_1)(x) - \int_{\mathbb{X}} g(x) \, \mathrm{d}(\mathbb{M}\mu_2)(x) \right| =$$

$$= \left| \int_{\mathbb{X}} g(x) \sum_i p_i \, \mathrm{d}(\mu_1 \circ f_i^{-1})(x) - \int_{\mathbb{X}} g(x) \sum_i p_i \, \mathrm{d}(\mu_2 \circ f_i^{-1})(x) \right| =$$

$$= \left| \int_{\mathbb{X}} \sum_i p_i g(f_i(y)) \, \mathrm{d}\mu_1(y) - \int_{\mathbb{X}} \sum_i p_i g(f_i(y)) \, \mathrm{d}\mu_2(y) \right|$$

Let $\hat{g} := \sum_i p_i g \circ f_i$.

$$|\hat{g}(x) - \hat{g}(y)| = \left| \sum_i p_i \left(g(f_i(x)) - g(f_i(y)) \right) \right| \leq$$

$$\leq \sum_i p_i \left| g(f_i(x)) - g(f_i(y)) \right| \leq$$

$$\leq \sum_i p_i \cdot 1 \cdot |f_i(x) - f_i(y)| \leq$$

$$\leq \left(\sum_i p_i c_i \right) |x - y|.$$

We can see the $\hat{g}$ is Lipschitz with factor $\sum_i p_i c_i$ for any $g \in \mathrm{Lip}_1$. Thus we get

$$\sup_{g \in \mathrm{Lip}_1} \left\{ \left| \int_{\mathbb{X}} \sum_i p_i g(f_i(y)) \, \mathrm{d}\mu_1(y) - \int_{\mathbb{X}} \sum_i p_i g(f_i(y)) \, \mathrm{d}\mu_2(y) \right| \right\} =$$

$$= \sup_{\hat{g} \in \mathrm{Lip}_{(\sum_i p_i c_i)}} \left\{ \left| \int_{\mathbb{X}} \hat{g}(y) \, \mathrm{d}\mu_1(y) - \int_{\mathbb{X}} \hat{g}(y) \, \mathrm{d}\mu_2(y) \right| \right\} \leq$$

$$\leq \sum_i p_i c_i \cdot \sup_{\hat{g}} \left\{ \left| \int_{\mathbb{X}} \frac{\hat{g}(y)}{\sum_i p_i c_i} \, \mathrm{d}\mu_1(y) - \int_{\mathbb{X}} \frac{\hat{g}(y)}{\sum_i p_i c_i} \, \mathrm{d}\mu_2(y) \right| \right\} =$$

$$= \sum_i p_i c_i \cdot \sup_{g \in \mathrm{Lip}_1} \left\{ \left| \int_{\mathbb{X}} g(y) \, \mathrm{d}\mu_1(y) - \int_{\mathbb{X}} g(y) \, \mathrm{d}\mu_2(y) \right| \right\}$$

From this follows the theorem. $\qquad\qquad\square$

In order to see the connection between the invariant measure μ of an IFSP and the geometric attractor of the IFS (without the probabilities), we define the *support of a measure*.

Definition 2.34 (Support of a measure)**.** *The* support *of the measure μ on $\mathbb{X}$ is*

$$supp(\mu) = \mathbb{X} \setminus \bigcup \{U : U \text{ is an open set with } \mu(U) = 0\}.$$

Theorem 2.35. *Let $\{f_i, p_i\}$ be an IFSP with contraction factors c_i, where $p_i > 0$ and $c_i < 1$ for each i, and let μ be the invariant measure. Then the support of μ is equal to the attractor A of the IFS $\{f_i\}$.*

For the proof see [4, p.56].

The following theorem gives the connection between the invariant measure μ and an appropriate measure P on the codespace Σ. Before we state this theorem let us define the measure P on Σ as the product measure given by the measure $\mathbb{P}(x = i) = p_i$ on each factor $\{0, 1, \ldots, N\}$. We know that the address map π is a continuous mapping from Σ to $\mathbb{X}$. Now let us define the push-forward measure of P:

Definition 2.36. *The* push-forward measure *of P via π is the measure $\pi_\#(P)$ defined on $\mathbb{X}$ and given by $\pi_\#(P)(B) = P(\pi^{-1}(B))$ for each $B \subseteq \mathbb{X}$ Borel set. It is a Borel measure since π is continuous, thus the preimage of a Borel set is a Borel set.*

Theorem 2.37. *The invariant measure μ of an IFSP $\{f_i, p_i\}$ is the push-forward measure $\pi_\#(P)$ where P is the natural product measure defined on the code space Σ via the address map $\pi : \Sigma \to \mathbb{X}$.*

Proof. Define $\tau_k : \Sigma \to \Sigma$ by $\tau_k(i_0, i_1, \ldots) = (k, i_0, i_1, \ldots)$. Then it is clear that $\pi \circ \tau_k = f_k \circ \pi$ and thus $\tau_k^{-1} \circ \pi^{-1} = \pi^{-1} \circ f_k^{-1}$. By the definition of the product measure, we have $P(\tau_k(B)) = p_k P(B)$ for any $B \subseteq \Sigma$. Hence we

can see that

$$\mathbb{M}\pi_\#(P)(B) = \sum_k p_k \pi_\#(P)(f_k^{-1}(B)) = \sum_k p_k P(\pi^{-1}(f_k^{-1}(B))) =$$

$$= \sum_k p_k P(\tau^{-1}(\pi^{-1}(B))) =$$

$$= \sum_k P(\tau_k(\tau_k^{-1}(\pi^{-1}(B)))) = \sum_k P(\pi^{-1}(B)) =$$

$$= P(\pi^{-1}(B)) = \pi_\#(P)(B).$$

Thus $\pi_\#(P)$ is invariant under $\mathbb{M}$, so it must be equal to μ. □

27

Chapter 3

Chaos game

In this part we will refer to the articles [3],[6].

We have seen one way to generate the attractor of an IFS: start from an initial set and iterate the mapping on $\mathbb{H}(\mathbb{X})$ which is induced by the IFS. In this mode we composed the IFS maps from the "outside" to the "inside": $f_{i_1} \circ \cdots \circ f_{i_n}(x) \to \pi(\bar{\imath})$. Another method is the chaos game, where the composition of the maps is in the opposite order. It generates a random sequence of points that approximate the attractor. The algorithm is the following:

(i) Choose an initial point $x_0 \in \mathbb{X}$.

(ii) Choose $i_1 \in \{1, 2, \ldots, N\}$ with the probabilities $\{p_1, \ldots, p_N\}$, compute $x_1 = f_{i_1}(x_0)$ and plot it.

(iii) For the general step, choose $i_{n+1} \in \{1, 2, \ldots, N\}$ with the probabilities, compute $x_{n+1} = f_{i_{n+1}}(x_n) = f_{i_{n+1}} \circ \cdots \circ f_{i_1}(x_0)$ and plot it.

Often, the first few points are not between the points of the attractor. One solution is to wait some number of iteration before plotting the points, another is to start with $x_0 \in A$ (the fixed point of f_1 will be a suitable choice).

Theorem 3.1. *Let $\mathbb{X}$ be a compact metric space and $\{f_i, p_i\}$ be a contractive IFSP. Choose some $x_0 \in \mathbb{X}$, and define the sequence x_n by selecting $\bar{\imath} \in \Sigma$*

according to the probability measure P and setting

$$x_n = f_{i_n} \circ \cdots \circ f_{i_1}(x_0).$$

Let μ be the invariant measure for the IFSP. Then we have

$$\lim_n \frac{1}{n} \sum_{k \leq n} g(x_k) = \int_{\mathbb{X}} g(x) \, \mathrm{d}\mu(x)$$

for any continuous $g : \mathbb{X} \to \mathbb{R}$ and for P almost all $\bar{\imath} \in \Sigma$.

Sometimes the restriction to continuous functions is not convenient. With a complete $\mathbb{X}$ metric space the theorem is true for every $g \in L^1(\mu)$.

For the proof see [4, p.66].

Corollary 3.2. *Let $\mathbb{X}$ be a complete metric space and $\{f_i, p_i\}$ be a contractive IFSP. Then for all B Borel sets, the sequence generated by the chaos game, $x_n = f_{i_n} \circ \cdots \circ f_{i_1}(x_0)$, satisfies*

$$\mu(B) = \lim_n \frac{1}{n} \#\{1 \leq k \leq n : x_k \in B\} \tag{3.1}$$

for μ almost all x_0 and for P almost all $\bar{\imath} \in \Sigma$.

Corollary 3.3. *Let $\{f_i\}$ be an IFS with the attractor A, and let $\{x_n\}$ be the sequence defined by the chaos game above. Then*

$$\bigcap_{N=1}^{\infty} \overline{\bigcup_{k=N}^{\infty} \{x_k\}} = A. \tag{3.2}$$

Proof. The $\bigcap \overline{\bigcup\{x_k\}} \subseteq A$ containing is clear, since A is compact and $x_0 \in A$. For the other direction, let $x \in A$. Then $\forall \delta \ \mu(B_\delta(x)) > 0$ because of Theorem 2.35. From this and (3.1) follows that $x_k \in B_\delta(x)$ infinitely often. It means that x is an accumulation point of $\bigcup\{x_k\}$, thus

$$x \in \bigcap_{N=1}^{\infty} \overline{\bigcup_{k=N}^{\infty} \{x_k\}}.$$

$\square$

The chaos game representation of an n-step Markov chain is the following:

(i) Choose an initial point $x_0 \in \mathbb{X}$.

(ii) Choose $i_1, \ldots, i_n$ with the transition probabilities $\{p_{1\ldots1}, \ldots, p_{N\ldots N}\}$, compute $x_1 = f_{i_1}(x_0)$ and plot it.

(iii) For the general step, choose x_{k+1} with the transition probabilities $P_{i_{k-n+1}\ldots i_k, i_{k+2-n}\ldots i_{k+1}}$, compute $x_{k+1} = f_{i_{k+1}}(x_k)$ and plot it.

Chapter 4

DNA representation

At [7], there are available the nucleotide sequences of many species (both animals and plants), coded in FASTA, which is a text-based format. Every genome sequence contains many gene sequences. They are separated with header lines, which include the names of the sequences and some additional information. We dropped these parts, thus we got one long list from the acid bases. In FASTA format they use not only 'A', 'C', 'T' and 'G', but for the unidentified nucleic acids there are other letters, for example 'N', which means that it is one of the previous four. Since there is negligible amount of them, we dropped these letters from the sequences. In the Mus musculus' genome sequence, the number of unidentified nucleic acids are less than 0.1%.
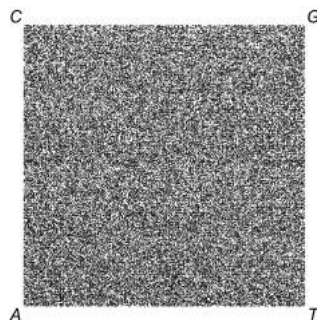

Figure 4.1: 0-step approximation of Mus musculus with Markov chains. Made with Wolfram Mathematica

In the approximation of the Mus musculus' sequence, the 0-step Markov chain is not suitable because the numbers of the four bases are roughly equal, so we get a homogene square, as we can see in Fig. 4.1.

The 1 and 2-step transition probabilities we got from the original genome sequence by the following:

$$p_{ij} = \lim_{n \to \infty} \frac{\#\{l \le n : (X_{l-1}, X_l) = (i, j)\}}{\#\{l \le n - 1 : X_l = i\}} = \lim_{n \to \infty} \frac{\sum\limits_{l=1}^{n} \chi_{ij}(X_{l-1}, X_l)}{\sum\limits_{l=0}^{n-1} \chi_i(X_l)}$$

$$p_{ijk} = \lim_{n \to \infty} \frac{\#\{l \le n : (X_{l-2}, X_{l-1}, X_l) = (i, j, k)\}}{\#\{l \le n - 1 : (X_{l-1}, X_l) = (i, j)\}} = \lim_{n \to \infty} \frac{\sum\limits_{l=2}^{n} \chi_{ijk}(X_{l-2}, X_{l-1}, X_l)}{\sum\limits_{l=1}^{n-1} \chi_{ij}(X_{l-1}, X_l)},$$

where X_i denotes nucleotide sequence's elements, and $\chi_{i_1 \dots i_k}$ is a k-step indicator function:

$$\chi_{i_1 \dots i_k}(X_1, \dots, X_k) = \begin{cases} 1 & \text{if } (X_1, \dots, X_k) = (i_1, \dots, i_k), \\ 0 & \text{else.} \end{cases}$$

The average of the frequency of a subsequence $(i_1, \dots, i_k)$ is

$$S_n^{i^k}(X) = \frac{1}{n} \sum_{l=k-1}^{n} \chi_{i_1 \dots i_k}(X_{l-k+1}, \dots, X_l),$$

if $n \gg k$. In our case $k = 2$ and 3. Since the measure μ is ergodic, from Theorem 2.22, we have that $\lim\limits_{n \to \infty} S_n^{i^k}(X) = \int \chi_{i^k}(\bar{\jmath}) \, d\mu(\bar{\jmath}) = \mu[\bar{\imath}^k] = \prod_{l=0}^{k} p_{i_l}$. In the sequences, every 2 and 3 long permutation of the acid bases occurs, so if they are Markov chains, they have to be irreducible and aperiodic. Thus there must exist the spectral gap of the transition matrix, therefore there exists $c > 0$ such that $\mathbb{D}^2(S_n^{i^k}) \le \frac{c}{n}$. From this and the Chebisev inequality we get $\mu(|S_n - \int \chi \, d\mu| \ge \varepsilon) \le \frac{c}{\varepsilon n} \to 0$ as $n \to \infty$. Apriori we do not know the precise value of this constant without any further information. So we chose $n = 180\,000$. It is also confirmed by the fact that the 1 and 2-step Markov chains' pictures are both quite similar to the original one, as we can see in Fig. 4.2.
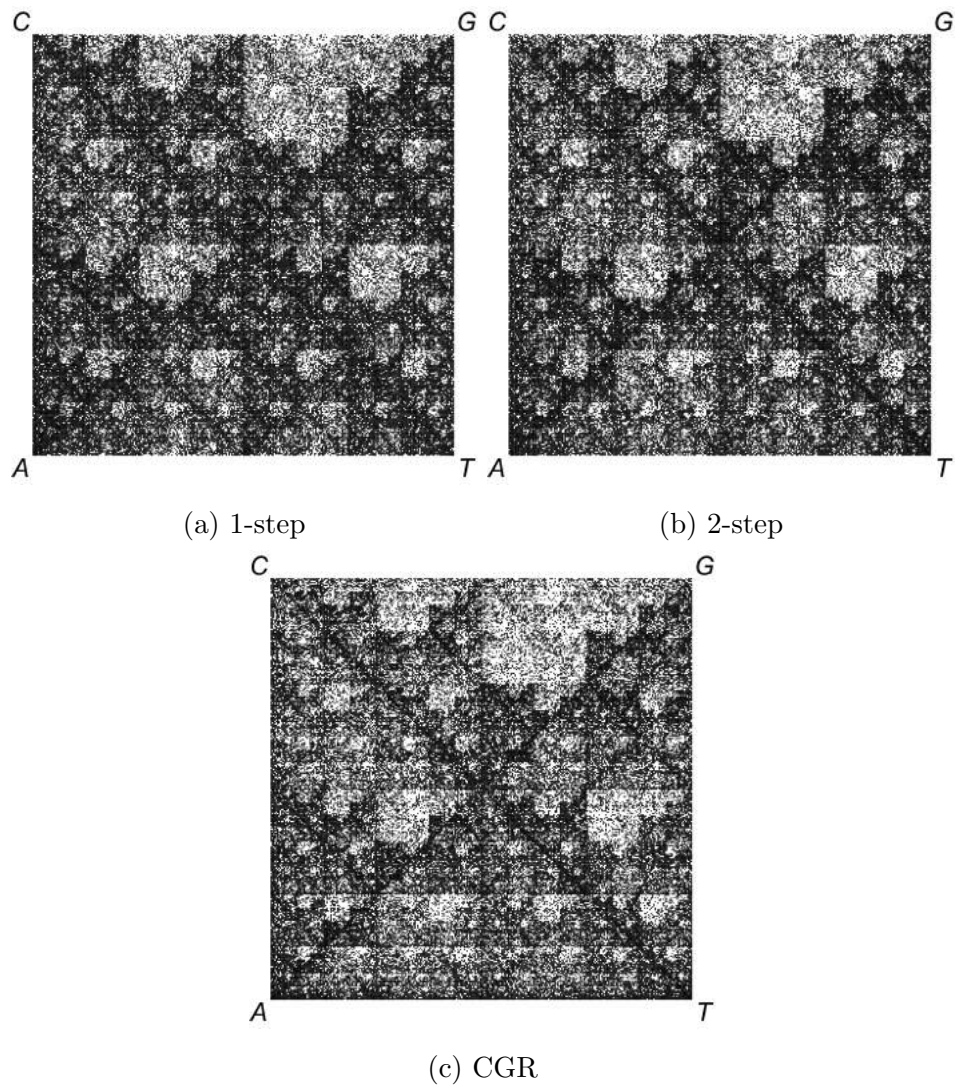
(a) 1-step

(b) 2-step

(c) CGR

Figure 4.2: The first 200 000 acid bases of Mus musculus represented with CGR and its approximations with Markov chains. Made with Wolfram Mathematica

	A	T	G	C
A	0,298	0,2240	0,2730	0,2050
T	0,187	0,2810	0,2850	0,2470
G	0,269	0,2220	0,2680	0,2410
C	0,313	0,313	0,096	0,278

Table 4.1: The transition matrix of the 1-step Markov chain of the Mus musculus' approximation

	A	T	G	C
AA	0,305	0,203	0,297	0,195
AT	0,3000	0,2640	0,2060	0,2300
AG	0,3080	0,1890	0,2950	0,2080
AC	0,258	0,204	0,326	0,212
TA	0,2154	0,2711	0,2851	0,2284
TT	0,1808	0,3156	0,2375	0,2661
TG	0,1694	0,2555	0,3350	0,2400
TC	0,1301	0,2804	0,3424	0,2471
GA	0,276	0,196	0,28	0,248
GT	0,2300	0,2370	0,2950	0,2380
GG	0,3006	0,2260	0,2386	0,2349
GC	0,237	0,185	0,329	0,249
CA	0,346	0,289	0,096	0,269
CT	0,2788	0,3386	0,0880	0,2947
CG	0,2950	0,3086	0,1057	0,2907
CC	0,356	0,298	0,113	0,233

Table 4.2: The transition matrix of the 2-step Markov chain of the Mus musculus' approximation

In Table 4.2, 4.3 and 4.4, we compacted the transition matrices. In the header row, we left the first bases, thus we did not indicate the events with zero probability.

If we consider the nucleotide sequence of the Beta vulgaris (carrot), we can see that it is not homogene, the CGR of the first and last 200 000 bases are different, see in Fig. 4.3. But we tried to approximate the different parts of the sequence. They are quite similar to the original pictures, thus we can

(a) First CGR   (b) Last CGR
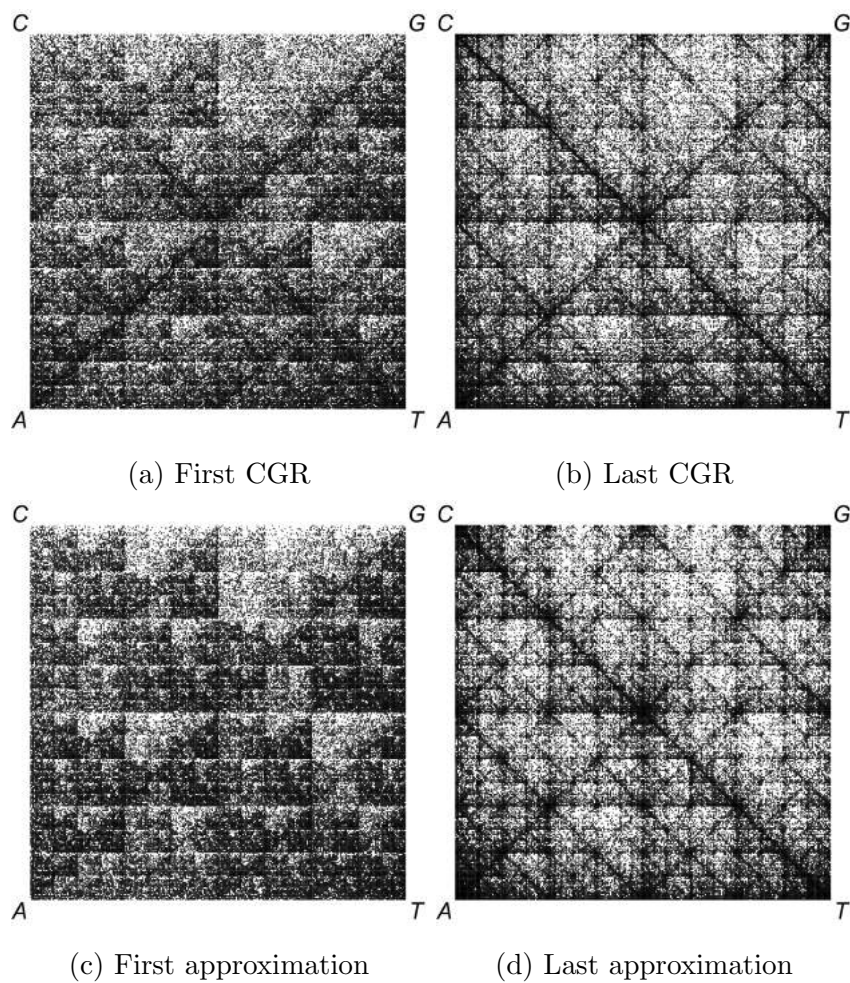
(c) First approximation   (d) Last approximation

Figure 4.3: The first and last 200 000 acid bases of Beta vulgaris represented with CGR and its approximations with Markov chains. Made with Wolfram Mathematica

draw the conclusion that while the whole sequence can not be approach with Markov chain, the different subsequences can be. See the transition matrices in Table 4.3 and 4.4.

	A	T	G	C
AA	0,264	0,262	0,293	0,181
AT	0,23	0,37	0,19	0,21
AG	0,31	0,32	0,24	0,14
AC	0,3	0,3	0,2	0,2
TA	0,17	0,28	0,32	0,23
TT	0,15	0,3	0,31	0,24
TG	0,19	0,34	0,26	0,21
TC	0,18	0,32	0,3	0,2
GA	0,32	0,2	0,25	0,23
GT	0,26	0,24	0,29	0,21
GG	0,34	0,30	0,20	0,17
GC	0,3	0,29	0,24	0,17
CA	0,35	0,32	0,09	0,24
CT	0,304	0,341	0,12	0,235
CG	0,31	0,42	0,07	0,20
CC	0,4	0,3	0,1	0,2

Table 4.3: The transition matrix of the 2-step Markov chain of the Beta vulgaris' approximation (from the first 200 000 bases)

0	A	T	G	C
AA	0,52	0,196	0,137	0,147
AT	0,37	0,341	0,122	0,167
AG	0,451	0,212	0,202	0,135
AC	0,41	0,24	0,14	0,21
TA	0,289	0,368	0,151	0,192
TT	0,17	0,48	0,15	0,2
TG	0,21	0,38	0,23	0,18
TC	0,17	0,40	0,15	0,28
GA	0,33	0,19	0,3	0,18
GT	0,27	0,252	0,293	0,185
GG	0,23	0,16	0,46	0,15
GC	0,25	0,19	0,37	0,19
CA	0,332	0,254	0,1	0,314
CT	0,237	0,357	0,087	0,319
CG	0,275	0,264	0,135	0,326
CC	0,20	0,22	0,10	0,48

Table 4.4: The transition matrix of the 2-step Markov chain of the Beta vulgaris' approximation (from the last 200 000 bases)

Chapter 5

Conclusions

We demonstrated that empirically certain parts of the DNA sequences can be approximated by 1 and 2-step Markov chains, but proving it is beyond the scope of this BSc thesis. We do not have proper estimation of the standard deviation, thus we can not define the value of n. The comparison of the original and the simulated sequences in Hausdorff metric may not be effective. It would be $200\,000 \times 200\,000$ computation two times to get their distance.

It propounds further questions. Can we use this method to improve gene mapping?

Jeffrey also propounded some open questions in [1]. One of them is the following: "Mathematically characterize the sparse sequences that produce the 'double scoop' pattern, which to date has been found only in non-oncogene vertebrate sequences and in some genes from viruses that can infect vertebrates?"
We can represent this 'double scoop' pattern superficially as a Markov chain. Its transition probability matrix is in Table 5.1, and its CGR is in Fig. 5.1. Thus we get the conclusion that G rarely comes after C. If we comparise this transition matrix with the 1-step Markov chain's transition matrix, we can see that the probability values approximately match, the average difference is 0.035.

	A	T	G	C
A	1/4	1/4	1/4	1/4
T	1/4	1/4	1/4	1/4
G	1/4	1/4	1/4	1/4
C	1/3	1/3	0	1/3

Table 5.1: The transition matrix of the Markov chain with $P_{CG} = 0$
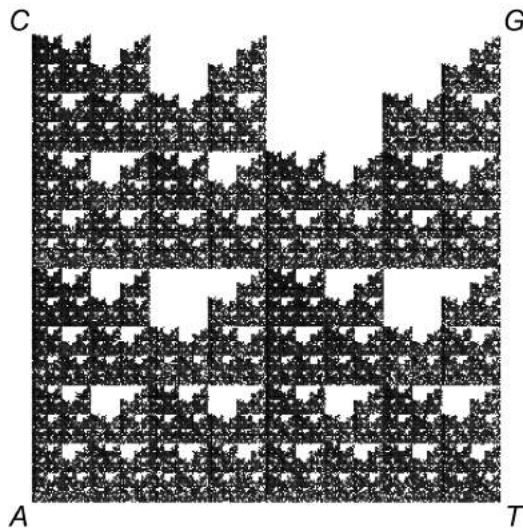

Figure 5.1: Representation of 1-step Markov chain with $P_{CG} = 0$. Made with Wolfram Mathematica

Bibliography

[1] H Joel Jeffrey. Chaos game representation of gene structure. *Nucleic Acids Research*, 18(8):2163–2170, 1990.

[2] David Koslicki and Daniel J Thompson. Coding sequence density estimation via topological pressure. *Journal of mathematical biology*, 70(1-2):45–69, 2015.

[3] Michael F Barnsley and Krzysztof Leśniak. The chaos game on a general iterated function system from a topological point of view. *International Journal of Bifurcation and Chaos*, 24(11):1450139, 2014.

[4] H. Kunze, D.L. Torre, F. Mendivil, and E.R. Vrscay. *Fractal-Based Methods in Analysis*. Mathematics and Statistics. Springer US, 2011.

[5] P. Walters. *An Introduction to Ergodic Theory*. Graduate Texts in Mathematics. Springer New York, 2000.

[6] Michael F Barnsley and Andrew Vince. The chaos game on a general iterated function system. *Ergodic Theory and Dynamical Systems*, 31(04):1073–1079, 2011.

[7] J. Quackenbush et al. TGI team. Tgi databases, 2014.