Substitution Operators

Andréa V. Rocha,[*] Alexandre B. Simas[†] and André Toom[‡]

[*]Departamento de Estatística, Universidade Federal da Paraíba,
Cidade Universitária, 58059-900, João Pessoa-PB, Brasil
[†]Departamento de Matemática, Universidade Federal da Paraíba,
Cidade Universitária, 58059-900, João Pessoa-PB, Brasil
[‡]Departamento de Estatística, Universidade Federal de Pernambuco,
Cidade Universitária, 50740-540, Recife-PE, Brasil

Abstract

We study a new kind of random processes with discrete time. In fact we study a certain class of operators called *substitution operators* whose iterations generate a process. Our operators act on probability measures on a configuration space $\mathcal{A}^{\mathbb{Z}}$, where $\mathcal{A}$ is a finite set called *alphabet*, elements of which are called *letters*. Elements of $\mathcal{A}^{\mathbb{Z}}$ are bi-infinite sequences of letters. Let us call any finite sequence of letters a *word*. *Length* of a word is the number of letters in it. Informally speaking, action of our operator consists in the following: any occurence of a word G in a configuration may be substituted by another word H with a certain probability ρ. Many well-known operators fit into this description with G and H having equal lenghts. Our main novelty is that the lengths of G and H may be different. This makes our operators non-linear and causes our main difficulties. Our main task is to define such operators rigorously (which is not trivial since our space is $\mathbb{Z}$ rather than its finite segment), prove some of their properties and use them to study invariant measures.

Keywords: variable-length random processes, operators, local interaction, ergodicity.

[*]E-mail: andrea@de.ufpb.br
[†]E-mail: alesimas@impa.br
[‡]Corresponding author. E-mail: toom@de.ufpe.br

1 Informal Introduction

The bulk of modern studies of locally interacting particle processes is based on the assumption that the set of sites, called the space, does not change in the process of interaction. Elements of this space, called components, may be in different states, e.g. 0 and 1, often interpreted as absence vs. presence of a particle, and may go from one state to another, which may be interpreted as birth or death of a particle, but the sites themselves do not appear or disappear in the process of functioning. Operators and processes which do not create or eliminate sites will be called *constant-length* ones.

However, in various areas of knowledge we deal with long sequences of components, which are subject to some local random transformations, which may change their lengths. The simplest and the most well-known of such transformations are often called "insertion" and "deletion" and are widely discussed in informatics and molecular biology (see e. g. [14, 15], where one can find more references). In such cases we use the phrase *variable-lenght processes* and our goal is to provide a rigorous definition of some class of variable-length processes with infinite space and study their properties.

Besides works of our group [11, 12, 13, 8, 9, 10] we found only several works on similar processes (see e.g. [1, 2, 4, 5, 6, 7]), which did much to emphasize connections of such processes with modern physics; however these works seem to contain no attempts to define variable-length processes with infinite space.

According to our knowledge, the first rigorous definitions of some non-trivial variable-length operators with infinite space were presented in [12, 13, 8] and the only rigorous definition of a large class of variable-length processes has been presented in [10]. This article is based on [10], where one can find some details omitted here. Our processes have discrete time and therefore can be defined in terms of operators acting on probability measures, which we call *substitution operators* or *SO* for short.

By $\#(S)$ we denote the cardinality of any finite set S. Throughout this article $\mathcal{A}$ is a non-empty finite set called *alphabet*. Its elements are called *letters* and finite sequences of letters are called *words*. The number of letters in a word W is called its *length* and denoted by $|W|$. Any letter may be considered as a word of length one. There is the empty word, denoted by Λ, whose length is zero. The set of words in a given alphabet $\mathcal{A}$ is called *dictionary* and denoted by $\mathrm{dic}(\mathcal{A})$. We denote by $\mathbb{Z}$ the set of integer numbers and $\mathcal{A}^{\mathbb{Z}}$ the set of bi-infinite (that is, infinite in both directions) sequences whose terms are elements of $\mathcal{A}$. We denote by $\mathcal{M}$ the set of translation-invariant probability measures on $\mathcal{A}^{\mathbb{Z}}$ that is on the σ-algebra generated by cylinders.

A generic SO acts from $\mathcal{M}$ to $\mathcal{M}$ roughly as follows. Given two words G and H (where G must satisfy a certain condition of self-avoiding which will be presented below) and a real number $\rho \in [0, 1]$, an SO, informally speaking, substitutes every occurrence of the word G in any configuration by the word H with a probability ρ (and leaves this occurrence unchanged with a probability $1 - \rho$). Our definition can be used only if all the occurrences of G in any configuration do not overlap, and this is why we need a special assumption about G.

Before going into formal details let us present a short synopsis of our work. Our first task is to define SO. We do it in several stages. In section 2 we define how measures may be approximated by words. Then we introduce random words, which also can approximate measures. In sections 3 and 4 we define how SO act on words and random words. This allows us to introduce *extension*, that is the coefficient, by which is multiplied the length of a word when a SO is applied to it. We do it in section 5. Extension, in its turn, allows us to define how SO act on measures. However, we found it too complicated to define directly how an arbitrary SO acts on measures. For this reason in section 6 we present a short list of *basic* SO (including insertion and deletion mentioned above) and define how they act on measures. Then in section 7 we represent an arbitrary SO as a composition of several basic SO and use this representation to define how an arbitrary SO acts on measures. Thus SO are completely defined. Based on this theoretical preparation, we study some properties of SO. A major difficulty in dealing with SO is that they are in general non-linear unlike the bulk of random processes studied till now. However, we found another property, which sometimes is as good as linearity: in section 8 we introduce *segment-preserving* operators and prove that all our SO have this property. In addition to this, in section 9 we prove that all our SO are continuous, which allows us to prove that each of them has at least one invariant measure. Using [12, 8], we prove that a certain operator has at least two invariant measures, which contributes to the study of one-dimensional non-ergodicity.

2 Formal Introduction

Let us denote by $\mathbb{A}$ the discrete topology on $\mathcal{A}$. We consider probability measures on the σ-algebra $\mathbb{A}^{\mathbb{Z}}$ on the product space $\mathcal{A}^{\mathbb{Z}}$ endowed with the topology - product of discrete topologies on all the copies of $\mathcal{A}$. Since $\mathcal{A}$ is finite, it is compact in the discrete topology, and by Tychonoff's compact theorem, $\mathcal{A}^{\mathbb{Z}}$ also is compact.

As usual, shifts on $\mathbb{Z}$ generate shifts on $\mathcal{A}^{\mathbb{Z}}$ and shifts on $\mathbb{A}^{\mathbb{Z}}$. We call a measure μ on $\mathcal{A}^{\mathbb{Z}}$ *uniform* if it is invariant under all shifts. In this case for any word $W = (a_1, \dots, a_n)$, where $a_1, \dots, a_n$ is the sequence of letters that forms the word W, the right side of

$$\mu(W) = \mu(a_1, \dots, a_n) = \mu(s_{i+1} = a_1, \dots, s_{i+n} = a_n)$$

is one and the same for all $i \in \mathbb{Z}$, whence we may use the left side as a shorter denotation.

We denote by $\mathcal{M}$ the set of uniform probability measures on $\mathcal{A}^{\mathbb{Z}}$. Since $\mathcal{A}^{\mathbb{Z}}$ is compact, $\mathcal{M}$ is also compact. Any uniform measure is determined by its values on all the words and it is a probability, that is normalized measure if its value on the empty word equals 1. So we may define a measure in $\mathcal{M}$ by its values on words. In order for values $\mu(W)$ to form a uniform probability measure, it is necessary and sufficient that: all the numbers $\mu(W)$ must be non-negative, μ on the empty word must equal one and for any letter a and any word W we must have

$$\mu(W) = \sum_{a \in \mathcal{A}} \mu(W, a) = \sum_{a \in \mathcal{A}} \mu(a, W),$$

where (W, a) and (a, W) are concatenations of the word W and the letter a in the two possible orders.

We assume that our alphabet contains no brackets or commas. Given any finite sequence of words $(W_1, \dots, W_k)$ (perhaps separated by commas or put in brackets), we denote by $\text{concat}(W_1, \dots, W_k)$ and call their *concatenation* the word obtained by writing all these words one after another in that order in which they are listed, all brackets and commas eliminated. In particular, W^n means concatenation of n words, everyone of which is a copy of W. If $n = 0$, the word W^n is empty, $W^0 = \Lambda$.

Given two words $W = (a_1, \dots, a_m)$ and $V = (b_1, \dots, b_n)$, where $|W| \leq |V|$, we call the integer numbers in the interval $[0,\ n-m]$ *positions* of W in V. We say that W *enters* V at a position k if

$$\forall\, i \in \mathbb{Z}\ :\ 1 \leq i \leq m \Rightarrow a_i = b_{i+k}.$$

We call a word W *self-overlapping* if there is a word V such that $|V| < 2 \cdot |W|$ and W enters V at two different positions. A word is called *self-avoiding* if it is not self-overlapping. In particular, the empty word, every word consisting of one letter and every word consisting of two different letters are self-avoiding.

It is known that self-avoiding words are not very rare: in fact for any alphabet with at least two letters the number of self-avoiding words of length n divided by the number of all words of length n tends to a positive limit when $n \to \infty$ and this limit tends to one when the number of letters in the alphabet tends to infinity [3].

We denote by $\text{freq}\,(W \ \text{ in } \ V)$ the frequency of W in V, that is, the number of positions at which W enters V. If W is the empty word, it enters any word V at $|V|+1$ positions. If $|W| \leq |V|$, we call the *relative frequency* of a word W in a word V and denote by $\text{rel.freq}\,(W \ \text{ in } \ V)$ the number of positions at which W enters V divided by the total number of positions of W in V, that is, the fraction

$$\text{rel.freq}\,(W \ \text{ in } \ V) = \frac{\text{freq}\,(W \ \text{ in } \ V)}{|V| - |W| + 1}. \tag{1}$$

Notice that the relative frequency of the empty word in any word is 1. If $|W| > |V|$, the set of positions of W in V is empty and the relative frequency of W in V is zero by definition.

We call a *pseudo-measure* any map $\mu : \text{dic}(\mathcal{A}) \to \mathbb{R}$. In particular, any measure $\mu \in \mathcal{M}$ is a pseudo-measure if it is defined by its values on words.

Definition 2.1. For every word $V \in \text{dic}(\mathcal{A})$ we define the corresponding pseudo-measure, denoted by meas^V and defined by the rule $\text{meas}^V(W) = \text{rel.freq}\,(W \ \text{ in } \ V)$ for every word W.

Definition 2.2. We say that a sequence (V_n) of words $V_1,\ V_2,\ V_3, \dots \in \text{dic}(\mathcal{A})$ converges to a measure $\mu \in \mathcal{M}$ if for every word $W \in \text{dic}(\mathcal{A})$ the relative frequency of W in V_n tends to $\mu(W)$ as $n \to \infty$, that is, if $\text{meas}^{V_n}(W)$ tends to $\mu(W)$ as $n \to \infty$.

Remark 2.3. Notice that since the relative frequencies of all W in a given V are zeros for all W longer than V, the convergence in the definition 2.2 is possible only if the length of V_n tends to ∞ as $n \to \infty$.

Definition 2.4. Given a real number $\varepsilon > 0$ and a natural number r, a word V is said to $(\varepsilon,\ r)$-*approximate* a measure $\mu \in \mathcal{M}$ if for every word $W \in \mathrm{dic}(\mathcal{A})$,

$$|W| \leq r \Rightarrow |\text{rel.freq}\,(W \text{ in } V) - \mu(W)| \leq \varepsilon.$$

Lemma 2.5. *A sequence (V_n) of words converges to a measure μ if and only if for any positive $\varepsilon > 0$ and any natural r there is n_0 such that for every $n \geq n_0$ the word V_n $(\varepsilon,\ r)$-approximates μ.*

Proof in one direction: Suppose that (V_n) converges to μ. We want to prove that

$$\left.\begin{array}{c} \forall\,\varepsilon > 0 \ \ \forall\, r \in N \quad \exists\, n_0 \ \ \forall\, n \geq n_0, \ \forall\, W \in \mathrm{dic}(\mathcal{A}): \\ |W| \leq r \ \Rightarrow \ |\text{rel.freq}\,(W \text{ in } V_n) - \mu(W)| \leq \varepsilon. \end{array}\right\} \tag{2}$$

Let us choose W such that $0 < |W| \leq r$. Since (V_n) converges to μ,

$$\lim_{n\to\infty} \text{rel.freq}\,(W \text{ in } V_n) = \mu(W),$$

that is

$$\forall\,\varepsilon' > 0 \ \ \exists\, n_W \ \ \forall\, n \geq n_W : |\text{rel.freq}\,(W \text{ in } V_n) - \mu(W)| \leq \varepsilon'.$$

Taking $\varepsilon' = \varepsilon$ and n_0 equal to the maximum of n_W over all those non-empty W, whose length does not exceed r, we obtain (2).

In the other direction the proof is straightforward. *Lemma 2.5 is proved.*

Theorem 2.6. *For any $\mu \in \mathcal{M}$, any $\varepsilon > 0$ and any natural r there is a word which $(\varepsilon,\ r)$-approximates μ.*

Proof: If $\#(\mathcal{A}) = 1$, the theorem is trivial. So let $\#(\mathcal{A}) > 1$. Let us introduce parameters

$$s = \left]\frac{4r}{\varepsilon}\right[, \quad N = (\#(\mathcal{A}))^s \quad \text{and} \quad Q = \left]\frac{4N}{\varepsilon}\right[.$$

Hence

$$\frac{r}{s} \le \frac{\varepsilon}{4} \quad \text{and} \quad \frac{N}{Q} \le \frac{\varepsilon}{4}.$$

Notice that there are N words in $\mathrm{dic}(\mathcal{A})$, whose length equals s, and we denote them by $U_1,\ U_2, \dots, U_N$.

Furthermore, for any position k of W in U_i, that is, for any $k \in [1, s-|W|+1]$, we define

$$\mathrm{freq}\,(W \ \text{ in } \ U_i)_k = \begin{cases} 1 & \text{if } W \text{ enters } U_i \text{ at the position } k, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\sum_{k=1}^{s-|W|+1} \mathrm{freq}\,(W \ \text{ in } \ U_i)_k = \mathrm{freq}\,(W \ \text{ in } \ U_i) \tag{3}$$

and

$$\sum_{i=1}^{N} (\mathrm{freq}\,(W \ \text{ in } \ U_i)_k \cdot \mu(U_i)) = \mu(W).$$

Summing this over k yields

$$\sum_{k=1}^{s-|W|+1} \sum_{i=1}^{N} (\mathrm{freq}\,(W \ \text{ in } \ U_i)_k \cdot \mu(U_i)) = (s-|W|+1) \cdot \mu(W).$$

Hence

$$\sum_{i=1}^{N} \left(\mu(U_i) \cdot \sum_{k=1}^{s-|W|+1} \mathrm{freq}\,(W \ \text{ in } \ U_i)_k \right) = (s-|W|+1) \cdot \mu(W). \tag{4}$$

Replacing (3) by (4) gives

$$\sum_{i=1}^{N} (\mathrm{freq}\,(W \ \text{ in } \ U_i) \cdot \mu(U_i)) = (s-|W|+1) \cdot \mu(W). \tag{5}$$

Further, for every $i = 1, \dots, N$ we denote

$$p_i = [Q \cdot \mu(U_i)]\,,$$

where Q was defined in the beginning of the proof. Hence

$$Q \cdot \mu(U_i) - 1 < p_i \le Q \cdot \mu(U_i). \tag{6}$$

For every i from 1 to N we take p_i copies of U_i and define V as their concatenation in any order, for instance

$$V = \mathrm{concat}(U_1^{p_1}, \dots, U_N^{p_N}).$$

Let us check that the word V has the desired property, namely $(\varepsilon,\ r)$-approximates the measure μ. Since $\mu(U_1) + \cdots + \mu(U_N) = 1$, summing (6) over $i = 1,\ \ldots,\ N$ gives

$$Q - N < \sum_{i=1}^{N} p_i \leq Q. \tag{7}$$

Let us estimate the relative frequency of W in V. First from below: In the present case the numerator of that fraction is

$$\text{freq}\,(W \ \text{ in } \ V) \geq \sum_{i=1}^{N} \text{freq}\,(W \ \text{ in } \ U_i) \cdot p_i$$

and the denominator is

$$|V| - |W| + 1 \leq |V| = s \cdot \sum_{i=1}^{N} p_i \leq s \cdot Q.$$

Hence, using equation (5), we obtain

$$\begin{aligned}
&\text{rel.freq}\,(W \ \text{ in } \ V) \geq \frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} (\text{freq}\,(W \ \text{ in } \ U_i) \cdot p_i) \geq \\
&\frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} (\text{freq}\,(W \ \text{ in } \ U_i) \cdot (Q \cdot \mu(U_i) - 1)) = \\
&\frac{1}{s} \cdot \sum_{i=1}^{N} (\text{freq}\,(W \ \text{ in } \ U_i) \cdot \mu(U_i)) - \frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} \text{freq}\,(W \ \text{ in } \ U_i) \geq \\
&\frac{s - |W| + 1}{s} \cdot \mu(W) - \frac{s \cdot N}{s \cdot Q} \geq \left(1 - \frac{r}{s}\right) \cdot \mu(W) - \frac{N}{Q} \geq \\
&\left(1 - \frac{\varepsilon}{4}\right) \cdot \mu(W) - \frac{\varepsilon}{4} \geq \mu(W) - \varepsilon.
\end{aligned} \tag{8}$$

Now let us estimate the relative frequency of W in V from above. The numerator of the fraction (1) is not greater than

$$\sum_{i=1}^{N} (\text{freq}\,(W \ \text{ in } \ U_i) + |W|) \cdot p_i \leq \sum_{i=1}^{N} (\text{freq}\,(W \ \text{ in } \ U_i) + |W|) \cdot Q \cdot \mu(U_i)$$

and the denominator is not less than

$$|V| - |W| + 1 \geq s \cdot \sum_{i=1}^{N} p_i - r \geq s \cdot (Q - N) - r.$$

Since $\#(\mathcal{A}) \geq 2$, $\quad r \geq 1$ and we may assume that $\varepsilon \leq 1$,

$$\frac{r}{Q \cdot N} \leq \frac{r \cdot \varepsilon}{4 \cdot N^2} \leq \frac{s \cdot \varepsilon^2}{16 \cdot N^2} \leq \frac{s \cdot \varepsilon^2}{16 \cdot 4^s} = \left(\frac{s}{4^s}\right) \cdot \frac{\varepsilon^2}{16} \leq \frac{\varepsilon^2}{16}.$$

Therefore

$$\frac{s \cdot Q}{s \cdot (Q - N) - r} = \frac{1}{1 - \frac{N}{Q} - \frac{r}{Q \cdot N}} \leq \frac{1}{1 - \frac{\varepsilon}{4} - \frac{\varepsilon^2}{16}} \leq 1 + \frac{\varepsilon}{2}.$$

Thus, using (6), (7) and (8), we get

$$\begin{aligned}
&\text{rel.freq}\,(W \ \text{ in } \ V) \leq \\
&\left(1 + \frac{\varepsilon}{2}\right) \cdot \frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} ((\text{freq}\,(W \ \text{ in } \ U_i) + |W|) \cdot Q \cdot \mu(U_i)) \leq \\
&\left(1 + \frac{\varepsilon}{2}\right) \cdot \frac{1}{s} \cdot \left(\sum_{i=1}^{N} (\text{freq}\,(W \ \text{ in } \ U_i) \cdot \mu(U_i)) + r \cdot \sum_{i=1}^{N} \mu(U_i) \right) \leq \\
&\left(1 + \frac{\varepsilon}{2}\right) \cdot \frac{1}{s} \cdot (s \cdot \mu(W) + r) \leq \left(1 + \frac{\varepsilon}{2}\right) \cdot \left(\mu(W) + \frac{r}{s}\right) \leq \\
&\left(1 + \frac{\varepsilon}{2}\right) \cdot \left(\mu(W) + \frac{\varepsilon}{4}\right) \leq \mu(W) + \varepsilon.
\end{aligned}$$

Theorem 2.6 is proved.

Corollary 2.7. *For any $\mu \in \mathcal{M}$ there is a sequence of words which converges to it.*

Proof: Due to the previous theorem 2.6, for every natural n we can find a word V_n which $(1/n,\ n)$-approximates μ. Note that if a word $(\varepsilon,\ r)$-approximates a measure, then it $(\varepsilon',\ r')$-approximates the same measure for any $\varepsilon' \geq \varepsilon$ and $r' \leq r$. Therefore the sequence (V_n) converges to μ. *Corollary 2.7 is proved.*

Remark 2.8. One of our referees noticed that Corollary 2.7 should have been published somewhere, even in a more general form, but we found no reference and present our own proof.

We define a *random word* X in an alphabet $\mathcal{A}$ as a random variable on $\text{dic}(\mathcal{A})$ which is concentrated on a finite subset of $\text{dic}(\mathcal{A})$. A random word is determined by its components $P(X = V)$, that is probabilities that $X = V$, whose sum is 1, the set $\{V : P(X = V) > 0\}$ being finite. We denote by Ω the set of random words in the alphabet $\mathcal{A}$.

Definition 2.9. We define the mean length of any random word X as

$$E|X| = \sum_{V \in \text{dic}(\mathcal{A})} P(X = V) \cdot |V|.$$

Definition 2.10. We define the mean frequency of a word W in a random word X as

$$E[\text{freq}\,(W \ \text{ in } \ X)] = \sum_{V \in \text{dic}(\mathcal{A})} P(X = V) \cdot \text{freq}\,(W \ \text{ in } \ V).$$

Definition 2.11. We define the mean relative frequency of a word W in a random word X as

$$\text{rel.freq}_{\,E}(W \ \text{ in } \ X) = \frac{E[\text{freq}\,(W \ \text{ in } \ X)]}{E|X| - |W| + 1}. \tag{9}$$

For any random word X we define the corresponding pseudo-measure meas^X by the rule

$$\text{meas}^X(W) = \text{rel.freq}_{\,E}(W \ \text{ in } \ X) \quad \text{for every word} \quad W.$$

Definition 2.12. We say that a sequence (X_n) of random words $X_1, \ X_2, \ X_3, \dots \in \Omega$ converges to a measure $\mu \in \mathcal{M}$ if for every word $W \in \text{dic}(\mathcal{A})$ the mean relative frequency of W in X_n tends to $\mu(W)$ as $n \to \infty$, that is if $\text{meas}^{X_n}(W)$ tends to $\mu(W)$ as $n \to \infty$.

Definition 2.13. Given a positive number $\varepsilon > 0$ and a natural number r, we say that a random word X $(\varepsilon, \ r)$*-approximates* a measure $\mu \in \mathcal{M}$ if for every non-empty word $W \in \text{dic}(\mathcal{A})$,

$$|W| \leq r \Rightarrow |\text{rel.freq}_{\,E}(W \ \text{ in } \ X) - \mu(W)| \leq \varepsilon.$$

Theorem 2.14. *For any $\mu \in \mathcal{M}$, any $\varepsilon > 0$ and any natural r there is a random word which $(\varepsilon, \ r)$-approximates μ.*

Proof: It could be obtained as a corollary of theorem 2.6 by considering the random word as a distribution concentrated on a single word. One could also adapt the proof of theorem 2.6 by considering a random word X with $P(X = U_i) = p_i$, with p_i and U_i being the same as in the proof of theorem 2.6. *Theorem 2.14 is proved.*

Corollary 2.15. *For any $\mu \in \mathcal{M}$ there is a sequence of random words which converges to it.*

Proof: Analogous to the proof of corollary 2.7. *Corollary 2.15 is proved.*

3 SO Acting on Words

A generic SO is determined by two words G and H, where G is self-avoiding, and a real number $\rho \in [0,1]$. We denote this operator by $(G \xrightarrow{\rho} H)$. The informal idea of this operator is that it substitutes every entrance of the word G in a long word by the word H with a probability ρ or leaves it unchanged with a probability $1-\rho$ independently of states and fate of all the other components. Following some articles in this area, we write operators on the right side of objects (words, measures) on which they act.

Our goal in this section is to define a general SO, which is denoted by $(G \xrightarrow{\rho} H)$, acting on words. If G and H are empty, the operator $(G \xrightarrow{\rho} H)$ changes nothing. Now let G or H be non-empty. Let us define how $(G \xrightarrow{\rho} H)$ acts on an arbirtary word V. Let us denote $N = \text{freq}\,(G \text{ in } V)$, i.e. N is the number of entrances of G in V. Since G is self-avoiding, these entrances do not overlap. Further, let $i_1, \ldots, i_N \in \{0,1\}$ and let us denote by $V(i_1, \ldots, i_N)$ the word obtained from V after replacing each entrance of G by the word H in those positions i_j where $i_j = 1$, the others left unchanged. We may, therefore, define the SO $(G \xrightarrow{\rho} H)$ as follows: the random word obtained from the word V is concentrated on the words $V(i_1, \ldots, i_N)$, where $i_1, \ldots, i_N \in \{0,1\}$ with probabilities

$$\mathbb{P}\Big(V(G \xrightarrow{\rho} H) = V(i_1, \ldots, i_N)\Big) \;=\; \rho^k \cdot (1-\rho)^{N-k}, \quad \text{where} \quad k = \sum_j i_j.$$

Now let us extend this definition to random words. Let us define the result of application of $(G \xrightarrow{\rho} H)$ to a random word X. Let X equal the words $V_1, \ldots, V_n$ with positive probabilities $P(X = V_j)$. We define $X(G \xrightarrow{\rho} H)$ as the random word which equals the words $V_j(i_1, \ldots, i_N)$ with probabilities

$$\mathbb{P}(X = V_j) \cdot \rho^{\sum_j i_j} \cdot (1-\rho)^{N - \sum_j i_j}.$$

Lemma 3.1. *For any non-empty word V and $\rho \in [0,1]$ we can express the mean length of the random word $V(G \xrightarrow{\rho} H)$ in the following simple way*

$$E|V(G \xrightarrow{\rho} H)| = |V| + \rho \cdot (|H| - |G|) \cdot \text{freq}\,(G \text{ in } V).$$

Proof: Let us evaluate the mean length (definition 2.9) of the random word $V(G \xrightarrow{\rho} H)$. We begin by noting that

$$|V(G \xrightarrow{\rho} H)| = |V| + (|H| - |G|) \cdot k,$$

where $k \sim \text{Bin}(N,\ \rho)$, where $N = \text{freq}\,(G \ \text{ in } \ V)$. Therefore

$$E|V(G \xrightarrow{\rho} H)| = |V| + (|H| - |G|) \cdot E(k),$$

which is equal to

$$E|V(G \xrightarrow{\rho} H)| = |V| + \rho \cdot (|H| - |G|) \cdot \text{freq}\,(G \ \text{ in } \ V).$$

Lemma 3.1 is proved.

Lemma 3.2. *For any random word X and a number $\rho \in [0,1]$ we can express the mean length of the random word $X(G \xrightarrow{\rho} H)$ in the following simple way*

$$E|X(G \xrightarrow{\rho} H)| = E|X| + \rho \cdot (|H| - |G|) \cdot E[\text{freq}\,(G \ \text{ in } \ X)].$$

Proof. Suppose that the possible values of X are the words $V_1, \ldots, V_n$. Then we note that

$$E|X(G \xrightarrow{\rho} H)| = \sum_{j=1}^{n} E|V_j(G \xrightarrow{\rho} H)|P(X = V_j).$$

Now, we use lemma 3.1 to obtain

$$\begin{aligned} E|X(G \xrightarrow{\rho} H)| &= \sum_{j=1}^{n} [|V_j| + \rho \cdot (|H| - |G|)\text{freq}\ (G \ \text{ in } \ V_j)]P(X = V_j) \\ &= E|X| + \rho \cdot (|H| - |G|)\ E[\text{freq}\,(G \ \text{ in } \ X)]. \end{aligned}$$

Lemma 3.2 is proved.

4 SO Acting on Random Words

Remember our notations: $\mathcal{A}$ is an alphabet, Ω is the set of random words on $\mathcal{A}$, $\mathcal{M}$ is the set of uniform probability measures on $\text{dic}(\mathcal{A})$.

Proposition 4.1. *Let X_n be a sequence of random words. If X_n converges to a pseudo-measure μ, then, μ is, in fact, a measure.*

Proof: Let us choose a word W and suppose that the sequence (X_n) converges, thus the limit $\lim_{n\to\infty} \text{rel.freq}_E(W \ \text{ in } \ X_n)$ exists. So let us define the following map having the set of all words as its domain:

$$\mu(W) = \lim_{n\to\infty} \text{rel.freq}_E(W \ \text{ in } \ X_n).$$

We want to prove that μ is indeed a measure. In other words, we want to prove for any word W that $0 \le \mu(W) \le 1$ and also that $\sum_a \mu(W,a) = \sum_a \mu(a,W) = \mu(W)$.

It is easy to see that $0 \le \text{rel.freq}_E(W \ \text{ in } \ X_n) \le 1$. Then

$$0 \le \lim_{n\to\infty} \text{rel.freq}_E(W \ \text{ in } \ X_n) \le 1.$$

Therefore

$$0 \le \mu(W) \le 1.$$

We still have to show that $\sum_a \mu(W,a) = \sum_a \mu(a,W) = \mu(W)$. To do so, note initially that $|(W,a)| = |(a,W)| = |W| + 1$.

Let us take first the case when a is on the right side, that is we show that $\sum_a \mu(W,a) = \mu(W)$. Let V be any word in $\text{dic}(\mathcal{A})$. If $a \neq b$, then (W,a) must enter V in a different position than that of (W,b). Moreover, if W enters V in a position which is not the last one, that is, if W does not enter V at the position $|V| - |W|$, there must exist a letter, say c, at the right side of W, such that (W,c) still enters V at the same position. Now we can make two remarks. First, the number of entrances of W in V is always greater or equal than the sum over all the letters a of the numbers of entrances of (W,a) in V, for if (W,a) enters V, then W also enters V. Second, if W enters V at a non-last position, then (W,c) enters V for some c, as explained before. Therefore

$$0 \le \text{freq}\,(W \ \text{ in } \ V) - \sum_a \text{freq}\,((W,a) \ \text{ in } \ V) \le 1$$

for any word V. Then, multiplying the above expression by $P(X_n = V)$, we get

$$0 \le P(X_n = V) \cdot \text{freq}\,(W \ \text{ in } \ V) - P(X_n = V) \cdot \sum_a \text{freq}\,((W,a) \ \text{ in } \ V) \le P(X_n = V).$$

Thus, summing over all words V, and noting that the set $\{V : X_V > 0\}$ is finite, yields

$$0 \le E[\text{freq}\,(W \ \text{ in } \ X_n)] - \sum_a E[\text{freq}\,((W,a) \ \text{ in } \ X_n)] \le 1.$$

Now, dividing by $(E|X_n| - |W| + 1)$ gives

$$0 \leq \text{rel.freq}_E(W \ \text{ in } \ X_n) - \sum_a \frac{E[\text{freq}\,((W,a) \ \text{ in } \ X_n)]}{E|X_n| - |W| + 1} \leq \frac{1}{E|X_n| - |W| + 1},$$

which yields

$$0 \leq \text{rel.freq}_E(W \ \text{ in } \ X_n) - \sum_a \text{rel.freq}_E((W,a) \ \text{ in } \ X_n) \frac{E|X_n| - |W|}{E|X_n| - |W| + 1} \leq \frac{1}{E|X_n| - |W| + 1},$$

since $1/(E|X_n| - |W| + 1) \to 0$ as $n \to \infty$, because $E|X_n| \to \infty$ as $n \to \infty$, and by the same reason,

$$\frac{E|X_n| - |W|}{E|X_n| - |W| + 1}$$

tends to 1 as $n \to \infty$. Therefore,

$$0 \leq \lim_{n\to\infty} \text{rel.freq}_E(W \ \text{ in } \ X_n) - \sum_a \lim_{n\to\infty} \text{rel.freq}_E((W,a) \ \text{ in } \ X_n) \leq 0$$

that is

$$\mu(W) = \sum_a \mu(W,a).$$

The argument for a on the left side is analogous. Thus, the map $\mu(\cdot)$ is indeed a measure.

Proposition 4.1 is proved.

Definition 4.2. We say that a map $P : \Omega \to \Omega$ is *consistent* if the following condition holds: for any $\mu \in \mathcal{M}$ and any sequence of random words $(X_n) \to \mu$ the limit $\lim_{n\to\infty}(X_n P)$ exists and is one and the same for all sequences $X_n \to \mu$.

Definition 4.3. Given any consistent map $P : \Omega \to \Omega$ and any $\mu \in \mathcal{M}$, we define μP , that is the result of application of P to μ, as the measure (see Proposition 4.1, note also that it is unique according to definition 4.2), to which $(X_n P)$ converges for all $(X_n) \to \mu$, and we call it *limit of consistent operators.*

Lemma 4.4. *Let P_1, $P_2 : \Omega \to \Omega$ be consistent operators. Then their composition is also consistent.*

Proof: Consider any sequence of random words (X_n) converging to μ. Then, the sequence of random words $Q_n = X_n P_1$ tends to μP_1 (following definition 4.3), hence $Q_n P_2$ tends to $\mu P_1 P_2$. *Lemma 4.4 is proved.*

5 Extension

Definition 5.1. For any $\mu \in \mathcal{M}$ and any $P : \Omega \to \Omega$ we define *extension* of μ under P as the limit

$$\operatorname{Ext}(\mu|P) = \lim_{n\to\infty} \frac{E|X_n P|}{E|X_n|}$$

for any sequence of random words $(X_n) \to \mu$ if this limit exists and is one and the same for all sequences (X_n) which tend to μ.

Informally speaking, extension of a measure μ under operator P is that coefficient by which P multiplies the length of a word approximating μ.

Lemma 5.2. *Suppose that $P_1,\ P_2 : \Omega \to \Omega$ have extensions for all measures and P_1 is consistent. Then their composition P_1P_2 also has extension for all measures and*

$$\forall\ \mu : \operatorname{Ext}(\mu|P_1P_2) = \operatorname{Ext}(\mu|P_1) \times \operatorname{Ext}(\mu P_1|P_2).$$

Proof. Since we are assuming that P_1 is consistent, we have by definition 4.3 that $X_nP_1 \to \mu P_1$ as $n \to \infty$ for any sequence (X_n) of random words converging to μ. Thus, since we are assuming that P_2 has extension for all measures, definition 5.1 implies that

$$\operatorname{Ext}(\mu P_1|P_2) = \lim_{n\to\infty} \frac{E|V_n P_2|}{E|V_n|},$$

for any sequence of random words (V_n) converging to μP_1. Thus, since $X_nP_1 \to \mu P_1$, as seen in the beginning of the proof,

$$\operatorname{Ext}(\mu P_1|P_2) = \lim_{n\to\infty} \frac{E|X_n P_1 P_2|}{E|X_n P_1|}.$$

Therefore

$$\lim_{n\to\infty} \frac{E|X_nP_1P_2|}{E|X_n|} = \lim_{n\to\infty} \left(\frac{E|X_nP_1P_2|}{E|X_nP_1|} \cdot \frac{E|X_nP_1|}{E|X_n|} \right) =$$

$$\lim_{n\to\infty} \frac{E|X_nP_1P_2|}{E|X_nP_1|} \cdot \lim_{n\to\infty} \frac{E|X_nP_1|}{E|X_n|} = \operatorname{Ext}(\mu P_1|P_2) \cdot \operatorname{Ext}(\mu|P_1).$$

The above expression implies that the extension of μ resulting from application of a composition P_1P_2 exists and equals

$$\operatorname{Ext}(\mu|P_1P_2) = \operatorname{Ext}(\mu P_1|P_2) \times \operatorname{Ext}(\mu|P_1).$$

Lemma 5.2 is proved.

Now let us show that every measure in $\mathcal{M}$ has an extension under every SO and provide an explicit expression for it.

Proposition 5.3. *If* $(G \xrightarrow{\rho} H)$ *is a SO acting on random words, then the extension of any* $\mu \in \mathcal{M}$ *under this operator exists and equals*

$$\mathrm{Ext}(\mu|(G \xrightarrow{\rho} H)) = 1 + \rho \cdot (|H| - |G|) \cdot \mu(G).$$

Proof: We know from lemma 3.2 that

$$E|X_n(G \xrightarrow{\rho} H)| = E|X_n| + \rho \cdot (|H| - |G|) \cdot E[\mathrm{freq}\,(G \text{ in } X_n)].$$

Dividing the above expression by $E|X_n|$ yields

$$\frac{E|X_n(G \xrightarrow{\rho} H)|}{E|X_n|} = 1 + \rho(|H| - |G|) \cdot \mathrm{rel.freq}_E(G \text{ in } X_n)\frac{E|X_n| - |G| + 1}{E|X_n|}.$$

Therefore

$$\lim_{n\to\infty} \frac{E|V_n(G \xrightarrow{\rho} H)|}{|V_n|} = 1 + \rho(|H| - |G|) \cdot \mu(G).$$

Proposition 5.3 is proved.

6 Basic SO Acting on Measures

Given a measure μ and a triple $(G,\ \rho,\ H)$, a generic SO acting on measures is also denoted by $(G \xrightarrow{\rho} H)$, where G and H are words, G is self-avoiding and $\rho \in [0,\ 1]$. Informally speaking, this operator substitutes every entrance of the word G by the word H with a probability ρ or leaves it unchanged with a probability $1 - \rho$ independently of states and fate of the other components.

Now we want to define a general SO acting on measures. However, it is too difficult to do it in a straightforward way. Instead, we shall introduce several simple operators acting on random words, prove their consistency and represent a general SO acting on random words as a composition of those operators.

Recall the definition of consistent operators in definition 4.2. If both G and H are empty, our operator $(G \xrightarrow{\rho} H)$ leaves all measures unchanged by definition. Leaving this trivial case aside, we assume that at least one of the words G and H is non-empty.

Let us define several small classes of operators acting on random words, which we call *basic operators* and prove that all of them are consistent. In doing this we follow our previous setup of consistent operators to define how they act on measures (see definition 4.3).

Basic operator 1: Conversion $(g \xrightarrow{\rho} h)$ is the only linear operator in our list. For any two different letters $g,\ h \in \mathcal{A}$, we define *conversion from* g *to* h as a map from Ω to Ω. The conversion operator changes each occurrence of the letter g into the letter h with probability $\rho \in [0,1]$ or does not change it with probability $1-\rho$ independently of the states of the other occurrences. In various sciences a similar transformation is often called *substitution*.

Lemma 6.1. *The basic operator conversion is consistent.*

Proof: Let (V_n) be a sequence of words converging to μ. We know that the extension of this operator equals 1, that is

$$Ext(\mu\,|(\,g \xrightarrow{\rho} h)) = 1.$$

Therefore it is sufficient to verify that the following limit exists for all words W:

$$\lim_{n\to\infty} \frac{E[\text{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|},$$

since, from the expression of the extension of this operator, we have the identity

$$\lim_{n\to\infty} \frac{E[\text{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{E|V_n(g \xrightarrow{\rho} h)| - |W| + 1} = \lim_{n\to\infty} \frac{E[\text{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|}.$$

Indeed, denoting $m = \text{freq}\,(g \ \text{ in } \ V_n)$, it is easy to see that

$$\text{freq}\,(g \ \text{ in } \ V_n(g \xrightarrow{\rho} h)) \sim \text{Bin}(m,\ 1-\rho),$$

whence

$$E[\text{freq}\,(g \ \text{ in } \ V_n(g \xrightarrow{\rho} h))] = (1-\rho)\ \text{freq}\,(g \ \text{ in } \ V_n).$$

This yields

$$\lim_{n\to\infty} \frac{E[\text{freq}\,(g \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|} \ = \ (1-\rho) \lim_{n\to\infty} \frac{\text{freq}\,(g \ \text{ in } \ V_n)}{|V_n|} = (1-\rho)\cdot\mu(g).$$

After similar calculations we obtain

$$\lim_{n\to\infty} \frac{E[\text{freq}\,(h \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|}.$$

It is easy to see that

$$\text{freq}\,(h \ \text{ in } \ V_n(g \xrightarrow{\rho} h)) = \text{freq}\,(h \ \text{ in } \ V_n) + K,$$

where $K \sim \text{Bin}(m,\rho)$ represents the number of copies of the letter g that turned into h, and $m = \text{freq}\,(g \ \text{ in } \ V_n)$. Therefore

$$E[\text{freq}\,(h \ \text{ in } \ V_n(g \xrightarrow{\rho} h))] = \text{freq}\,(h \ \text{ in } \ V_n) + \rho\cdot\text{freq}\,(g \ \text{ in } \ V_n),$$

whence

$$\lim_{n\to\infty} \frac{E[\text{freq}\,(h \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|} = \lim_{n\to\infty} \frac{\text{freq}\,(h \ \text{ in } \ V_n)}{|V_n|} + \rho \lim_{n\to\infty} \frac{\text{freq}\,(g \ \text{ in } \ V_n)}{|V_n|} = \mu(h) + \rho\cdot\mu(g).$$

For any letter e different from g and h

$$\lim_{n\to\infty} \frac{E[\text{freq}\,(e \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|} = \lim_{n\to\infty} \frac{\text{freq}\,(e \ \text{ in } \ V_n)}{|V_n|} = \mu(e).$$

Thus we define how this operator acts on μ:

$$\mu(g \xrightarrow{\rho} h)(g) = \lim_{n\to\infty} \frac{E[\text{freq}\,(g \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|},$$

$$\mu(g \xrightarrow{\rho} h)(h) = \lim_{n\to\infty} \frac{E[\text{freq}\,(h \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|}$$

and

$$\mu(g \xrightarrow{\rho} h)(e) \ = \ \lim_{n\to\infty} \frac{E[\text{freq}\,(e \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|}.$$

Then, if we let $F(g|g) = 1-\rho$, $F(h|g) = \rho$ and $F(h|h) = F(e|e) = 1$, we have

$$\mu(g \xrightarrow{\rho} h)(g) = F(g|g)\mu(g) \ = \ (1-\rho)\cdot\mu(g),$$

$$\mu(g \xrightarrow{\rho} h)(h) = F(h|h)\mu(h) + F(h|g)\mu(g) \ = \mu(h) + \rho\cdot\mu(g),$$

$$\mu(g \xrightarrow{\rho} h)(e) = F(e|e)\ \mu(e) = \mu(e).$$

More generally, given a word $W = (a_1, \ldots, a_k)$, we have by similar calculations:

$$\lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|} = \sum_{b_1,\ldots,b_k \in \mathcal{A}} \left(\prod_{i=1}^{k} F(a_i|b_i) \times \mu(b_1, \ldots, b_k) \right),$$

where

$$F(a|b) = \begin{cases} 1-\rho & \text{if} \quad b = g \quad \text{and} \quad a = g, \\ \rho & \text{if} \quad b = g \quad \text{and} \quad a = h, \\ 0 & \text{if} \quad b = g \quad \text{and} \quad a \quad \text{is neither} \quad g \quad \text{nor} \quad h, \\ 1 & \text{if} \quad b \neq g \quad \text{and} \quad a = b, \\ 0 & \text{if} \quad b \neq g \quad \text{and} \quad a \neq b. \end{cases}$$

Lemma 6.1 is proved.

Now we can use consistency of this operator to define the conversion operator acting on any measure μ applied in any word $W = (a_1, \ldots, a_k)$:

$$\mu(g \xrightarrow{\rho} h)(W) = \sum_{b_1,\ldots,b_k \in \mathcal{A}} \left(\prod_{i=1}^{k} F(a_i|b_i) \times \mu(b_1, \ldots, b_k) \right).$$

Basic operator 2: Insertion $(\Lambda \xrightarrow{\rho} h)$. Insertion of a letter $h \notin \mathcal{A}$ into a random word in the alphabet $\mathcal{A}$ with a rate $\rho \in [0,1]$ means that a letter h is inserted with probability ρ between every two neighbor letters independently from other places. This term is used in molecular biology and computer science with a similar meaning [15].

We already know that the extension of any μ for this operator equals

$$\mathrm{Ext}(\mu|(\Lambda \xrightarrow{\rho} h)) = 1 + \rho.$$

Lemma 6.2. *The basic operator insertion is consistent.*

Proof: Let (V_n) be a sequence of words converging to some measure μ. We need to prove the following equation for any word W:

$$\lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(\Lambda \xrightarrow{\rho} h))]}{E|V_n(\Lambda \xrightarrow{\rho} h)| - |W| + 1} = \frac{1}{\mathrm{Ext}(\mu|(\Lambda \xrightarrow{\rho} h))} \lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(\Lambda \xrightarrow{\rho} h))]}{|V_n|}. \qquad (10)$$

First let us prove that the limits in the left and right sides of (10) exist.

Now, let a word W be in the alphabet $\mathcal{A}' = \mathcal{A} \cup \{h\}$. If W contains two consecutive appearances of h, then $E[\mathrm{freq}\,(W \ \text{ in } \ V_n(\Lambda \xrightarrow{\rho} h))] = 0$, otherwise:

$$E[\mathrm{freq}\,(W \ \text{ in } \ V_n(\Lambda \xrightarrow{\rho} h))] =$$

$$\sum_{i_j \in \{0,1\}} \mathrm{freq}\,(W \ \text{ in } \ V_n(i_1, \ldots, i_{|V_n|+1})) \cdot \rho^{\sum_j i_j} \cdot (1-\rho)^{|V_n|+1-\sum_j i_j}, \tag{11}$$

where $V_n(i_1, \ldots, i_{|V_n|+1})$ is the word obtained from V_n after inserting the letter h in those positions where $i_j = 1$. Further, let M be the number of pairs of consecutive letters in W, both of which are not h. It is clear that if $\sum_j i_j < \mathrm{freq}\,(h \ \text{ in } \ W)$ or if $|V_n|+1-\sum_j i_j < M$, then $\mathrm{freq}\,(W \ \text{ in } \ V_n(i_1, \ldots, i_{|V_n|+1})) = 0$. Let also

$$R = \{x \in \mathbb{N}; \mathrm{freq}\,(h \ \text{ in } \ W) \leq x \leq |V_n| + 1 - M\},$$

W' be the word obtained from W by deleting all the entrances of the letter h, and let $f_h(W) = \mathrm{freq}\,(h \ \text{ in } \ W)$. Note also that

$$\sum_{\substack{i_j \in \{0,1\} \\ \sum_j i_j = f_h(W)}} \mathrm{freq}\,(W \ \text{ in } \ V_n(i_1, \ldots, i_{|V_n|+1})) = \mathrm{freq}\,(W' \ \text{ in } \ V_n), \tag{12}$$

and, more generally, for $0 \leq k \leq |V_n| + 1 - M - f_h(W)$, we have

$$\sum_{\substack{i_j \in \{0,1\} \\ \sum_j i_j = f_h(W)+k}} \mathrm{freq}\,(W \ \text{ in } \ V_n(i_1, \ldots, i_{|V_n|+1})) = \binom{|V_n| + 1 - M - f_h(W)}{k} \mathrm{freq}\,(W' \ \text{ in } \ V_n). \tag{13}$$

Now we can simplify the expression (11) to obtain

$$\begin{aligned}
& E[\mathrm{freq}\,(W \ \text{ in } \ V_n(\Lambda \xrightarrow{\rho} h))] \\
= \ & \rho^{f_h(W)} \cdot (1-\rho)^M \sum_{\substack{i_j \in \{0,1\} \\ \sum_j i_j \in R}} \mathrm{freq}\,(W \ \text{ in } \ V_n(i_1, \ldots, i_{|V_n|+1})) \cdot \rho^{\sum_j i_j - f_h(W)} \cdot (1-\rho)^{|V_n|+1-\sum_j i_j - M} \\
= \ & \rho^{f_h(W)} \cdot (1-\rho)^M \cdot \mathrm{freq}\,(W' \ \text{ in } \ V_n) \times \\
& \sum_{k=0}^{|V_n|+1-M-f_h(W)} \binom{|V_n| + 1 - M - f_h(W)}{k} \cdot \rho^k \cdot (1-\rho)^{|V_n|+1-M-k-f_h(W)} \\
= \ & \mathrm{freq}\,(W' \ \text{ in } \ V_n) \cdot \rho^{\mathrm{freq}\,(h \text{ in } W)} \cdot (1-\rho)^M.
\end{aligned}$$

Hence it is easy to conclude that both limits in (10) exist. Now the fact that these limits are equal comes from the definition and existence of extension of this operator. *Lemma 6.2 is proved.*

Then we use consistency of this operator to define how operator $(\Lambda \xrightarrow{\rho} h)$ acts on any measure μ. We define the result of its application to an arbitrary word W by:

$$\begin{aligned}\mu(\Lambda \xrightarrow{\rho} h)(W) &= \frac{1}{\mathrm{Ext}(\mu|(\Lambda \xrightarrow{\rho} h))}\ \mu(W')\ \rho^{\mathrm{freq}\,(h \text{ in } W)}\ (1-\rho)^M \\ &= \frac{1}{1+\rho}\ \mu(W')\ \rho^{\mathrm{freq}\,(h \text{ in } W)}\ (1-\rho)^M.\end{aligned}$$

Basic operator 3: Deletion $(g \xrightarrow{\rho} \Lambda)$. Deletion of a letter $g \in \mathcal{A}$ with some probability $\rho \in [0,1)$ in a random word means that each occurrence of g disappears with probability ρ or remains unchanged with probability $1-\rho$ independently from the other occurrences. This term is also used in sciences with a similar meaning [15]. The extension of any measure μ under this operator is

$$\mathrm{Ext}(\mu|(g \xrightarrow{\rho} \Lambda)) = 1 - \rho \cdot \mu(g).$$

Lemma 6.3. *The basic operator deletion is consistent.*

Proof. Let (V_n) be a sequence of words converging to a measure μ. We need to prove that for any word W

$$\lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))]}{E|V_n(g \xrightarrow{\rho} \Lambda)| - |W| + 1} = \frac{1}{\mathrm{Ext}(\mu|(g \xrightarrow{\rho} \Lambda))} \lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))]}{|V_n|}, \tag{14}$$

but first we need to prove that both limits in (14) exist. Let $W = (a_0, \ldots, a_m)$ be any word with $|W| = m$, and $N_n = \mathrm{freq}\,(g \ \text{ in } \ V_n)$. Then from definition 2.10

$$E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))] = \sum_{k;\, j_1, \ldots, j_k} \mathrm{freq}\,(W \ \text{ in } \ V_n(k; j_1, \ldots, j_k)) \cdot \rho^k \cdot (1-\rho)^{N_n - k},$$

where $V(k; j_1, \ldots, j_k)$ is the word obtained from V by deletion of k letters g from positions $j_1, \ldots, j_k$.

Let $M = \mathrm{freq}\,(g \ \text{ in } \ W)$ and note that if $M > N_n$, then $E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))] = 0$. Fix some k in $\{0, \ldots, N_n - M\}$, and note that the equations (12) and (13), written in the context of the deletion operator, provide the following equation:

$$\sum_{j_1,\ldots,j_k} \mathrm{freq}\,(W \ \text{ in } \ V_n(k; j_1,\ldots,j_k))$$

$$= \sum_{n_1+\cdots+n_{m+1}\le k} \binom{N_n - M - (n_1+\cdots+n_{m+1})}{k-(n_1+\cdots+n_{m+1})} \mathrm{freq}\,((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \ \text{ in } \ V_n).$$

Multiplying the above expression by $\rho^k \cdot (1-\rho)^{N_n-k}$, summing over k and inverting the order of summation on the right-hand side of the equation yields

$$\sum_{k=0}^{N_n-M} \sum_{j_1,\ldots,j_k} \mathrm{freq}\,(W \ \text{ in } \ V_n(k; j_1,\ldots,j_k)) \cdot \rho^k \cdot (1-\rho)^{N_n-k}$$

$$= \sum_{n_1+\cdots+n_{m+1}\le N_n-M} \ \sum_{k=n_1+\cdots+n_{m+1}}^{N_n-M} \mathrm{freq}\,((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \ \text{ in } \ V_n)$$

$$\times \ \binom{N_n - M - (n_1+\cdots+n_{m+1})}{(n_1+\cdots+n_{m+1})-k} \cdot \rho^k \cdot (1-\rho)^{N_n-k}$$

$$= \sum_{n_1+\cdots+n_{m+1}\le N_n-M} \ \sum_{j=0}^{N_n-M-(n_1+\cdots+n_{m+1})} \mathrm{freq}\,((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \ \text{ in } \ V_n)$$

$$\times \ \binom{N_n - M - (n_1+\cdots+n_{m+1})}{j} \cdot \rho^{n_1+\cdots+n_{m+1}+j} \cdot (1-\rho)^{N_n-j-(n_1+\cdots+n_{m+1})}.$$

Then, we can use the previous equation to obtain that

$$E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))]$$

$$= (1-\rho)^M \sum_{k=0}^{N_n-M} \sum_{j_1,\ldots,j_k} \mathrm{freq}\,(W \ \text{ in } \ V_n(k; j_1,\ldots,j_k)) \cdot \rho^k \cdot (1-\rho)^{N_n-k-M}$$

$$= \sum_{n_1+\cdots+n_{m+1}\le N_n-M} \mathrm{freq}\,((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \ \text{ in } \ V_n) \cdot \rho^{n_1+\cdots+n_m+n_{m+1}} \cdot (1-\rho)^M$$

$$\times \sum_{j=0}^{N_n-M-(n_1+\cdots+n_{m+1})} \binom{N_n - M - (n_1+\cdots+n_{m+1})}{j} \cdot \rho^j \cdot (1-\rho)^{N_n-M-j-(n_1+\cdots+n_{m+1})}$$

$$= \sum_{n_1+\cdots+n_{m+1}\le N_n-M} \mathrm{freq}\,((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \ \text{ in } \ V_n) \cdot \rho^{n_1+\cdots+n_m+n_{m+1}} \cdot (1-\rho)^M.$$

Therefore

$$\frac{E[\mathrm{freq}\,(W \;\text{ in }\; V_n(g \stackrel{\rho}{\to} \Lambda))]}{|V_n|}$$

$$= \sum_{n_1+\cdots+n_{m+1}\le N_n-M} \frac{\mathrm{freq}\,((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \;\text{ in }\; V_n)\cdot \rho^{n_1+\cdots+n_{m+1}}\cdot(1-\rho)^M}{|V_n|}$$

$$= \sum_{n_1,\ldots,n_{m+1}=0}^{\infty} \mathrm{I}_{\{n_1+\cdots+n_{m+1}\le N_n-M\}} \frac{\mathrm{freq}\,((g^{n_1}, a_1, g^{n_2}, \ldots, a_m, g^{n_{m+1}}) \;\text{ in }\; V_n)\cdot \rho^{n_1+\cdots+n_{m+1}}\cdot(1-\rho)^M}{|V_n|},$$

where I_A is the indicator function of the set A; in the last identity the indicator function was used to avoid dependence of n in the index of summation. Clearly, $\mathrm{I}_{\{n_1+\cdots+n_{m+1}\le N_n\}} \stackrel{n\to\infty}{\longrightarrow} 1$ (i.e. it converges to the function, which is identically equal to 1). Also

$$\left| \mathrm{I}_{\{n_1+\cdots+n_{m+1}\le N_n-M\}} \frac{\mathrm{freq}\,((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \;\text{ in }\; V_n)}{|V_n|} \right| \le 1.$$

Since

$$\sum_{n_1,\ldots,n_{m+1}=0}^{\infty} \rho^{n_1+\cdots+n_{m+1}}\cdot(1-\rho)^M < \infty \quad \text{if} \quad \rho<1,$$

we may conclude from the dominated convergence theorem that

$$\lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \;\text{ in }\; V_n(g \stackrel{\rho}{\to} \Lambda))]}{|V_n|}$$

$$= \lim_{n\to\infty} \sum_{n_1,\ldots,n_{m+1}=0}^{\infty} \mathrm{I}_{\{n_1+\cdots+n_{m+1}\le N_n\}} \frac{\mathrm{freq}\,((g^{n_1}, a_1, \ldots, a_m, g^{n_{m+1}}) \;\text{ in }\; V_n)\cdot \rho^{n_1+\cdots+n_{m+1}}\cdot(1-\rho)^M}{|V_n|}$$

$$= \sum_{n_1,\ldots,n_{m+1}=0}^{\infty} \lim_{n\to\infty} \mathrm{I}_{\{n_1+\cdots+n_{m+1}\le N_n\}} \frac{\mathrm{freq}\,((g^{n_1}, a_1, \ldots, a_m, g^{n_{m+1}}) \;\text{ in }\; V_n)\cdot \rho^{n_1+\cdots+n_{m+1}}\cdot(1-\rho)^M}{|V_n|}$$

$$= \sum_{n_1,\ldots,n_{m+1}=0}^{\infty} \mu(g^{n_1}, a_1, \ldots, a_m, g^{n_{m+1}})\cdot \rho^{n_1+\cdots+n_{m+1}}\cdot(1-\rho)^M$$

Hence it is easy to conclude that both limits in (14) exist. In addition, we notice that the equality in (14) follows from the definition of extension. *Lemma 6.3 is proved.*

Now let us use consistency of this operator to define how the operator $(g \stackrel{\rho}{\to} \Lambda)$ acts on an arbitrary measure μ:

$$\begin{aligned}
&\mu(g \xrightarrow{\rho} \Lambda)(W) \\
= &\ \frac{1}{\mathrm{Ext}(\mu|(g \xrightarrow{\rho} \Lambda))} \sum_{n_1,\dots,n_{m+1}=0}^{\infty} \mu(g^{n_1}, a_1, \dots, a_m, g^{n_{m+1}}) \cdot \rho^{n_1+\cdots+n_{m+1}} \cdot (1-\rho)^M \\
= &\ \frac{1}{1-\rho\cdot\mu(g)} \sum_{n_1,\dots,n_{m+1}=0}^{\infty} \mu(g^{n_1}, a_1, \dots, a_m, g^{n_{m+1}}) \cdot \rho^{n_1+\cdots+n_{m+1}} \cdot (1-\rho)^M,
\end{aligned}$$

for all words $W = (a_1, \dots, a_m)$ and $M = \mathrm{freq}\,(g \text{ in } W)$.

Basic operator 4: Compression $(G \xrightarrow{1} h)$. Given a non-empty self-avoiding word G in an alphabet $\mathcal{A}$ and a letter $h \notin \mathcal{A}$, *compression from* G *to* h is the following map from $\Omega(\mathcal{A})$ to $\Omega(\mathcal{A}')$, where $\mathcal{A}' = \mathcal{A} \cup \{h\}$ and $\Omega(\mathcal{A})$ is the set of random words on the alphabet $\mathcal{A}$: each occurrence of the word G is replaced by the letter h with probability 1. The extension of any measure μ under this operator is

$$\mathrm{Ext}(\mu|(G \xrightarrow{1} h)) = 1 - (|G|-1)\cdot\mu(G).$$

Lemma 6.4. *The basic operator compression is consistent.*

Proof: Let (V_n) be a sequence of words converging to μ. We need to prove that

$$\lim_{n\to\infty} \frac{\mathrm{freq}\,(W \text{ in } V_n(G \xrightarrow{1} h))}{|V_n(G \xrightarrow{\rho} h)| - |W| + 1} = \frac{1}{\mathrm{Ext}(\mu|(G \xrightarrow{1} h))} \lim_{n\to\infty} \frac{\mathrm{freq}\,(W \text{ in } V_n(G \xrightarrow{1} h))}{|V_n|}. \tag{15}$$

Let us first prove that both limits in (15) exist. Let W be a word in the alphabet $\mathcal{A}'$. If there exist words U and V, with $|U| < |G|$ and $|V| < |G|$, satisfying $\mathrm{freq}\,(G \text{ in } W) < \mathrm{freq}\,(G \text{ in } \mathrm{concat}(U,W,V))$, then $\mathrm{freq}\,(W \text{ in } V_n) = 0$. Otherwise, notice that

$$\mathrm{freq}\,(W \text{ in } V_n(G \xrightarrow{1} h)) = \mathrm{freq}\,(W' \text{ in } V_n) - \mathrm{freq}\,(W \text{ in } G)\cdot\mathrm{freq}\,(G \text{ in } V_n),$$

where W' is the word obtained from W by replacing every letter h by the word G. Then

$$\begin{aligned}
\lim_{n\to\infty} \frac{\mathrm{freq}\,(W \text{ in } V_n(G \xrightarrow{1} h))}{|V_n|} &= \lim_{n\to\infty} \frac{\mathrm{freq}\,(W' \text{ in } V_n) - \mathrm{freq}\,(W \text{ in } G)\,\mathrm{freq}\,(G \text{ in } V_n)}{|V_n|} \\
&= \mu(W') - \mathrm{freq}\,(W \text{ in } G)\cdot\mu(G).
\end{aligned}$$

Hence it is easy to see that both limits in (15) exist. Now the equation (15) follows from the definition of extension. *Lemma 6.4 is proved.*

Now we use consistency of this operator to define how operator $(G \xrightarrow{1} h)$ acts on any measure μ:

$$\mu(G \xrightarrow{1} h)(W) = \frac{\mu(W') - \mathrm{freq}\,(W \ \text{ in } \ G) \cdot \mu(G)}{\mathrm{Ext}(\mu|(G \xrightarrow{1} h))} = \frac{\mu(W') - \mathrm{freq}\,(W \ \text{ in } \ G) \cdot \mu(G)}{1 - (|G| - 1) \cdot \mu(G)}.$$

Basic operator 5: Decompression $(g \xrightarrow{1} H)$. Given a non-empty self-avoiding word H in an alphabet $\mathcal{A}$ and a letter $g \notin \mathcal{A}$, *decompression of* g *to* H is the following map from $\Omega(\mathcal{A}')$ to $\Omega(\mathcal{A})$, where $\mathcal{A}' = \mathcal{A} \cup \{g\}$ and, again, $\Omega(\mathcal{A})$ is the set of random words on the alphabet $\mathcal{A}$: every occurrence of the letter g is replaced by the word H with probability 1. The extension of any measure μ for this operator is

$$\mathrm{Ext}(\mu|(g \xrightarrow{1} H)) = 1 + (|H| - 1) \cdot \mu(g).$$

Lemma 6.5. *The basic operator decompression is consistent.*

Proof: Let (V_n) be a sequence of words converging to the measure μ. We need to prove that for any word W

$$\lim_{n\to\infty} \frac{\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{1} H))}{|V_n(g \xrightarrow{\rho} H)| - |W| + 1} = \frac{1}{\mathrm{Ext}(\mu|(g \xrightarrow{1} H))} \lim_{n\to\infty} \frac{\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{1} H))}{|V_n|}. \tag{16}$$

Let us first prove that the limit in the right side of (16) exists. First let us consider the decompression of the letter g into the word (h_1, h_2) with probability 1, where the letters h_1 and h_2 are different and do not belong to the alphabet $\mathcal{A}$. The extension for this operator equals $1 + \mu(g)$. Now let us compute the following limit

$$\lim_{n\to\infty} \frac{\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{1} (h_1, h_2)))}{|V_n|}.$$

Let W be a word in the alphabet $\mathcal{A} \cup \{h_1, \ h_2\}$. We define a new word W' as the concatenation $W' = \mathrm{concat}(U, W, V)$, where

$$U = \begin{cases} h_1 & \text{if the first letter of } W \text{ is } h_2, \\ \Lambda & \text{otherwise} \end{cases} \quad \text{and} \quad V = \begin{cases} h_2 & \text{if the last letter of } W \text{ is } h_1. \\ \Lambda & \text{otherwise.} \end{cases}$$

After that we turn each entrance of the word (h_1, h_2) in W' into the letter g and denote the resulting word by W''. (We may do it since the word (h_1, h_2) is self-avoiding.) Now, if W'' contains any entrance of h_1 or h_2 it means that $\text{freq}\,(W \text{ in } V_n(g \xrightarrow{1} (h_1, h_2))) = 0$, and therefore the above limit equals zero. Otherwise,

$$\text{freq}\,(W \text{ in } V_n(g \xrightarrow{1} (h_1,\ h_2))) = \text{freq}\,(W'' \text{ in } V_n),$$

and thus,

$$\lim_{n\to\infty} \frac{\text{freq}\,(W \text{ in } V_n(g \xrightarrow{1} (h_1, h_2)))}{|V_n|} = \lim_{n\to\infty} \frac{\text{freq}\,(W'' \text{ in } V_n)}{|V_n|} = \mu(W'').$$

Therefore, if W'' contains any entrance of h_1 or h_2, we define $\mu(g \xrightarrow{1} (h_1, h_2)) = 0$, otherwise, we define:

$$\mu(g \xrightarrow{1} (h_1, h_2))(W) = \frac{\mu W''}{\text{Ext}(\mu|(g \xrightarrow{1} (h_1, h_2)))} = \frac{\mu(W'')}{1 + \mu(g)}.$$

Now, we will define the decompression of a letter g into the word $(h_1, \ldots, h_k)$ with probability 1, where the letters $h_1, \ldots, h_k$ are all different from each other and do not belong to the alphabet $\mathcal{A}$. Let us then define how the operator $(g \xrightarrow{1} (h_1, \ldots, h_k))$ acts on the measure μ by induction in k. The case $k = 2$ was treated above. Now, let us take $k > 2$ and a letter s not belonging to $\mathcal{A}$. Then for any word W in the alphabet $\mathcal{A} \cup \{h_1, \ldots, h_k\}$

$$\text{freq}\,(W \text{ in } V_n(g \xrightarrow{1} (h_1, \ldots, h_k))) = \text{freq}\,(W \text{ in } (V_n(g \xrightarrow{1} (h_1, s)))(s \xrightarrow{1} (h_2, \ldots, h_k))),$$

and then we can prove by induction that the following limits exist and the following equality holds:

$$\lim_{n\to\infty} \frac{\text{freq}\,(W \text{ in } V_n(g \xrightarrow{1} (h_1, \ldots, h_k)))}{|V_n|} =$$
$$\lim_{n\to\infty} \frac{\text{freq}\,(W \text{ in } V_n(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k)))}{|V_n|}.$$

Now it is easy to see that the limits in the equation (16) exist.

We will now use consistency of this operator to define how it acts on an arbitrary measure. Let $V_n \to \mu$

when $n \to \infty$ and assume that $h_1, \ldots, h_k, s$ do not belong to $\mathcal{A}$. Then

$$\begin{aligned}
&\mathrm{Ext}(\mu|(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k))) \\
= &\lim_{n\to\infty} \frac{|V_n(g \xrightarrow{1} (h_1, s)(s \xrightarrow{1} (h_2, \ldots, h_k))))}{|V_n|} \\
= &\lim_{n\to\infty} \frac{|V_n(g \xrightarrow{1} (h_1, s))| + (|H| - 2)\mathrm{freq}\,(s \ \text{ in } \ V_n(g \xrightarrow{1} (h_1, s)))}{|V_n|} \\
= &\lim_{n\to\infty} \frac{|V_n(g \xrightarrow{1} (h_1, s))| + (|H| - 2)\mathrm{freq}\,(g \ \text{ in } \ V_n)}{|V_n|} \\
= &\lim_{n\to\infty} \frac{|V_n| + (|H| - 1)\mathrm{freq}\,(g \ \text{ in } \ V_n)}{|V_n|} \\
= &\mathrm{Ext}(\mu|(g \xrightarrow{1} (h_1, \ldots, h_k))).
\end{aligned}$$

After that, we define how the operator $(g \xrightarrow{1} (h_1, \ldots, h_k))$ acts on an arbitrary measure μ in the following inductive way:

$$\mu(g \xrightarrow{1} (h_1, \ldots, h_k)) = \mu(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k)),$$

we can check this claim by noting that:

$$\begin{aligned}
&\mu(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k))(W) \\
= &\lim_{n\to\infty} \frac{\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k)))}{\mathrm{Ext}(\mu|(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k)))} \\
= &\lim_{n\to\infty} \frac{\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{1} (h_1, \ldots, h_k))}{\mathrm{Ext}(\mu|(g \xrightarrow{1} (h_1, \ldots, h_k)))} \\
= &\mu(g \xrightarrow{1} (h_1, \ldots, h_k))(W).
\end{aligned}$$

This is a composition of the decompression from g to (h_1, s) and the decompression from s into $(h_2, \ldots, h_k)$.

Finally, we can define the decompression operator acting on a measure μ. It transforms a letter g into an arbitrary word $H = (s_1, \ldots, s_k)$ with no restrictions on letters $s_1, \ldots, s_k$. First, we use the decompression from g to a word $(h_1, \ldots, h_k)$, where all the letters $h_1, \ldots, h_k$ are different from each other and do not belong to the alphabet $\mathcal{A}$. Further, we perform k conversions, each with probability 1, from h_i to s_i for all $i = 1, \ldots, k$. *Lemma 6.5 is proved.*

7 Compositions of Basic Operators

The main goal of this section is to give a general definition of SO $(G \overset{\rho}{\to} H)$ acting on measures. We shall do it by representing an arbitrary SO as a composition of several basic operators, which we have defined in the previous section.

Theorem 7.1. *Let $(G \overset{\rho}{\to} H)$, where G is self-avoiding, be a SO acting on words. Let also $(G \overset{\rho}{\to} \Lambda)$ and $(\Lambda \overset{\rho}{\to} H)$ be SO acting on words, where Λ is the empty word, s, g, h are different letters not belonging to $\mathcal{A}$, and $\rho \in [0,1]$ (and $\rho < 1$ for the operator $(G \overset{\rho}{\to} \Lambda)$). Then, for any words V and W:*

$$E[\text{freq}\,(W \;\text{ in }\; V(G \overset{\rho}{\to} H))] = E[\text{freq}\,(W \;\text{ in }\; V(G \overset{1}{\to} h)(h \overset{\rho}{\to} s)(s \overset{1}{\to} H)(h \overset{1}{\to} G))], \tag{17}$$

$$E[\text{freq}\,(W \;\text{ in }\; V(G \overset{\rho}{\to} \Lambda))] = E[\text{freq}\,(W \;\text{ in }\; V(G \overset{1}{\to} g)(g \overset{\rho}{\to} \Lambda)(g \overset{1}{\to} G))], \tag{18}$$

and

$$E[\text{freq}\,(W \;\text{ in }\; V(\Lambda \overset{\rho}{\to} H))] = E[\text{freq}\,(W \;\text{ in }\; V(\Lambda \overset{\rho}{\to} h)(h \overset{1}{\to} H))]. \tag{19}$$

Also

$$E|V(G \overset{\rho}{\to} H)| = E|V(G \overset{1}{\to} h)(h \overset{\rho}{\to} s)(s \overset{1}{\to} H)(h \overset{1}{\to} G)|, \tag{20}$$

$$E|V(G \overset{\rho}{\to} \Lambda)| = E|V(G \overset{1}{\to} g)(g \overset{\rho}{\to} \Lambda)(g \overset{1}{\to} G)|, \tag{21}$$

and finally,

$$E|V(\Lambda \overset{\rho}{\to} H)| = E|V(\Lambda \overset{\rho}{\to} h)(h \overset{1}{\to} H)|. \tag{22}$$

Proof of theorem 7.1: Observe that for any word V, the distributions of the random words

$$V(G \overset{\rho}{\to} H) \quad \text{and} \quad V(G \overset{1}{\to} h)(h \overset{\rho}{\to} s)(s \overset{1}{\to} H)(h \overset{1}{\to} G)$$

are one and the same. Therefore the mean frequency and the mean length are the same for both random words. The same argument holds for the other cases. *Theorem 7.1 is proved.*

Now let us state several corollaries, which will allow us to define SO on measures.

Corollary 7.2. *Let the SO*

$$(G \overset{\rho}{\to} H), \quad (G \overset{\rho}{\to} \Lambda), \quad (\Lambda \overset{\rho}{\to} H)$$

act on words. Here G *is a self-avoiding word,* $s, g, h \notin \mathcal{A}$, *and* $\rho \in [0,1]$ *(and* $\rho < 1$ *for the operator* $(G \xrightarrow{\rho} \Lambda)$ *). Then, for any words* V, W

$$\text{rel.freq}_E(W \text{ in } V(G \xrightarrow{\rho} H)) = \text{rel.freq}_E(W \text{ in } V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G)), \tag{23}$$

$$\text{rel.freq}_E(W \text{ in } V(G \xrightarrow{\rho} \Lambda)) = \text{rel.freq}_E(W \text{ in } V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G)), \tag{24}$$

and

$$\text{rel.freq}_E(W \text{ in } V(\Lambda \xrightarrow{\rho} H)) = \text{rel.freq}_E(W \text{ in } V(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H)). \tag{25}$$

Therefore, if (V_n) *is a sequence of words, then*

$$V_n(G \xrightarrow{\rho} H) \text{ converges } \iff V_n(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G) \text{ converges,} \tag{26}$$

$$V_n(G \xrightarrow{\rho} \Lambda) \text{ converges } \iff V_n(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G) \text{ converges} \tag{27}$$

and

$$V_n(\Lambda \xrightarrow{\rho} H) \text{ converges } \iff V_n(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H) \text{ converges.} \tag{28}$$

Proof: straightforward. *Corollary 7.2 is proved.*

To imitate an arbitrary operator $(G \xrightarrow{\rho} H)$, where G, H are words in an alphabet $\mathcal{A}$ and G is self-avoiding we first compress (with probability 1) each entrance of the word G into a letter h, which is introduced especially for this purpose and does not belong to $\mathcal{A}$. Then with probability ρ we turn each letter h into a letter $s \neq h$ which also does not belong to $\mathcal{A}$. After that we decompress (with probability 1) the letter s into a word H and decompress the letter h into a word G. We proceeded analogously to imitate the other operators.

Corollary 7.3. *For any words* G, H, *where* G *is self-avoiding, and any* $\rho \in [0,1]$ *(where* $\rho < 1$ *if* $H = \Lambda$ *), the operator* $(G \xrightarrow{\rho} H)$ *acting on words is consistent.*

Proof: The identities (23), (24) and (25) yield that the SO is the composition of basic operators described in the last section, and each basic operator is consistent. Thus, by lemma 4.4, their composition is also consistent. Thus $(G \xrightarrow{\rho} H)$ is consistent. *Corollary 7.3 is proved.*

In view of the above corollary, we have the following definition:

Definition 7.4. We define $\mu(G \xrightarrow{\rho} H)$, that is the result of application of the operator $(G \xrightarrow{\rho} H)$ to a measure $\mu \in \mathcal{M}$ (following definition 4.3) by

$$\mu(G \xrightarrow{\rho} H) = \lim_{n\to\infty} V_n(G \xrightarrow{\rho} H)),$$

where V_n is a sequence converging to μ.

Corollary 7.5. *Consider the operator $(G \xrightarrow{\rho} H)$ acting on measures, where G is self-avoiding and $\rho \in [0,1]$ ($\rho < 1$ if $H = \Lambda$). Then the following identities hold for any $s, g, h \notin \mathcal{A}$:*

$$\mu(G \xrightarrow{\rho} H) = \mu(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G),$$

$$\mu(G \xrightarrow{\rho} \Lambda) = \mu(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G),$$

$$\mu(\Lambda \xrightarrow{\rho} H) = \mu(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H).$$

Proof: It is a straightforward consequence of corollary 7.2. *Corollary 7.5 is proved.*

8 Segment-Preserving Operators

For any two measures μ, ν we denote by convex(μ, ν) their convex hull , that is

$$\text{convex}(\mu,\nu) = \{k\mu + (1-k)\nu \,|\, 0 \le k \le 1\}. \tag{29}$$

Lemma 8.1. *Let (V_n) and (W_n) be sequences of words converging to measures μ and ν respectively. Let the following limit exist*

$$\lim_{n\to\infty} \frac{|V_n|}{|V_n| + |W_n|} = L.$$

Then the sequence $\text{concat}(V_n,\ W_n)$ *converges to the measure $L \cdot \mu + (1-L)\cdot\nu$ when $n \to \infty$.*

Proof: We clearly have $|\text{concat}(V_n,\ W_n)| = |V_n| + |W_n|$ and also

$$\begin{aligned}\text{freq}\,(W \text{ in } V_n) + \text{freq}\,(W \text{ in } W_n) &\le \text{freq}\,(W \text{ in } \text{concat}(V_n, W_n)) \\ &\le \text{freq}\,(W \text{ in } V_n) + \text{freq}\,(W \text{ in } W_n) + 1.\end{aligned}$$

Hence,

$$\frac{\text{freq}\,(W \quad \text{in} \quad V_n)}{|V_n|+|W_n|} + \frac{\text{freq}\,(W \quad \text{in} \quad W_n)}{|V_n|+|W_n|} \leq \text{rel.freq}\,(W \quad \text{in} \quad \text{concat}(V_n, W_n)) \leq$$

$$\left(\frac{\text{freq}\,(W \quad \text{in} \quad V_n)}{|V_n|+|W_n|} + \frac{\text{freq}\,(W \quad \text{in} \quad W_n)}{|V_n|+|W_n|} + \frac{|W|}{|V_n|+|W_n|}\right) \cdot \frac{|V_n|+|W_n|}{|V_n|+|W_n|-|W|+1}.$$

Therefore

$$\frac{|V_n|-|W|+1}{|V_n|+|W_n|} \cdot \text{rel.freq}\,(W \quad \text{in} \quad V_n) + \frac{|W_n|-|W|+1}{|V_n|+|W_n|} \cdot \text{rel.freq}\,(W \quad \text{in} \quad W_n)$$

$$\leq \quad \text{rel.freq}\,(W \quad \text{in} \quad \text{concat}(V_n, W_n))$$

$$\leq \quad \left(\frac{|V_n|-|W|+1}{|V_n|+|W_n|} \cdot \text{rel.freq}\,(W \quad \text{in} \quad V_n) + \frac{|W_n|-|W|+1}{|V_n|+|W_n|} \cdot \text{rel.freq}\,(W \quad \text{in} \quad W_n)\frac{|W|}{|V_n|+|W_n|}\right)$$

$$\times \frac{|V_n|+|W_n|}{|V_n|+|W_n|-|W|+1}.$$

But

$$\frac{|V_n|-|W|+1}{|V_n|+|W_n|} \cdot \text{rel.freq}\,(W \quad \text{in} \quad V_n) + \frac{|W_n|-|W|+1}{|V_n|+|W_n|} \cdot \text{rel.freq}\,(W \quad \text{in} \quad W_n)$$

$$\to L \cdot \mu(W) + (1-L) \cdot \nu(W) \quad \text{as} \quad n \to \infty$$

and

$$\frac{|W|}{|V_n|+|W_n|} \to 0 \quad \text{as} \quad n \to \infty.$$

Also

$$\frac{|V_n|+|W_n|}{|V_n|+|W_n|-|W|+1} \to 1 \quad \text{as} \quad n \to \infty.$$

Thus the left and right sides of the above inequality tend to $L \cdot \mu(W) + (1-L) \cdot \nu(W)$. Then

$$\lim_{n\to\infty} \text{rel.freq}\,(W \quad \text{in} \quad \text{concat}(V_n, W_n)) = L \cdot \mu + (1-L) \cdot \nu.$$

That is, the sequence of words $\text{concat}(V_n,\ W_n)$ converges to the measure $L\mu + (1-L)\nu$. *Lemma 8.1 is proved.*

Lemma 8.2. *Let* (V_n) *be a sequence of words such that* $(V_n) \to \mu$ *and* (k_n) *a sequence of natural (therefore positive) numbers. Then* $V_n^{k_n} \to \mu$ *as* $n \to \infty$.

Proof follows immediately from the inequalities:

$$k_n \cdot \text{freq}\,(W \;\text{ in }\; V_n) \leq \text{freq}\,(W \;\text{ in }\; V_n^{k_n}) \leq k_n \cdot \text{freq}\,(W \;\text{ in }\; V_n) + k_n \cdot |W|$$

and from the fact that $|V_n^{k_n}| = k_n |V_n|$. *Lemma 8.2 is proved.*

Theorem 8.3. *For any* $L \in [0,1]$ *and any measures* μ *and* ν *there is a sequence of words* (V_n) *converging to* μ *and another sequence of words* (W_n) *converging to* ν, *such that* $\text{concat}(V_n,\ W_n)$ *converges to* $L \cdot \mu + (1-L) \cdot \nu$.

Proof: Take any sequences of words (V_n) and (W_n) such that $(V_n) \to \mu$ and $(W_n) \to \nu$ as $n \to \infty$. Then we construct a sequence of pairs $(\widetilde{V}_n,\ \widetilde{W}_n)$, such that $\widetilde{V}_n \to \mu,\ \widetilde{W}_n \to \nu$ and $|\widetilde{V}_n| = |\widetilde{W}_n|$. Indeed, let us consider

$$\widetilde{V}_n = V_n^{t_n}, \quad \text{where} \quad t_n = |W_n| \quad \text{and} \quad \widetilde{W}_n = W_n^{u_n}, \quad \text{and} \quad u_n = |V_n|.$$

Then we get $|\widetilde{V}_n| = |V_n| \cdot |W_n| = |\widetilde{W}_n|$.

Now we need to obtain a new sequence of pairs $(\widehat{V}_n,\ \widehat{W}_n)$, such that $\widehat{V}_n \to \mu$, $\widehat{W}_n \to \nu$ and

$$\text{concat}(\widehat{V}_n, \widehat{W}_n) \to L \cdot \mu + (1-L) \cdot \nu.$$

To do so, let $r \geq 0$ be given as $r = 1/L - 1$, and $r = +\infty$ if $L = 0$. Further, consider $r_n > 0$ a sequence of positive rational numbers such that $r_n \to r$. Let us write r_n in a more convenient way: $r_n = p_n/q_n$, where $p_n,\ q_n > 0$ are natural numbers. Then we take

$$\widehat{V}_n = \widetilde{V}_n^{q_n} \quad \text{and} \quad \widehat{W}_n = \widetilde{W}_n^{p_n}.$$

Noting that $|\widehat{V}_n| = q_n \cdot |\widetilde{V}_n|$, $\quad |\widehat{W}_n| = p_n \cdot |\widetilde{W}_n|$ and $|\widetilde{V}_n| = |\widetilde{W}_n|$. Thus $\widetilde{V}_n = V_n^{t_n + q_n}$, $\widetilde{W}_n = W_n^{u_n + p_n}$ and therefore, by lemma 8.2, $\widehat{V}_n \to \mu$ and $\widehat{W}_n \to \nu$. We also get that

$$\frac{|\widehat{V}_n|}{|\widehat{V}_n| + |\widehat{W}_n|} = \frac{1}{1 + p_n/q_n} \to \frac{1}{1+r} = L,$$

and by lemma 8.1, we have that

$$\mathrm{concat}(\widehat{V}_n, \widehat{W}_n) \to \frac{1}{1+r} \cdot \mu + \frac{r}{1+r} \cdot \nu = L\mu + (1-L)\nu.$$

Theorem 8.3 is proved.

The following definition gives us a useful property, which all SO have. This property is trivially satisfied for linear operators, but is not true in general.

Definition 8.4. An operator $P : \mathcal{M}_{\mathcal{A}} \to \mathcal{M}_{\mathcal{B}}$, where $\mathcal{A}$ and $\mathcal{B}$ are alphabets, is called *segment-preserving* if

$$\forall \mu,\ \nu \in \mathcal{M}_{\mathcal{A}} \quad \lambda \in \mathrm{convex}(\mu,\ \nu) \Rightarrow \lambda P \in \mathrm{convex}(\mu P,\ \nu P),$$

where $\mathrm{convex}(\mu,\ \nu)$ was defined in (29).

Theorem 8.5. *Every SO* $(G \xrightarrow{\rho} H)$ *is segment-preserving and*

$$(L \cdot \mu + (1-L) \cdot \nu)(G \xrightarrow{\rho} H) = \widetilde{L} \cdot \mu(G \xrightarrow{\rho} H) + (1-\widetilde{L}) \cdot \nu(G \xrightarrow{\rho} H)$$

for any measures $\mu,\ \nu$, *where*

$$\widetilde{L} = \frac{L \cdot \mathrm{Ext}(\mu|(G \xrightarrow{\rho} H))}{L \cdot \mathrm{Ext}(\mu|(G \xrightarrow{\rho} H)) + (1-L) \cdot \mathrm{Ext}(\nu|(G \xrightarrow{\rho} H))}. \tag{30}$$

Proof: The proof will be done by first obtaining a similar result for words, then going to the limit and finally proving it for measures. The key tool is theorem 8.3. Let $L \in (0,1)$. Due to theorem 8.3 we can take two sequences of words (V_n) and (W_n) converging to μ and ν, respectively, such that

$$\mathrm{concat}(V_n, W_n) \to L \cdot \mu + (1-L) \cdot \nu, \quad \text{as} \quad n \to \infty.$$

We want to show that the SO $(G \xrightarrow{\rho} H)$ satisfies this:

$$(L \cdot \mu + (1-L) \cdot \nu)(G \xrightarrow{\rho} H) = \widetilde{L} \cdot \mu(G \xrightarrow{\rho} H) + (1-\widetilde{L}) \cdot \nu(G \xrightarrow{\rho} H).$$

Let us choose any word W. Then

$$E[\mathrm{freq}\,(W \ \text{ in } \ V_n(G \xrightarrow{\rho} H))] + E[\mathrm{freq}\,(W \ \text{ in } \ W_n(G \xrightarrow{\rho} H))] \leq$$

$$E[\mathrm{freq}\,(W \ \text{ in } \ \mathrm{concat}(V_n, W_n)(G \xrightarrow{\rho} H))] \leq$$

$$E[\mathrm{freq}\,(W \ \text{ in } \ V_n(G \xrightarrow{\rho} H))] + E[\mathrm{freq}\,(W \ \text{ in } \ W_n(G \xrightarrow{\rho} H))] + 1.$$

Note that

$$\frac{1}{E|\mathrm{concat}(V_n,\ W_n)(G \stackrel{\rho}{\to} H)|} \stackrel{n\to\infty}{\longrightarrow} 0.$$

Therefore, to prove the convergence of the sequence of the words $\mathrm{concat}(V_n,\ W_n)(G \stackrel{\rho}{\to} H)$, it is sufficient to look at the limit values of

$$\frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(G \stackrel{\rho}{\to} H))] + E[\mathrm{freq}\,(W \ \text{ in } \ W_n(G \stackrel{\rho}{\to} H))]}{E|\mathrm{concat}(V_n,\ W_n)(G \stackrel{\rho}{\to} H)|}.$$

Notice further that

$$E|\mathrm{concat}(V_n,\ W_n)(G \stackrel{\rho}{\to} H)| \ = \ |V_n| + |W_n| + \rho(|H| - |G|)\cdot \mathrm{freq}\,(G \ \text{ in } \ \mathrm{concat}(V_n,\ W_n))$$

and furthermore that

$$\begin{aligned}\mathrm{freq}\,(G \ \text{ in } \ V_n) + \mathrm{freq}\,(G \ \text{ in } \ W_n) \ &\leq \ \mathrm{freq}\,(G \ \text{ in } \ \mathrm{concat}(V_n, W_n)) \\ &\leq \ \mathrm{freq}\,(G \ \text{ in } \ V_n) + \mathrm{freq}\,(G \ \text{ in } \ W_n) + 1.\end{aligned}$$

Due to the analogies between several parts of our argument, we will examine in detail only some cases and omit the others since they are analogous to those studied below. We want to sandwich the middle part of (31) between two values, say a_n and b_n, which we shall choose in an appropriate way:

$$a_n \ \leq \ \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(G \stackrel{\rho}{\to} H))] + E[\mathrm{freq}\,(W \ \text{ in } \ W_n(G \stackrel{\rho}{\to} H))]}{E|\mathrm{concat}(V_n, W_n)(G \stackrel{\rho}{\to} H)|} \ \leq \ b_n. \qquad (31)$$

First let us care about the right inequality in (31). To choose appropriate values of b_n we use the following inequalities:

$$\begin{aligned}&\frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(G \stackrel{\rho}{\to} H))] + E[\mathrm{freq}\,(W \ \text{ in } \ W_n(G \stackrel{\rho}{\to} H))]}{E|\mathrm{concat}(V_n,\ W_n)(G \stackrel{\rho}{\to} H)|} \\ = \ &\frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(G \stackrel{\rho}{\to} H))] + E[\mathrm{freq}\,(W \ \text{ in } \ W_n(G \stackrel{\rho}{\to} H))]}{|V_n| + |W_n| + \rho(|H| - |G|)\ \mathrm{freq}\,(G \ \text{ in } \ \mathrm{concat}(V_n, W_n))} \\ \leq \ &\frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(G \stackrel{\rho}{\to} H))] + E[\mathrm{freq}\,(W \ \text{ in } \ W_n(G \stackrel{\rho}{\to} H))]}{|V_n| + |W_n| + \rho(|H| - |G|)\cdot(\mathrm{freq}\,(G \ \text{ in } \ V_n) + \mathrm{freq}\,(G \ \text{ in } \ W_n))}.\end{aligned}$$

These inequalities suggest us to choose

$$b_n = \frac{E[\text{freq}\,(W \ \text{ in } \ V_n(G \xrightarrow{\rho} H))] + E[\text{freq}\,(W \ \text{ in } \ W_n(G \xrightarrow{\rho} H))]}{|V_n| + |W_n| + \rho(|H| - |G|) \cdot (\text{freq}\,(G \ \text{ in } \ V_n) + \text{freq}\,(G \ \text{ in } \ W_n))}. \tag{32}$$

It is evident that with these b_n the right inequality (31) holds. Analogously we can choose a_n to satisfy the left inequality in (31).

Now let us check the limiting behavior of b_n. We begin by checking the limiting behavior of the following quantity:

$$\frac{E[\text{freq}\,(W \ \text{ in } \ V_n(G \xrightarrow{\rho} H))]}{|V_n| + |W_n| + \rho(|H| - |G|) \cdot (\text{freq}\,(G \ \text{ in } \ V_n) + \text{freq}\,(G \ \text{ in } \ W_n))}.$$

The limit behavior of the other part included in b_n is obtained by simply replacing each entry of V_n by W_n, and each entry of W_n by V_n in the above expression. Therefore for the second case we will just give the resulting expression. Thus, we begin with:

$$\begin{aligned}
&\frac{E[\text{freq}\,(W \ \text{ in } \ V_n(G \xrightarrow{\rho} H))]}{|V_n| + |W_n| + \rho(|H| - |G|)(\text{freq}\,(G \ \text{ in } \ V_n) + \text{freq}\,(G \ \text{ in } \ W_n))} \\
= \ &\frac{|V_n|}{|V_n| + |W_n| + \rho(|H| - |G|)(\text{freq}\,(G \ \text{ in } \ V_n) + \text{freq}\,(G \ \text{ in } \ W_n))} \times S_n \times M_n \\
= \ &\frac{|V_n| + |W_n|}{|V_n| + |W_n| + \rho(|H| - |G|)(\text{freq}\,(G \ \text{ in } \ V_n) + \text{freq}\,(G \ \text{ in } \ W_n))} \times L_n \times S_n \times M_n,
\end{aligned}$$

where

$$L_n = \frac{|V_n|}{(|V_n| + |W_n|)} \quad \rightarrow \quad L,$$

$$S_n = \frac{E|V_n(G \xrightarrow{\rho} H)|}{|V_n|} \quad \rightarrow \quad \text{Ext}(\mu | G \xrightarrow{\rho} H)$$

$$M_n = \frac{E[\text{freq}\,(W \ \text{ in } \ V_n(G \xrightarrow{\rho} H))]}{E|V_n(G \xrightarrow{\rho} H)|} \quad \rightarrow \quad \mu(G \xrightarrow{\rho} H)(W).$$

Going on, we have

$$\frac{|V_n| + |W_n|}{|V_n| + |W_n| + \rho(|H| - |G|)(\mathrm{freq}\,(G \ \text{ in } \ V_n) + \mathrm{freq}\,(G \ \text{ in } \ W_n))} \times L_n \times S_n \times M_n$$

$$= \frac{L_n S_n M_n}{L_n \left[1 + \rho(|H| - |G|)\frac{\mathrm{freq}\,(G \ \text{ in } \ V_n)}{|V_n|}\right] + (1 - L_n)\left[1 + \rho(|H| - |G|)\frac{\mathrm{freq}\,(G \ \text{ in } \ W_n)}{|W_n|}\right]}$$

$$\overset{n\to\infty}{\longrightarrow} \frac{L \cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H))}{L \cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))}\mu(G \overset{\rho}{\to} H)(W)$$

By means of the analogous calculations, one can obtain:

$$\frac{E[\mathrm{freq}\,(W \ \text{ in } \ W_n(G \overset{\rho}{\to} H))]}{|V_n| + |W_n| + \rho(|H| - |G|)(\mathrm{freq}\,(G \ \text{ in } \ V_n) + \mathrm{freq}\,(G \ \text{ in } \ W_n))}$$

$$\overset{n\to\infty}{\longrightarrow} \frac{(1 - L) \cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))}{L \cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))}\nu(G \overset{\rho}{\to} H)(W).$$

Thus we have obtained the limiting behavior of b_n in the equation (32):

$$b_n \overset{n\to\infty}{\longrightarrow} \frac{L \cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H))}{L \cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))} \times \mu(G \overset{\rho}{\to} H)(W)$$

$$+ \frac{(1 - L) \cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))}{L \cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))} \times \nu(G \overset{\rho}{\to} H)(W).$$

Analogously, it is possible to find a sequence (a_n) satisfying equation (31) such that it has the same limit as (b_n), that is

$$a_n \overset{n\to\infty}{\longrightarrow} \frac{L \cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H))}{L \cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))} \times \mu(G \overset{\rho}{\to} H)(W)$$

$$+ \frac{(1 - L) \cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))}{L \cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))} \times \nu(G \overset{\rho}{\to} H)(W).$$

For instance, we may choose

$$a_n = \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(G \overset{\rho}{\to} H))] + E[\mathrm{freq}\,(W \ \text{ in } \ W_n(G \overset{\rho}{\to} H))]}{|V_n| + |W_n| + \rho(|H| - |G|)(\mathrm{freq}\,(G \ \text{ in } \ V_n) + \mathrm{freq}\,(G \ \text{ in } \ W_n) + |G|)}.$$

We conclude that

$$\operatorname{concat}(V_n, W_n)(G \xrightarrow{\rho} H) \overset{n\to\infty}{\longrightarrow}$$

$$\frac{L \cdot \operatorname{Ext}(\mu|(G \xrightarrow{\rho} H))}{L \cdot \operatorname{Ext}(\mu|(G \xrightarrow{\rho} H)) + (1-L) \cdot \operatorname{Ext}(\nu|(G \xrightarrow{\rho} H))} \times \mu(G \xrightarrow{\rho} H)$$

$$+\frac{(1-L) \cdot \operatorname{Ext}(\nu|(G \xrightarrow{\rho} H))}{L \cdot \operatorname{Ext}(\mu|(G \xrightarrow{\rho} H)) + (1-L) \cdot \operatorname{Ext}(\nu|(G \xrightarrow{\rho} H))} \times \nu(G \xrightarrow{\rho} H).$$

Thus, applying theorem 8.3, since $\operatorname{concat}(V_n,\ W_n) \to L \cdot \mu + (1-L) \cdot \nu,$ we obtain

$$(L \cdot \mu + (1-L) \cdot \nu)(G \xrightarrow{\rho} H) \ = \ \widetilde{L} \cdot \mu(G \xrightarrow{\rho} H) + (1-\widetilde{L}) \cdot \nu(G \xrightarrow{\rho} H),$$

for all $L \in (0,1)$, where $\widetilde{L}$ was defined in (30). *Theorem 8.5 is proved.*

9 Continuity and Invariant Measures

For any $\mathcal{M}' \subset \mathcal{M}$ we say that an operator $P : \mathcal{M}' \to \mathcal{M}'$ is *continuous* if whenever a sequence $\mu_n \in \mathcal{M}'$ tends to $\lambda \in \mathcal{M}'$ (in the weak topology, i.e., convergence separately on every word), the sequence $\mu_n P$ tends to λP (the well-known sequential continuity).

Definition 9.1. A measure μ is called *invariant* for an operator P if $\mu P = \mu$.

The article [13] indicated the following corollary of the well-known fixed point theorems:

Theorem 9.2. *For any non-empty compact convex $\mathcal{M}' \subset \mathcal{M}$ any continuous operator $P : \mathcal{M}' \to \mathcal{M}'$ has an invariant measure.*

We now state a general result about continuity of consistent operators.

Theorem 9.3. *Let $P : \Omega \to \Omega$ be a consistent operator. Given any non-empty compact convex $\mathcal{M}' \subset \mathcal{M}$, let $P : \mathcal{M}' \to \mathcal{M}'$ be the limit operator defined on measures (see definition 4.3). Then P (defined on measures) is continuous.*

Proof: Let μ be a measure in $\mathcal{M}'$ and let (μ_n) be a sequence of measures in $\mathcal{M}'$ converging to μ. Then $\mu_n(W) \to \mu(W)$ as $n \to \infty$ for every word W.

Let (V_{n_k}) be a sequence of words converging to μ_n as $k \to \infty$. We claim that the sequence (V_{k_k}) converges to μ as $k \to \infty$. Let $\varepsilon > 0$ be any positive number. We can choose k such that

$$|\text{meas}^{V_{k_k}}(W) - \mu_k(W)| < \varepsilon/2 \quad \text{and} \quad |\mu_k(W) - \mu(W)| < \varepsilon/2.$$

Then

$$|\text{meas}^{V_{k_k}}(W) - \mu(W)| \leq |\text{meas}^{V_{k_k}}(W) - \mu_k(W)| + |\mu_k(W) - \mu(W)| \leq \varepsilon.$$

Hence (V_{k_k}) converges to μ as $k \to \infty$.

Now, since P is consistent (see definition 4.2), we have $V_{n_k}P \to \mu_n P$ as $k \to \infty$ and $V_{k_k}P \to \mu P$ as $k \to \infty$. Therefore for any fixed $\varepsilon > 0$ and large enough k we have

$$|\mu_k P(W) - \text{meas}^{V_{k_k}P}(W)| < \varepsilon/2 \quad \text{and} \quad |\mu P(W) - \text{meas}^{V_{k_k}P}(W)| < \varepsilon/2.$$

Therefore

$$|\mu_k P(W) - \mu P(W)| \leq |\mu_k P(W) - \text{meas}^{V_{k_k}P}(W)| + |\mu P(W) - \text{meas}^{V_{k_k}P}(W)| \leq \varepsilon.$$

Theorem 9.3 is proved.

Now we note that $\mathcal{M}$ is convex and compact and apply theorem 9.2 to conclude the following:

Corollary 9.4. *Let $P : \mathcal{M}' \to \mathcal{M}'$ be the limit of consistent operators (see definition 4.3), where $\mathcal{M}'$ is a closed and convex subset of $\mathcal{M}$. Then P has at least one invariant measure.*

Proof: Since $\mathcal{M}'$ is closed and $\mathcal{M}$ is compact, $\mathcal{M}'$ also is compact. Further, by theorem 9.3, the operator P is continuous, threfore by theorem 9.2 P has an invariant measure. *Corollary 9.4 is proved.*

The next corollary applies these results to SO:

Corollary 9.5. *Every SO $(G \overset{\rho}{\to} H)$ (where $\rho < 1$ if $H = \Lambda$) is continuous and has an invariant measure.*

Proof: Take any $(G \stackrel{\rho}{\to} H)$. By corollary 7.3, it is consistent. Therefore by theorem 9.3, it is continuous. Then, by corollary 9.4, $(G \stackrel{\rho}{\to} H)$ has an invariant measure. *Corollary 9.5 is proved.*

Remark 9.6. We note that in [13], the proof of continuity of the SO is different from ours, since it proves that the basic operators are quasi-local and therefore continuous, and further, that any composition of continuous operators is continuous.

10 A Large Class of Substitution Processes

We now introduce a large class of stochastic processes which contains as particular cases, for instance, the process defined in [12]. For all these processes we prove existence of at least one invariant measure.

Definition 10.1. Let $\mu \in \mathcal{M}$ and let $(G \stackrel{\rho}{\to} H)$ (where $\rho < 1$ if $H = \Lambda$) be a SO. We define the *discrete substitution process* (μ_n) starting at μ, by

$$\mu_n(W) = \mu(G \stackrel{\rho}{\to} H)^n(W) \quad \text{for every word} \quad W.$$

Definition 10.2. Let $\nu \in \mathcal{M}$, and $P_1, \ldots, P_j$ be a finite sequence of SO. Then we define the *generalized discrete substitution process* (ν_n), where $\nu_0 = \nu$, as follows:

$$\nu_n(W) = \nu(P_1 P_2 \cdots P_j)^n(W) \quad \text{for every word} \quad W.$$

It is easy to see that the process defined in [12] is a special case of our generalized discrete substitution process, and further, that the substitution process itself is a special case of the generalized substitution process.

We will now apply the results of the last section to these processes. In fact, we will apply them to an even more general class of processes, which we will call the *consistent processes* and define them as follows:

Definition 10.3. Let $P : \mathcal{M}' \to \mathcal{M}'$ be the limit of consistent operators (see definition 4.3 for the definition of limit of consistent operators), and let μ be a measure in $\mathcal{M}'$. Then we say that (μ_n) is a *consistent process* starting at μ if

$$\mu_n(W) = \mu P^n(W) \quad \text{for all words} \quad W.$$

Since every SO is consistent (see corollary 7.3), and a composition of several consistent operators is also consistent, the generalized discrete substitution process is consistent.

Theorem 10.4. *Let* $P : \mathcal{M}' \to \mathcal{M}'$ *be the limit of consistent operators (see definition 4.3), where* $\mathcal{M}'$ *is a convex and closed subset of* $\mathcal{M}$ *. Then* P *has an invariant measure.*

Proof: a straightforward application of corollary 9.4. *Theorem 10.4 is proved.*

Remark 10.5. Since any generalized discrete substitution process is a special case of consistent processes, every generalized discrete substitution process has at least one invariant measure.

Theorem 10.6. *Let us consider a generalized discrete substitution process* $\nu_n = \nu_0 P^n$ *, where* $P = P_1 P_2 \cdots P_j$ *as in definition 10.2. Let* $S \subset \mathbb{A}^{\mathbb{Z}}$ *be some subset of the* σ *-algebra* $\mathbb{A}^{\mathbb{Z}}$ *. Then, if* $\nu_n(c) \leq \delta$ *(respectivelly* $\nu_n(c) \geq \varepsilon$ *) for all* $c \in S$ *, then* P *has an invariant measure* μ *such that* $\mu(c) \leq \delta$ *(respectivelly* $\mu(c) \geq \varepsilon$*) for every* $c \in S$ *, where* $\delta,\ \varepsilon > 0$ *are some positive constants.*

Proof: Let $\mathcal{M}'$ denote the closure in $\mathcal{M}$ of the convex hull of the measures $\nu_0,\ \nu_1, \ldots$. Therefore $\mathcal{M}'$ is a non-empty convex closed subset of $\mathcal{M}$. Since $\mathcal{M}$ is compact, $\mathcal{M}'$ is also compact. We now apply corollary 9.5 to note that P is continuous. Further, the continuity together with theorem 8.5 yields that if $\tau \in \mathcal{M}'$, then τP also belongs to $\mathcal{M}'$. Therefore, by theorem 9.2 the operator P has an invariant measure μ in $\mathcal{M}'$. Since for every n and every $c \in S$, $\nu_n(c) \leq \delta$ (respectivelly $\nu_n(c) \geq \varepsilon$), a simple calculation shows that for every $\tau \in \mathcal{M}'$ and every $c \in S$ we have $\tau(c) \leq \delta$ (respectivelly $\tau(c) \geq \varepsilon$). Therefore the invariant measure μ satisfies $\mu(c) \leq \delta$ (respectivelly $\mu(c) \geq \varepsilon$) for every $c \in S$. *Theorem 10.6 is proved.*

11 Application to Toom's Process

We now consider the process studied in [12], which is a special case of the generalized substitution process defined in section 10. In this case our alphabet is $\mathcal{A} = \{\oplus, \ominus\}$, whose elements are called *plus* and *minus*. We consider two specific operators: *flip* denoted by Flip_β and *annihilation* denoted by Ann_α. Flip_β is a special case of the basic operator which we called *conversion* (see section 6). More precisely, Flip $_\beta$ is $(\ominus \stackrel{\beta}{\to} \oplus)$, which turns every minus into plus with probability β independently from the fate of other components. Ann $_\alpha$ is $((\oplus, \ominus) \stackrel{\alpha}{\to} \Lambda)$, which makes every entrance of the self-avoiding word $(\oplus, \ominus)$ disappear with probability $\alpha < 1$ independently from fates of the other components. We therefore consider the sequence of measures

$$\mu_n = \delta_\ominus(\mathrm{Flip}_\beta \mathrm{Ann}_\alpha)^n, \tag{33}$$

where δ_0 is the measure concentrated in the configuration, all of whose components are zeros. [12, 8] have proved the following:

Theorem 11.1. *For all* $\beta \in [0, 1]$ *and* $\alpha \in (0, 1)$ *the relative frequency of pluses in the measure* μ_n *does not exceed* $250 \cdot \beta/\alpha^2$ *for all* n *.*

Now, based on these results, we can prove more:

Theorem 11.2. *For all* $\beta \in [0,\ 1]$ *and* $\alpha \in (0,\ 1)$ *the operator* $\mathrm{Flip}_\beta \mathrm{Ann}_\alpha$ *has an invariant measure, whose relative frequency of pluses does not exceed* $250 \cdot \beta/\alpha^2$ *.*

Proof: We may use theorem 10.6 with $S = \{\oplus\}$ and $\delta = 250 \cdot \beta/\alpha^2$ since by theorem 11.1

$$\mu_n(\oplus) < 250 \cdot \beta/\alpha^2 \quad \text{for all} \quad n.$$

Therefore, by theorem 10.6 the operator $\mathrm{Flip}_\beta \mathrm{Ann}_\alpha$ has an invariant measure ν such that

$$\nu(\oplus) \le 250 \cdot \beta/\alpha^2.$$

Theorem 11.2 is proved.

Corollary 11.3. *For* $\beta < \alpha^2/250$ *, the process* (33) *has at least two different invariant measures.*

Proof: On one hand the measure $\delta_\oplus$ concentrated in "all pluses" is invariant for the operator Flip $_\beta$ Ann $_\alpha$. On the other hand, by theorem 11.2 above, this operator has an invariant measure, in which the relative frequency of pluses does not exceed $250 \cdot \beta/\alpha^2$. Thus, with appropriate α and β this operator has at least two different invariant measures. *Corollary 11.3 is proved.*

Acknowledgements

The authors are grateful to CNPq and CAPES (Brazil) for financial support. We also would like to thank two anonymous reviewers for their generous attention and sensible suggestions that helped us improve the overall quality of the article.

References

[1] Gairat, A., Malyshev, V. Menshikov, M. and Pelikh, K. (1995): Classification of Markov chains describing the evolution of random strings. Russ. Math. Surv. **50** 237–255.

[2] Gairat A. , Iasnogorodski, R. and Malyshev, V. (1996): Null recurrent string. Markov Process Related Fields, **2**, 427–460.

[3] Janson, Svante (2007). Private communication.

[4] Malyshev, V. (1997): Interacting strings of symbols. Uspekhi Mat. Nauk, **52**, 59–86 (in Russian). English translation: Russian Math. Surveys, **52**, 299–326.

[5] Malyshev, V. (1998): Stochastic Evolution via Graph Grammars. Diskret. Mat., **10**, 30–44 (in Russian). English translation: Discrete Math. Appl., **8**, 247–262.

[6] Malyshev, V. (1998): Random grammars. Russ. Math. Surv. **53** 345–370.

[7] Malyshev, V. (2002): Quantum evolution of words. Theoretical Computer Science **273** 263–269.

[8] Ramos, A. D. and Toom, A. (2008) An error correction. Letter to the editor. Journal of Stat. Physics, **131**, n. 1, april 2008, pp. 167-168.

[9] Ramos, A. D. and Toom, A. (2008) Chaos and Monte-Carlo approximations of the Flip-Annihilation process. Journal of Stat. Physics, **133**, n. 4, pp. 761-771.

[10] Rocha, A. V. (2009): Substitution operators. Doctoral thesis, UFPE. In English. Available at `http://www.de.ufpe.br/~toom/alunos/doutorado/andrea/tese-andrea.pdf`

[11] Toom, A. (2002) : Particle systems with variable length. Bulletin of Brasilian Math Society, v. 33, n. 3, Nov 2002, pp. 419-425.

[12] Toom, A. (2004): Non-ergodicity in a 1-D particle process with variable length. Journal of Stat. Physics, **115**, 895–924.

[13] Toom, A. (2007): Every continuous operator has an invariant measure. Journal of Stat. Physics, **129**, 555–566.

[14] Wikipedia (2010): Damerau-Levenshtein distance.

[15] Wikipedia (2010): Indel.