# Random Grammars

Vadim A. Malyshev

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Random Grammars*

Vadim Malyshev

**N ° 3187**

Juin 1997

THÈME 1

Rapport de recherche

ISSN 0249-6399

# Random Grammars

Vadim Malyshev

Thème 1 — Réseaux et systèmes
Projet Meval

Rapport de recherche n ° 3187 — Juin 1997 — 36 pages

**Abstract:** This is the first part in a series of papers, where we consider new connections between computer science and modern mathematical physics. Here we begin to study a class of "concrete" random processes covering most of well known processes, such as locally interacting processes, random fractals, random walks, queueing networks, random Turing machines, etc. Here we restrict ourselves to linear graphs. We establish existence and uniqueness of the dynamics in the thermodynamic limit and prove that this dynamics is clustering. We get ergodicity and non-recurrence conditions in a small perturbation region. We study invariant measures and large time fractal type behaviour for random context free grammars and languages.

**Key-words:** grammars, L-systems, thermodynamic limit, cluster expansion, branching process.

*(Résumé : tsvp)*

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33) 01 39 63 55 11 – Télécopie : (33) 01 39 63 53 30

# Grammaires Aléatoires

**Résumé :** On commence a étudier une nouvelle classe des processus stochastiques, qui proviennent de nouveaux liens entre informatique et physique mathématique.

**Mots-clé :** grammaires, L-systèmes, limite thermodynamique, processus de branchement, processus à interaction locale.

# Contents

**1 Definitions** **4**
- 1.1 Grammars . . . . . 4
- 1.2 Random Grammars and $L$-systems . . . . . 6
- 1.3 Semigroup Representations . . . . . 8

**2 Infinite String Dynamics** **10**
- 2.1 Cluster expansion . . . . . 11
- 2.2 Cluster dynamics . . . . . 15
- 2.3 Local Observer . . . . . 18

**3 Large Time Behaviour: Small Perturbations** **20**
- 3.1 Invariant measures . . . . . 20
- 3.2 Classification . . . . . 24

**4 Large Time Behaviour: Context Free Case** **26**
- 4.1 Invariant measures for grammars . . . . . 26
- 4.2 $L$-systems. . . . . 30
- 4.3 Fractal Correlation Functions . . . . . 31
- 4.4 Measures on Languages . . . . . 34

# 1 Definitions

Let us consider configurations $\omega = (s_x) \in S^Z$ on the integer lattice $Z$ with values in some finite set $S$. It is well known that $S^Z$ is the standard state space for processes with a local interaction (see [6]) which act on this state space by randomly updating symbols $s_x$ at sites $x$. We want to consider processes where one symbol $s_x$ can be replaced say, for example, by two symbols. The question is where to put these symbols without destroying space homogeneity. To have sufficient space for this on the lattice one would have to move apart all other symbols, producing new enumeration for an infinite number of them. This procedure is intractable because on a finite time interval we should do an infinite number of such renumerations.

We discuss here rigorous definitions of such processes in one-dimensional case (that is for linear graphs). We establish existence and uniqueness of dynamics in the thermodynamic limit and prove that this dynamics is clustering. We get ergodicity and nonrecurrence conditions in a small perturbation region. We study invariant measures and large time fractal type behaviour for random context free grammars.

We widely use here some ideas from cluster expansions technology which has proved to be strongest tool in mathematical physics but has not yet been used in computer science.

This section contains mainly definitions and the simplest results. The main results of section 2 are theorems 1, 1 constructing the thermodynamic limit of the dynamics for finite time. The proofs use essentially cluster expansion techniques introduced in the same section. In section 3 we consider the simplest process with a local interaction and its small perturbation where vertices can be produced and killed. Main results are theorems 2, proving convergence to a unique invariant measure for infinite string dynamics and 3 giving conditions for transience and ergodicity for finite string dynamics. It is worth notice that these conditions can only be given in terms of invariant measure for infinite system. Section 4 deals with context free grammars outside of small perturbation region. We consider the large time limit of correlation functions. We find invariant measures (theorems 5,6). For degenerate cases we define fractal behaviour of grammars and finiteness of the number of critical exponents (theorems 7,8). Also the asymptotic probability distribution on the set of sentences of a context free language is found (theorem 9).

## 1.1 Grammars

Consider a finite set $S$ which we call the alphabet. A string $\alpha$ is a linearly-ordered (or completely ordered) sequence of symbols from $S$. For finite strings

$$\alpha = x_1 ... x_n, \beta = y_1 ... y_m$$

their concatenation $\alpha\beta$ is

$$\alpha\beta = x_1 ... x_n y_1 ... y_m$$

.

Concatenation of infinite strings is defined similarly.

Let $n = \mid \alpha \mid = l(\alpha)$ be the length of $\alpha$. Let $e = \emptyset$ be an empty string, so that

$$e\alpha = \alpha e = \alpha$$

We call $\gamma$ a substring of $\alpha$ if $\alpha = \beta\gamma\delta$ for some strings $\beta, \delta$.

Denote $S^*$ to be the set of all finite strings over the alphabet $S$, including the empty one.

Let $U$ a finite number of elementary substitutions (productions), i.e. ordered pairs of finite strings $\gamma_i \to \delta_i, i = 1, ..., k$. A grammar theory considers trajectories, i.e. sequences of strings $\alpha_1, ..., \alpha_k$ such that for each $j = 1, 2, ..., k-1$ string $\alpha_{j+1}$ is obtained from $\alpha_j$ by deleting some substring $\gamma_i$ of $\alpha_j$ and appending $\delta_i$ instead, i.e. replacing $\gamma_i$ by $\delta_i$.

A pair $(S, U)$ we shall call grammar. We shall always use this general definition but note that in computer science the definition of grammar is more restrictive. We remind it here.

Grammar (with nonterminal symbols) is a 4-tuple $G = (W, V, U, n_0)$ where

(1) $W$ is a finite set (its elements are called nonterminal symbols, variables, or syntactic categories ), $V$ is also a finite set such that $V \bigcap W = \emptyset$.

(2) $U$ is a finite subset of productions, i.e. pairs $u = (\alpha \to \beta)$ where $\alpha$ is a word over $S = W \bigcup V$ containing at least one symbol from $W$, $\beta \in S^*$.

(3) $n_0 \in W$ is a distinguished symbol (initial sentential form).

Sentential form is defined by : $n_0$ is a sentential form and if $\alpha\beta\gamma$ is a sentential form and $\beta \to \delta \in U$ then $\alpha\delta\gamma$ is a sentential form.

A sentence generated by $G$ is a sentential form containing no $W$-symbols.

Simplest classes of grammars are the following. $G$ is said to be

(1) *Linear* if each production is of the form $n \to l\alpha m$, where $l, n, m \in W, \alpha \in V^*$. It is called right linear if each production is of the form $n \to \alpha m$, where $n, m \in W, \alpha \in V^*$.

(2) *Context-free* if each production is of the form $n \to \alpha$, where $n \in W, \alpha \in S^*$.

Language $L$ over $\Sigma$ is a set of strings over $\Sigma$. Concatenation (product) of languages $L_1L_2$ is the set of all strings $\alpha\beta, \alpha \in L_1, \beta \in L_2$. The closure of $L$ is $\bigcup_{n=0}^{\infty} L^n, L^n = LL^{n-1}, L^0 = \{\emptyset\}$.

The language $L(G)$ generated by $G$ is the set of all sentences generated by $G$. Language $L(U, \alpha)$ generated by $U$ and $\alpha$ is the minimal set of strings satisfying the following conditions :

(i) $\alpha \in L(U, \alpha)$;

(ii) if $\rho\beta\gamma \in L(U, \alpha)$ and $(\beta \to \delta) \in U$ then $\rho\delta\gamma \in L(U, \alpha)$.

Language $L(G)$ has type (i) iff $G$ has type (i), $i = 1, 2$.

Lindenmayer theory of $L$-systems is a parallel analog of Grammar Theory. It considers trajectories where ALL possible substitutions should be done simultaneously. This poses some restrictions because ambiguities can arise. That is why normally only the case when the left side of each production is one symbol only.

The following are classes $L$-systems are similar to subclasses of context-free grammars:

- OL-system has all productions of the form $s \to \alpha, s \in S$ and at least one production $s \to \alpha$ for each $s \in S$;
- DOL-system (deterministic OL-system) is an OL-system with exactly one production for each $s$;
- Context-sensitive $L$-systems have productions like

$$\alpha x\beta \to \alpha\gamma\beta$$

## 1.2 Random Grammars and $L$-systems

Computer science studies languages generated by grammars. Then random grammars should study probability measures on languages.

Random grammar is the following countable Markov chain. Assume that on $U$ the nonnegative function $q(\alpha \to \beta) = q(\alpha, \beta)$ is defined. Countable continuous time Markov chain $\mathcal{G}(U, q)$ with state space $S^*$ is defined by the following transition rates: for any string $\alpha\gamma\beta$ and any production $\gamma \to \delta$ the rate of the transition $\alpha\gamma\beta \to \alpha\delta\beta$ is equal to $q(\gamma, \delta)$.

We discuss the following basic problems:

- Thermodynamic limit for such processes, existence and uniqueness;

- Classification of such chains. We use martingales and cluster expansion to provide explicit necessary and sufficient conditions for ergodicity and recurrence in a "small perturbation region";
- We study large time behaviour in transient context free cases and show that it can be decomposed on invariant measure type behaviour and fractal type behaviour.

Stochastic $L$-systems are discrete time systems, they were introduced earlier. But it seems that the terminology of Markov processes was not even known to the authors, see for example [3]. As a consequence the authors rediscovered some elementary results from branching processes. When we do trajectories random we get asynchronous (continuous time) dynamics for grammars and synchronous (discrete time) dynamics for $L$-systems.

Random grammars and $L$-systems are known to have applications to programming languages and to biological growth models.

**Turtle Dynamics** One can also define a local sequential alternative of such processes. The state space is now the set of all pairs $(\alpha, x_i)$ where $\alpha$ is a string and $x_i$ is some of the symbols of $\alpha$. One can think about a particle which is situated at this specified symbol. Each transition (in discrete or continuous time) consists of the creation by the particle, instead of the symbol $x_i$ where it is situated, two, one or zero symbols and simultaneous jump of the particle to one of the created symbols or to the neighbours of $x_i$.

**Existence** The continuous time homogeneous countable Markov chain $\mathcal{G}(U, q)$ is non exploding, i.e. one can construct a process with a.s. finite number of jumps on any time interval $[0, T]$ for any initial state. It is quite obvious because sum of the rates of jumps increasing the length of the string is dominated by the rates of the pure birth process on $Z_+$ with transition rates $\lambda(n \to n + d) = Cn$, where $n$ is the length of the string, $C$ is maximum of the transition rates for our process and $d$ is the maximal difference of string lengths in the productions. But such pure birth process with linear growth of transition rates is known to be nonexploding.

**Classification** The complete classification for the case of context-free grammars is provided below. Put $W(\alpha)$ equal to the number of $W$-symbols in $\alpha$.

**Proposition 1** *Assume that is there is only one $W$-symbol $n$ . For context free grammars the process is ergodic (null-recurrent, transient) iff*

$$\sum_{\alpha}(W(\alpha)-1)q(n,\alpha)<0$$

*(correspondingly,* $=0,\ >0$*).*

*Proof.* Note that $W$-symbols behave as the particles in the simple branching process. Then the assertion follows from well known results about branching processes.

If $W$ contains more than one symbol then one can also obtain the ergodicity and recurrence conditions using branching process with several types of particles. Similarly one get the classification for a little bit more general case when we do not make difference between terminal and nonterminal symbols and $U$ consists only of substitutions $s \to \beta$ with $s \in S$.

## 1.3 Semigroup Representations

**Assumption 1** *To simplify notation we shall consider in this section only the case when all productions are of the type*

$$\gamma y \delta \to \gamma \beta \delta$$

*for some* $y \in S, \delta, \gamma, \beta$ *where* $l(\gamma)+l(\delta) \leq 1, l(\beta) \geq 0$*. In this subsection however we do not need this assumption. Moreover, in the cluster expansion we need only to take more separating symbols (instead of one, as below) to deal with the general case.*

Let $H$ be the generator of our Markov process. We decompose it in some way as

$$H = H_0 + V = H_0 + \sum_a V_a$$

Below we shall use three types of such splitting. Using the differential equation

$$\frac{dW(t)}{dt} = W(t)e^{H_0 t} V e^{-H_0 t} ds$$

for $W(t) = e^{(H_0+V)t}e^{-H_0 t}$ one can get the following expansion

$$\exp(tH) = \exp(tH_0) + \int_0^t \exp(Hs)V\exp(H_0(t-s))ds = \exp(tH_0) +$$

$$\sum_{n=1}^{\infty}\int_0^t\int_0^{s_1}...\int_0^{s_{n-1}}\exp(H_0 s_n)V\exp(H_0(s_{n-1}-s_n))...$$

$$...V\exp(H_0(t-s_1))ds_n...ds_1 = \exp(tH_0)+$$

$$\sum_{n=1}^{\infty}\int_0^t\int_0^{s_1}...\int_0^{s_{n-1}}\sum\exp(H_0 s_n)V_{a_n}\exp(H_0(s_{n-1}-s_n))...$$

$$...V_{a_1}\exp(H_0(t-s_1))ds_n...ds_1 \quad (1)$$

where the last sum runs over all sequences $a_1,...,a_n$.

**Operator Representation.** Here we take $H_0 = 0$. Symbol $a = (u,i)$ consists of a particular production $u = (\gamma y\delta \to \gamma\beta\delta)$ and an integer $i \geq 1$, $V_a$ is a linear operator in the Banach space $l_1(S^*)$ which acts in the following way. We write down its action on measures from the right. Denote $\delta_\alpha$ the point measure on $S^*$ with support on string $\alpha$. Then

$$\delta_\alpha V_a = q(u)(\delta_{\alpha_1\gamma\beta\delta\alpha_2} - \delta_\alpha)$$

if $\alpha = x_1...x_n = \alpha_1\gamma y\delta\alpha_2$ so that $x_i = y$ if $a = (u,i)$. Otherwise we put

$$\delta_\alpha V_a = 0$$

In this case the expansion can be rewritten in simpler way

$$\exp(Ht) = \sum_{n=0}^{\infty}\frac{t^n}{n!}\sum_{a_n,...,a_1}V_{a_n}...V_{a_1} \quad (2)$$

**Lemma 1** *Now apply the last term of the expansion (2) to some $\delta_\alpha$. Then the number of terms for given $n$ has an upper bound $C_1(l(\alpha)C^n$ for some constants $C, C_1 = C_1(l(\alpha))$. Then for small $t$ the series is a norm analytic function of $q(.)$ and $t$ in the space of measures.*

*Proof.* For given $a_1,...,a_k$ consider the number of operators $V_{a_{k+1}}$ giving nonzero contribution to $\delta_\alpha V_{a_1}...,V_{a_k}V_{a_{k+1}}$. It does not exceed $l(\alpha)+Ck$ where $l(\alpha)$ is the length of the initial string. This gives the desired estimate.

Note that the norms of $V_a$ are uniformly bounded. Thus we get analyticity for small $t$.

**Trajectory Representation.** We take $H_0$ to be the diagonal part of $H$, it has negative elements on the diagonal. Symbol $a$ has the same meaning but now we put

$$\delta_\alpha V_a = q(u)\delta_{\alpha_1\gamma\beta\delta\alpha_2}$$

if $\alpha = \alpha_1\gamma y\delta\alpha_2$, also

$$\delta_\alpha V_a = 0$$

otherwise.

Note that the terms in this expansion have a precise probabilistic meaning. They give the following formulae for the density function in the space of trajectories with respect to Lebesgue measure on the union of simplexes

$$\{(s_n, ..., s_1) : s_n < s_{n-1} < ... < s_1 < t\}$$

More exactly, matrix elements $(\delta_\alpha e^{Ht}, \delta_{\alpha_1})$ are sums of integrals of the following products of basic matrix elements $V_{\alpha\beta}$ of $V$ and of the diagonal part

$$(e^{-H_0 s_n})_{\alpha\alpha} V_{\alpha\alpha^n} (e^{-H_0(s_{n-1}-s_n)})_{\alpha^n\alpha^n} ... V_{\alpha^2\alpha^1} (e^{-H_0(t-s_1)})_{\alpha^1\alpha^1}$$

where we assume that the trajectory $\omega$ starts from $\alpha$ and has $n$ consecutive jumps to $\alpha_n, ..., \alpha_1$ at times $s_n < ... < s_1$ respectively.

## 2 Infinite String Dynamics

Infinite string here is defined as a string which is isomorphic (as a completely ordered set) to $Z$. For finite time we would like to define the analog of the process $\mathcal{G}(U, q)$ for infinite strings as the thermodynamic limit of processes $\mathcal{G}(U, q)$ starting with finite strings of length $N$. We study also how this limit is related to large time limit. The answer is not as straightforward as for locally interacting processes and depends on the problem one considers. We discuss various approaches to the thermodynamic limit.

In the standard approach to thermodynamic limit one is looking for the limit of correlation functions. The situation with our processes is more delicate because all vertices constantly die and reappear. One of the questions is how to specify a point in which we want to find (e.g. one-point) correlation function. Second, one cannot expect to define an honorable process if we renumerate the infinite string after each transition. In finite interval we shall have infinite number of transitions and there is no unique natural enumeration at time $t$. One could say that there is no coordinate system for strings.

There is however a straightforward way to reduce infinite string dynamics to a finite strings dynamics. If such reduction exists we say informally that the dynamics is clustering (or cluster). We give two rigorous incarnations of this intuitive idea.

## 2.1 Cluster expansion

**Probabilistic expansion.** Let us start with any initial infinite string. Fix some symbol $x_0$ of this string labelling it with number 0. We get then natural enumeration $s(x) : Z \to S$ that is a one-to-one mapping of $Z$ onto the string. Fix some pair $(U, q)$ and consider the sequence of Markov processes $\mathcal{G}^N$ on probability spaces $\Omega^N$ which are copies of the process $\mathcal{G}(U, q)$ starting with the finite substring $x_{-N}...x_N$ of the initial infinite string. We want to study the limit $N \to \infty$.

Let $\Omega^N_{ij,t} \subset \Omega^N, -N-1 \leq i < 0 \leq j \leq N+1$, be the event that symbols $x_i$ and $x_j$ of the initial string $x_{-N}...x_N$ are not changed in the time interval $[0, t]$ and moreover there are no symbols in the substring $x_{i+1}...x_{j-1}$ with this property. Note that $i = -N-1$ means that all symbols $x_k, k < 0$ are updated in this time interval. Similarly for $j = N+1$.

Then a.s.

$$\bigcup_{i,j:-N\leq i<0\leq j\leq N+1} \Omega^N_{ij,t} = \Omega$$

**Theorem 1** *Probabilities $P^N(t; i, j) = P(\Omega^N_{ij,t})$ tend to some limit $P(t; i, j)$ as $N \to \infty$. Moreover*

$$\sum_{(i,j):i<0\leq j} P(t; i, j) = 1$$

Define the following probabilities for the process $\mathcal{G}(U, q)$ on the interval $[0, t]$

- $P^N_l(x_i)$ - probability that starting with initial string $x_{-N}...x_i$ extreme right symbol $x_i$ is not updated
- $P^N_r(x_j)$ - probability that starting with initial string $x_j...x_N$ extreme left symbol $x_j$ is not updated
- $P(x_i, x_j)$ - probability that starting with initial string $x_i...x_j$ both extreme right and left symbols $x_i, x_j$ are not updated but all other symbols are updated.

**Lemma 2** *For all $i, j, x_i, x_j, N$ the following formula holds*

$$P^N(t; i, j) \doteq P(\Omega^N_{ij,t}) = P^N_l(x_i) P(x_i, x_j) P^N_r(x_j)$$

This is intuitively clear because until one of the symbols $x_i$ and $x_j$ is not updated the process runs as 3 independent processes - left, middle and right - due to our assumption about productions that a symbol can be updated only separately through left or right context. We shall give however more arguments which we shall need further for another expansion.

First of all, we introduce finite string dynamics with boundary conditions $\mathcal{G}(U, q_{b.c.}(x,y)), x, y \in S \bigcup \{e\}$, where $\{e\}$ means empty boundary conditions. More exactly, this is a Markov chain where rate functions $q_{b.c.}(x,y)$ coincide with the rate function for $\mathcal{G}(U,q)$ with the following exception: we add transitions for the extreme left symbol $x_l$

$$x_l\delta \to \alpha\delta$$

with rates

$$q_{b.c.}(x_l\delta \to \alpha\delta)) = q(xx_l\delta \to x\alpha\delta)$$

inherited from $\mathcal{G}(U,q)$. Similarly, we add the transitions for the extreme right symbol $x_r$

$$\gamma x_r \to \gamma\alpha$$

with rates $q(\gamma x_r y \to \gamma\alpha y)$.

Let $\omega$ be a trajectory for the process $\mathcal{G}^N$ satisfying our conditions that $x_i, x_j$ are not updated but other symvols between them are updated. These notupdated symbols subdivide $\omega$ onto three parts $\omega_{<i}, \omega_{>j}, \omega_{ij}$ correspondingly from the left of $x_i$, from the right of $x_j$ and in the middle.

By $P(\omega), P_{b.c.}(\omega_{<i}), ...$, we denote density functions for trajectories of the processes $\mathcal{G}^N, \mathcal{G}(U, q_{b.c.}(e, x_i)), ....$ Note that, for example, $\omega_{<i}$ may be considered as a trajectory of the process $\mathcal{G}(U, q_{b.c.}(e, x_i))$ with boundary conditions.

Let $\omega$ have jumps at times $s_k$ in the part $\omega_{<i}$ and at times $t_k$ in the part $\omega_{ij}$. Denote $y(s)$ the symbol of the string of the trajectory $\omega$ at time $s$, which is next to the left from the symbol $x_i$. Consider the conditional probability (for the process $\mathcal{G}(U,q)$) that $x_i, x_j$, are not updated under the condition that the trajectory $\omega$ (outside these two symbols) is given. It is equal to

$$P(x_i, x_j \mid \omega) = P(x_i \mid \omega_{<i}) P(x_i, x_j \mid \omega_{ij}) P(x_j \mid \omega_{>j})$$

where for example

$$P(x_i \mid \omega_{<i}) = \prod_{k=1}^{n+1} \exp(-(s_{k-1} - s_k)(\sum_{\beta} q(y(s_k)x_i \to y(s_k)\beta))) \tag{3}$$

where $s_0 = t, s_{n+1} = 0$. But at the same time

$$P_l^N(x_i) = \int P(x_i \mid \omega_{<i}) d\mu_{b.c.}(\omega_{<i})$$

The same equalities can be written for two other probabilities and also for the whole process we get

$$P^N(t; i, j) = \int P(x_i, x_j \mid \omega) d\mu_{b.c.}(\omega)$$

where $\mu_{b.c.}$ is our process with excluded rates which could update $x_i, x_j$. Now we note that $\mu_{b.c.}$ is a product of three independent processes with boundary conditions

$$\mu_{b.c.} = \mu_{(e,x_i)} \mu_{(x_i,x_j)} \mu_{(x_j,e)}$$

Using this and the last formulae we get the result.

**Operator Expansion.** Introduce a structure of noncommutative algebra in $l_1(S^*)$ in the following way. Note that $\delta_\alpha$ form a basis of this space. Define a multiplication of these elements in the following way:

$$\delta_\alpha \otimes \delta_\beta \to \delta_\alpha \star \delta_\beta \doteq \delta_{\alpha\beta}$$

For the initial finite string $\alpha^{(N)} = x_{-N} ... x_N$ put

$$\alpha_{<i} = x_{-N} ... x_{i-1}, \alpha_{i,j} = x_{i+1} ... x_{j-1}, \alpha_{>j} = x_{j+1} ... x_N$$

We say that a linear operator $A$ in $l_1(S^*)$ admits a cluster expansion if there exist $\mid S \mid$ operators $A_{<,x}, x \in S$, $\mid S \mid$ operators $A_{>,x}, x \in S$, and $\mid S \mid^2$ operators $A_{0,x,y}, x, y \in S$, such that for all $\alpha = \alpha^N = x_{-N} ... x_N$,

$$\delta_\alpha A = \sum_{i,j: -N \leq i < 0 \leq j \leq N} (\delta_{\alpha_{<i}} A_{<,x_i}) \star \delta_{x_i} \star (\delta_{\alpha_{i,j}} A_{O,x_i,x_j}) \star \delta_{x_j} \star (\delta_{\alpha_{>j}} A_{>,x_j})$$

Our goal now is to give explicit cluster expansion for the semigroup $\exp tH$

**Lemma 3** *The following algebraic identity holds*

$$\delta_{\alpha^{(N)}} \exp(tH) =$$

$$\sum_{i,j: -N \leq i < 0 \leq j \leq N} [\delta_{\alpha_{<i}} exp(tH_{(e,x_i)})] \star \delta_{x_i} \star [\delta_{\alpha_{i,j}} L_{i,j}] \star \delta_{x_j} \star [\delta_{\alpha_{>j}} \exp(tH_{(x_j,e)})] \qquad (4)$$

*where $H$ with indices mean generators of Markov chains with the corresponding boundary conditions. The central term is $L_{-10} = 1$ and otherwise*

$$\sum_{n=j-i-1}^{\infty} \frac{t^n}{n!} V_{a_n} ... V_{a_1}$$

*where the last summation is over all $a_n, .., a_1$ such that each symbol of the string $\alpha_{ij}$ is updated at least once. In $L_{ij}$ we take all $V_a$ from the Markov chain with boundary conditions $(x_i, x_j)$.*

*Proof.* Note that $L_{ij}$ can be considered as the restriction of one operator (connected kernel) to strings of length $j - i - 1$. To get the algebraic expansion rewrite the sum $\sum_{a_n,...,a_1}$ in the right hand side of formula (2) as

$$\sum_{a_n,...,a_1} = \sum_{ij} \sum_{a_n,...,a_1}^{(ij)}$$

where the last sum is over all $a_n, ..., a_1$ such that

- No $V_a$ factors for symbols $x_i$ and $x_j$;
- At least one factor $V_a$ for each symbol of the initial string between $x_i$ and $x_j$.

It is useful to come back to the expansion (1). Call diagram $G$ a sequence $V_{a_n}, ..., V_{a_1}$. Contribution $Q(G)$ of this diagram is the corresponding term in the expansion including integration over the simplex of the time variables, correspondingly to the order in the sequence $V_{a_n}, ..., V_{a_1}$. Now we define a partition

$$A_{left} \bigcup A_{middle} \bigcup A_{right} = \{n, ..., 1\}$$

where these three subsets do not intersect each other: for example, $k \in A_{left}$ iff $V_{a_n}$ acts from the left of $x_i$. Correspondingly, $V_{a_k}$ and time variables $s_k$ fall into three subsets.

With a given diagram $G$ we associate three other diagrams $G_l, G_m, G_r$ as follows. $G_l$ is a subsequence of $V_{a_n}, ..., V_{a_1}$ with indices belonging to $A_{left}$. To get $G_m$ we also take a subsequence of $V_{a_n}, ..., V_{a_1}$ with indices belonging to $A_{middle}$ and modify the indices in the following way. If $k \in A_{middle}$ then for $a_k = (u_k, i_k)$ we define $\bar{a}_k = (u_k, j_k), j_k = i_k - l(\alpha_{<i}(.)) - 1$ where $\alpha_{<i}(.)$ is the string to the left of $x_i$ at the corresponding time moment. That is for the action $\delta_\alpha V_{a_k}$ we count $j_k$ starting from the first symbol to the right after $x_i$. Remind that we alsways know where the

conserved symbol $x_i$ is. Similarly $G_r$ is defined but in this case we count starting from the first symbol to the right after $x_j$.

With given diagram $G$ we associate a class of diagrams $C(G)$ consisting of all diagrams which can be obtained from $G$ by all allowable permutations of $V_{a_k}$. Permutation is called allowable if the order of each pair $V_{a_i}, V_{a_k}$ belonging to the same class $A_{left}$, $A_{middle}$ or $A_{right}$ is not changed.

Then

$$\sum_{G \in C(G)} Q(G) = Q(G_l)Q(G_m)Q(G_r)$$

This follows from the commutativity of $V_{a_i}V_{a_k} = V_{a_k}V_{a_i}$ belonging to different classes, using separation of time variables $s_j$.

**Remark 1** *The last formula gives some interesting interplay between concatenation and shuffle algebras. Concatenation algebra was defined above, the definition of shuffle algebra see in [10].*

## 2.2 Cluster dynamics

**Operator version.**

**Proposition 2** *From the operator expansion we see that for any $\epsilon > 0$ there exists $n = n(\epsilon) > 0$ and operators $M_n$ such that for every initial string $\alpha = y_{-N}...y_N$ the norm of*

$$\delta_\alpha(\exp(Ht) - M_n)$$

*is less than $a^n$ for some $a < 1$.*

*Moreover for some operators $A^{(n)} = A^{(n)}_{0,y_i,y_j}$, every $N$ and every initial string $\alpha = y_{-N}...y_N$*

$$M_n =$$

$$\sum \delta_{y_{-N}...y_{i-1}} \exp(Ht) \star \delta_{y_i} \star \delta_{y_{i+1}...y_{j-1}} A^{(n)} \star \delta_{y_j} \star \delta_{y_{j+1}...y_N} \exp(Ht)$$

*where the sum is over all $i, j$ such that $-2n \leq i < -n, n \leq j \leq 2n$.*

*Proof.* By induction where on the first we put $M_n = 0$. On the second step take the operator expansion (4). All terms of it with $-2n \leq i < -n, n < j \leq 2n$ we add to $M_n$. On the next step we expand each term of the operator expansion with either $-n \leq i < 0 \leq j \leq 2n$ or $-2n \leq i < 0 \leq j \leq n$ further. We do it in the following way. Take for example the case $-n \leq i, n < j \leq 2n$ and write for $\exp(H_{e,x_i})$ an expansion similar to the basic expansion (4)

$$\delta_{\alpha<i} \exp(H_{e,x_i}) = \sum_{k:k<i} \delta_{\alpha<k} \exp(H_{e,x_k}) \star \delta_{x_k} \star \delta_{x_{k+1}\ldots x_{i-1}} L^1_{ki}$$

We add terms with $2n \leq k < n$ to $M_n$ and for $-n \leq k$ continue in a similar way. The induction expires after finite number of steps.

**Probabilistic version.** The central idea of the cluster expansion is to further expand $P_l^N(x_i)$ and $P_r^N(x_j)$ in the same spirit. One cannot do it with only positive terms in the cluster expansion.

Let denote now $p_i^N(x_i) = P_l^N(x_i)$. Then

$$p_i^N(\bar{x}_i) = 1 - p_i^N(x_i) = p_i^N(\bar{x}_i, x_{i-1}) + p_i^N(\bar{x}_i, \bar{x}_{i-1})$$

Index $i$ at $p_i^N$ means that we consider our process starting with the string $x_{-N}...x_i$, the bar means that the symbol under the bar was updated. For example, $p_i^N(\bar{x}_i, x_{i-1})$ means that $x_i$ is updated and $x_{i-1}$ not. Then using arguments similar to lemma 2 we get

$$p_i^N(\bar{x}_i, x_{i-1}) = p_{i-1}^N(x_{i-1})P(x_{i-1})$$

where $P(x_{i-k})$ is the probability that in the process started with the string $x_{i-k...}x_i$ symbol $x_{i-k}$ is not updated but other symbols are updated. Also

$$p_i^N(\bar{x}_i, \bar{x}_{i-1}) = \sum_{k=2}^{\infty} p_i^N(\bar{x}_i, ..., \bar{x}_{i-k+1}, x_{i-k}) = \sum_{k=2}^{\infty} P(x_{i-k}) p_{i-k}^N(x_{i-k})$$

Our final expansion is thus

$$p_i^N(x_i) = 1 - \sum_{k=1}^{\infty} P(x_{i-k}) p_{i-k}^N(x_{i-k}) \tag{5}$$

Note that $P(x_{i-k}) = O(t^k)$ and by iterating the expansion we get the exponentially convergent series for $p_i^N(x_i)$.

*Proof of theorem 1.* Using smallness of $t$ we get

$$P(\Omega_{ij,t}^N) < a^{|j-i|}$$

for some $a = a(t) < 1$ (this fact can be easily proven without cluster expansion).

Iterating the expansion (5) we get a convergent series for $p_i^N(x_i)$ with terms not dependent on $N$ up to terms of order $a^N$ for some $a < 1$. Exponential convergence of $p_i^N(x_i)$ as $N \to \infty$ follows.

**Theorem-Definition 1** *For any $t$ there exists cluster dynamics on the set of infinite strings. This means the following.*

- *For any initial string $\alpha$ a random point set $A(\omega, t, \alpha)$ is defined. It consists almost surely of infinite number of symbols $x_{k_i}(\omega) = x_{k_i}(\omega, \alpha, t)$ in $\alpha$ such that*

$$... < x_{k_i}(\omega)... < x_{k_j}(\omega) < ...$$

*and which are the only symbols not involved into substitutions during whole time interval $[0, t]$. Let us take the agreement that $k_0$ is the smallest nonnegative index. These sets satisfy the following properties:*

1. *For all $t_1 < t_2$ a.s.*
$$A(\omega, t_2) \subset A(\omega, t_1)$$
2. *Given $x_{k_0}(\omega) = x^0$ and $x_{k_{-1}}(\omega) = x^{-1}$ conditional distributions of $x_{k_i}(\omega) - x_{k_{i-1}}(\omega), i < 0$, are independent and identically distributed.*

*The same of course holds in the direction to the right.*

- *Take some symbols*
$$... < x^i < ... < x^j < ...$$
*Then the dynamics under the condition that for all $i$*
$$x_{k_i}(\omega) = x^i$$
*consists of independent ditributions $\mu_i$ on finite string trajectories (in-between $x^i$ and $x^{i-1}$). These distributions are the restrictions of the process $\mathcal{G}(U, q)$ with initial strings $x_{k_i}...x_{k_{i+1}}$ to the trajectories such that starting with initial string $x_i...x_j$ both extreme right and left symbols $x_i, x_j$ are not updated but all other symbols are updated.*

We get this dynamics as a limit of finite string dynamics for small $t$. We use the cluster expansion. Let $\alpha^{(N)} = x_{-N}...x_N$ be the substring (of the length $2N+1$) of the infinite initial string $\alpha$ (we again fix some symbol $x_0$ of the initial string). We already proved convergence of distributions of symbols $x^0$ and $x^{-1}$ and of trajectories in-between these symbols. Now exactly in the same way, given $x_{k_{-1}}(\omega) = x^{-1}$ we find the ditribution of $x_{k_{-2}}(\omega)$, i.e. of the first symbol from the left of $x^{-1}$ which is not updated. The cluster expansion is the same and we shall not repeat it here. By induction we find all other symbols. Independence of increments for this random point process is clear. All other statements are proven similarly to the statements for trajectories between $x^0$ and $x^{-1}$.

The fact that random sets $A(t,\omega)$ are infinite a.s. for all $t$ follows by covering $[0,t]$ by intervals of sufficiently small length $t_0$. All other properties of dynamics for arbitrary $t$ are easily obtained because they hold uniformly in all initial conditions for $t < t_0$.

### 2.3 Local Observer

We saw that, for finite time, sets $A(\omega,t)$ together with fixing a symbol of the initial string gave us a good reference frame. When $t$ becomes infinite this reference frame dissappears (as $A(\omega,t) \to \emptyset$) and we need some means to understand where we are situated on the string. There exists one way to do it once and for all. But we shall see now that this way is extremely non constructive.

Consider the set $S^Z$ of configurations or functions on $Z$ taking values in $S$. This is a topological space equipped with the product topology. Also the group of translations acts on this topological space. Infinite string can be identified with an equivalence class of functions up to translations (shifts). Equivalence class can contain one string (if the function is constant), finite number of strings (if the function is periodic), otherwise it contains countable number of functions, all of them different from each other. The set $\mathcal{E}$ of all equivalence classes becomes a topological space in the induced topology. In this topology two infinite strings are close if there exists sufficiently long substring common to both representatives.

For Gibbs fields on $Z$ the thermodynamic limit is defined using local functions. To define a notion of a local function in our case one needs to use Zermelo's axiom of choice, i.e. to choose one representative from each equivalence class, that is to fix some symbol for each string. One can say that to find zero point (coordinate system) in space one needs Zermelo's axiom of choice. After this one could consider local function on these representatives and extend it by translations to all functions in the equivalence class. This is obviously intractable and we shall use more constructive appoaches to find reference frames.

I think that these difficulties have some fundamental nature (but completely different from nonexistence of points in noncommutative geometry), especially in higher dimensional situations - one cannot fix a point in "space" independently of the past.

Reference point can be related to a local observer by putting him at some point of the initial string and defining some rules of his jumping in time. Then we can look at the correlation functions at the points close to the observer and far away from it. One can put also several such observers and study their mutual disposition.

If the initial string is finite then the simplest way is put observer always to one of the ends of the string (for example, to the left one). More general way is to fix again a symbol $x_0$ of the initial string and assign number 0 to it. Then all remaining elements get their numbers automatically. One can imagine that a local observer sits at site 0. Until this symbol is not updated the observer stays at the same point. When $x_0$ is updated with some substitution $x_0 \to \alpha$ the observer jumps (using some Markovian rule) to one of the symbols of $\alpha$ or to one of its neighbours $x_{-1}, x_1$ at the moment. Then number zero is prescribed to the symbol where the observer is and the remaining elements are reenumerated correspondingly. To escape dissapearance in case $\alpha = e$, the observer should jump to one of its neighbours.

For a given observer one can define local correlation functions. A local observer provides us with a zero point and thus with an enumeration at any time $t$. Then we look at $P(s(-k) = s_{-k}, ..., s(k) = s_k), k > O$, at time $t$. It is clear that in the generic situation the random field on $Z$ defined by these correlation functions will not be space homogeneous.

As an example consider the case with substitutions of the form

$$\alpha \to \beta, \mid \alpha \mid = 1, \mid \beta \mid = 2$$

( context free grammars without terminal symbols).

Let the observer jumps always to the right symbol of the string $\beta$. Then at point 0 we shall see a finite Markov chain with transition intensities

$$\mu_{xy} = \sum_z q(x \to zy)$$

Denote its stationary distribution by $\pi$. Then stationary distribution of the correlation function

$$p(x_{-1} = b_{-1}, x_0 = b_0) = \sum_x \pi(x) \frac{q(x \to b_{-1} b_0)}{\sum_\beta q(x \to \beta)}$$

Contrary to this, the limiting distribution of $x_1$ is independent from $x_0$ and is the limiting distribution of the finite Markov chain with intensities

$$\mu_{xy} = \sum_z q(x \to yz)$$

For the same example consider $k$ observers observers put at different points at time 0. Here it is evident that local functions at neighbourhoods of different local observers become mutually independent.

See [1] where the case of the observer sitting at the end of the string is studied in more difficult cases.

# 3 Large Time Behaviour: Small Perturbations

Consider an independent process with local interactions starting with an infinite string and defined in the following way. With rate 1, i.e. after exponential waiting time with mean 1, the symbol in a given vertex becomes $r$ with probability $p(r), \sum_{r \in S} p(r)$. Invariant measure for this process is Bernoulli sequence with probabilities $p(r)$ of symbols. We shall call such transitions independent transitions.

Consider also a small perturbation of this process. Assume there are also rates $c(s_v \to \alpha, s_{O(v)})$ with which any symbol $s_v$ is replaced by a word $\alpha$. These rates depend on the configuration $s_{O(v)}$ in the neighbourhood $O(v)$ of $v$, $\alpha$ can be either empty or consists of one or two symbols. We assume further that all these functions $c(.)$ (there is a finite number of them) are small enough. This set of parameters we shall call the small perturbation region.

## 3.1 Invariant measures

We shall give two definitions of limiting measures.

**Definition 1** *Start with some infinite string and consider any local observer (by definition it is always at vertex* $0$*). Consider the correlation functions* $P(x_n(t)...x_{n+k-1}(t) = \gamma)$ *where* $\gamma$ *is a string of length* $k$*. Any limiting point of these functions for* $t \to \infty, n = n(t) \to \infty$*, we shall call limiting correlation functions.*

Note that at least one limiting point exists due to compactness.

**Empirical distribution.** In the space homogeneous case we can use an alternative approach. Fix some numeration at time 0 and assume that the initial distribution on functions $Z \to S$ is a stationary random process $\eta$. Let us consider the sequence of substrings $\alpha_N = x_{-N}...x_N$ of the initial infinite string $\alpha$. For fixed $N$ consider finite dynamics $\alpha_N(t)$ started from $\alpha_N$ with the initial distribution on it taken as the restriction of $\eta$. Define empirical one point correlation functions $q(\gamma, t)$. For this take the string $\alpha_N(t)$ at time $t$ and the number $Q(\gamma; t, N)$ of substrings $\gamma$ in it.

**Lemma 4** *For small* $t$

$$\frac{1}{N} Q(\gamma; t, N) \to q(\gamma; t)$$

*as $N \to \infty$ and for all $k$*

$$\sum_{\gamma : l(\gamma)=k} q(\gamma;t) = 1$$

This follows from the cluster arguments developped in the previous section taking into account that strings close to the boundary give a vanishing contribution to $Q$.

**Definition 2** *Limiting correlation functions are any limiting points $\mu(\gamma)$ of $q(\gamma;t)$ as $t \to \infty$. We call correlation functions $q(\gamma;t)$ invariant if they do not depend on $t$. We call them translation invariant if they define (by Kolmogorov theorem, if we enumerate $\gamma$ as $\gamma = x_0 ... x_{k-1}$) translation invariant measure on $S^Z$.*

**Theorem 2** *In the small perturbation region limiting correlation functions are unique. They coincide with invariant correlation functions which are also unique.*

We shall get explicit series for the correlation functions. In particular, for one-point correlation functions we have

$$\mu(r) = p(r) + O(c(.))$$

We shall also prove exponential convergence to this invariant measure.

*Proof.* We shall use the same kind of expansion as formula (1) where again

$$H = H_0 + V = H_0 + \sum_a V_a$$

but here we take $H_0$ to be the generator of independent process, i.e. the rate matrix for the independent transitions. $V$ corresponds to the small perturbation and is not positive

$$\delta_\alpha V_a = c(u)(\delta_{\alpha_1 \gamma\beta\delta\alpha_2} - \delta_\alpha)$$

where $u$ is the transition $\gamma y \delta \to \gamma\beta\delta$. Note two properties of $V_a$:

1. Its norm is $2c(.)$. Thus the total variation decreases not less than $2c(.)$ times.
2. For any measure $\mu$ the total charge of the measure $\mu V_a$ is zero.

We can write

$$V_a = V_a^+ + V_a^-$$

where

$$\delta_\alpha V_a^+ = c(u)\delta_{\alpha_1 \gamma\beta\delta\alpha_2}, \delta_\alpha V_a^- = -c(u)\delta_\alpha$$

We shall use the expansion (1). Note that in the interval $[s_k, s_{k-1}]$ the vertices do not change and we can write

$$\exp(H_0\Delta s) = \prod_v \exp(H_{0,v}\Delta s) = \prod_v (P_0 + W_v(\Delta s))$$

where $\Delta s = s_{k-1} - s_k$ and $\exp(H_{0,v}\Delta s)$ is independent updating process in one vertex, $P_0$ is the linear operator which transforms any probability distribution on $S$ to the distribution $p(r)$ on $S$.

The norm of their difference $W_v(\Delta s) = \exp(H_{0,v}\Delta s) - P_0$ tends to zero exponentially fast as $s \to \infty$. Choose $d$ so that $W_v(d) < \epsilon$ .

Now we introduce diagrams, i.e. directed graphs in $R_+ \times V$ where time $R_+$ will be vertical direction, and countable set $V$ - horizontal. Diagrams will label the terms of the expansion. The vertices of diagrams are labelled by $(s, v)$. At time 0 the set of vertices coincides with the set of symbols (vertices) of the initial string. We draw a vertical line from each vertex $(0, v)$ until it meets some $V_a^+$ or $V_a^-$ which update this vertex, at time $s_k$. Assume that vertical lines are directed to smaller times.

In both cases we draw a new vertex $(s_k, v)$. From it we draw horizontal lines to other new vertices $(s_k, v_i)$ ((in the direction from $(s_k, v_i)$ ) )for all symbols $x_{v_i}$ belonging to the neighbourhood of $v$ which gave rise to the transition. In the first case (for $V^+$) we draw also new vertices $(s_k, w)$ which appeared instead of the updated symbol $x_v$. We connect these new vertices by horizontal lines with vertex $(s_k, v)$ (in the direction to $(s_k, v)$ ).

By induction we continue in this way until we come to time $T$ slice.

Contribution of the diagram

$$Q(G) = \int \prod_{lines} Q(l) \prod_{vertices} Q((s_k, v))$$

where only vertices $(s_k, v)$ will give contribution. They are equal to $V_a^{\pm}$.

Contribution of each vertical line of length smaller than $d$ is $\exp(H_0\Delta s)$. Contribution of each vertical line of length greater than $d$ is either $P_0$ or $W(s_+ - s_-)$ where $s_+$ is the time coordinate of the upper vertex of the line, $s_-$ - lower coordinate. Contribution of each horizontal line is 1.

As it is standard in cluster expansions for large $T$ we first take $c(.)$ sufficiently small and show (by appropriate resummation) that the radius of convergence (analyticity region) is independent of $T$.

Take some vertex $g = (T, v)$ at time $T$ sufficiently large. Let $v$ be on distance exactly $L$ from the left end of the string at time $T$ and sufficiently far away from both ends of the string at time $T$. We assume that the string is finite and its length $N$ is much greater than $L$ and $T$. We shall prove that the one-point correlation function at this vertex converges to a limit independently of how $T$ and $L$ tend to infinity.

To prove this we shall define for each diagram $G$ a graph $\mathcal{T}(G)$, a cluster of this fixed vertex. $\mathcal{T}(G)$ is defined as the maximal connected directed subgraph of $G$ containing $g = (T, v)$ itself and not containing any line with contribution $P_0$, and moreover for each vertex there is only one ingoing line (from bigger times).

Now fix some $\mathcal{T}$, we define its contribution as

$$Q(\mathcal{T}) = \sum_{G:\mathcal{T}(G)=\mathcal{T}} Q(G)$$

Resummation also will give us a simpler formula to calculate $Q(\mathcal{T})$. Consider the lowest vertices of $\mathcal{T}$, that is the vertices from which there are no more outgoing lines belonging to $\mathcal{T}$. They can be of two kinds: we denote vertices lying on time slice zero by $V_0$ and the rest by $V_1$. We remember that in $G$ just below this vertex $v \in V_1$ there is a line $l$ with contribution $P_0$. If under it there is a vertex not on slice zero, then we make resummation using $V_a P_0 = 0$. Thus only lines $l$ ending at slice zero are left. They provide the joint probability distribution in the vertices if the tree just above lines with $P_0$ which is independent Bernoulli. Assume first that there is no vertices in $\mathcal{T}$ on time slice zero. Then the contribution $Q(\mathcal{T})$ of $\mathcal{T}$ is

$$\mu_B Q(\mathcal{T}) = \mu_B \prod_l Q(l) \prod_v Q_v$$

where $Q(l)$ is either $W(\Delta s)$ or $e^{H_0 \Delta s}$ and $Q_v$ is $V_a^{\pm}$.

**Lemma 5** *For any $\epsilon > 0$ sufficiently small there is such $c_0$ that for all $c(.) < c_0$ we have*

$$norm(\mu_B Q(\mathcal{T})) < (\epsilon)^{l(\mathcal{T})+n(\mathcal{T})}$$

*where $l\mathcal{T}$ is the sum of lengths of lines of $\mathcal{T}$ with lengths greater than $d$, $n(\mathcal{T})$ - number of vertices of $\mathcal{T}$.*

Similarly, if $\mathcal{T}$ has vertices on time slice zero, one can prove that their contribution tends to zero as $\epsilon^T$ when $T \to \infty$. Thus,using standard methods of summation over trees one can prove the following theorem.

**Theorem 3** *We have the following formula for the one point limiting correlation function*

$$p(.) = \sum_{\mathcal{T}} Q(\mathcal{T})$$

*Correlation functions at time $t$ converge to it exponentially fast.*

**Corollary 1** *Let us consider infinite string process and arbitrary local observer. For any $L_1, L_2 \in Z$ denote $P(\gamma_i, L_i, t_i), i = 1, 2$, the probabiliies of events that at times $t_1, t_2 = t_1 + \tau$ the string starting from $L_i$ is $\gamma_i$ $P(\gamma_1, L_1, t_1; \gamma_2, L_2, t_2)$ the probability of their intersection. Then the following limits exist*

$$\lim_{t_i \to \infty} P(\gamma_i, L_i, t_i) = \mu(\gamma_i)$$

$$P(\gamma_1, L_1; \gamma_2, L_2; \tau) = \lim_{t_1 \to \infty} P(\gamma_1, L_1, t_1; \gamma_2, L_2, t_2)$$

*and the covariances*

$$P(\gamma_1, L_1; \gamma_2, L_2; \tau) - \mu(\gamma_1)\mu(\gamma_2)$$

*tend to zero exponentially fast with $\tau$.*

The proof of this theorem is along the same lines using cluster expansion introduced above.

## 3.2 Classification

Intuitively, the strategy of getting stability results is the following. A given vertex can produce instantaneously zero, one or two vertices. But the production rates depends also on the environment. We already know that some limiting "local" invariant measure will be established in the system after some time. This limiting measure appears to be the unique limiting measure $\mu$ for infinite string process. Exactly this invariant measure will give us infinitesimal mean production rates. Thus the stability condition should be the following.

**Theorem 4** *For fixed $v$ we put $c(s_{O(v)}, \alpha) = c(s_v \to \alpha, s_{O(v)})$. We call*

$$M = \sum_{s_{O(v)}} \mu(s_{O(v)})(-c(s_{O(v)}, \emptyset) + \sum_{r,s} c(s_{O(v)}, rs))$$

*the infinitesimal mean production rate. The chain is ergodic if $M < 0$, transient if $M > 0$.*

We only sketch the proof for the transient case, ergodic is quite similar. The case when already

$$M_0 = \sum_{s_{O(v)}} \mu_0(s_{O(v)})(-c(s_{O(v)}, \emptyset) + \sum_{r,s} c(s_{O(v)}, rs)) > O$$

for Bernoulli measure $\mu_0$ invariant with respect to independent transitions. More delicate is the case when $M_0 = 0$, i.e. only second order terms are positive.

We will use Lyapounov function (see [11]), which is taken to be just the number of vertices in the string.

Assume thus that $M > 0$. We shall prove that there exists such $\tau > 0$, $N > 0$ and $\epsilon > 0$ such that for any string $\alpha(0) = \alpha, l(\alpha) > N$, at time 0 we have

$$E(l(\alpha(\tau)) \mid \alpha(0) = \alpha) - l(\alpha) > \epsilon l(\alpha)$$

We shall choose $1 << \tau << N$. So take arbitrary string of length $N$ at time 0, all probabilities will be conditional under the condition that this string is fixed. Let $l(t)$ be (random) length of the string at time $t$. Take some $T$ so that at time $T$ the distribution were sufficiently close to the invariant measure. Note that during time interval $[0, T]$ the decrease of the length is of order $c(.)T$.

From our assumption it follows that

$$E(l(T+1) - l(T)) > \epsilon l(T)$$

for some small $\epsilon$. Then take $\tau$ that during time interval $[T, \tau]$ we have considerable increase of the length, i.e.

$$E \sum_{k=1}^{[\tau - T]} (l(T+k) - l(T+k-1)) > \epsilon l(T)(\tau - T)$$

From this the result follows.

**Corollary 2** *Null recurrent case is of Lebesgue measure zero in the parameter space.*

It is an interesting question of whether the last statement holds in the general situation (without assumption about smallness of some parameters).

**Conjecture 1** *If $M = 0$ then the chain is null recurrent.*

*About the proof.* The case $M = 0$ is more complicated. It is difficult to find an exact martingale here necessary for criteria of null-recurrence (see [11]). Instead one could use change of measure for different times and coupling of finite and infinite string dynamics.

**Corollary 3** *(Classification for Infinite Dynamics) In the small perturbation region transient case for finite string holds iff for the corresponding infinite strings anyone of the following conditions take place:*

1. *For any two local observers the distance between them tends to $\infty$ with positive probability;*

2. *There exist nonzero bound for the density of survived observers uniform in the initial distribution. More exactly, if at time zero we have the density of observers on the initial string say $\rho_0$ then at all times the density is larger than $c\rho_0$ for some constant $c > 0$.*

# 4 Large Time Behaviour: Context Free Case

I did not go to maximal generality here, because there are too many types of degeneration of the considered processes. This is not too difficult but sufficiently boring work. But I tried to give interesting examples and considered some cases which seemed typical.

## 4.1 Invariant measures for grammars

**One nonterminal symbol.** First we consider context-free grammars and assume that $\mid W \mid = 1$. Fix $x_0$ in the initial string. Then the initial string becomes a configuration on $Z$. Denote by $\mathcal{M}$ the class of measures on $S^Z$ such that

$$\liminf \frac{1}{N} \#\{i : x_i \in W, -N \leq i \leq N\} > 0$$

**Theorem 5** *Assume transience for the countable case. Assume also that in the productions $w \to \alpha$ the strings $\alpha$ cannot have a substring $ww$.*

*Then the invariant measure is unique in $\mathcal{M}$, translation invariant (we defined it earlier for empirical correlation functions) and has exponential decay of correlations.*

*Convergence, starting from some measure in $\mathcal{M}$, of the correlation functions to those of the invariant measure is exponentially fast.*

**Remark 2** *Assume that the countable case is ergodic. Then it is easy to see that in nondegenerate cases there is an infinite number of extremal invariant measures. One should just take at time* 0 *special translation invariant distributions with sufficiently many final symbols on the initial strings with fixed point.*

*If the countable case is null-recurrent. then the invariant measure (empirical) in* $\mathcal{M}$ *also exists and seems to be unique in nondegenerate cases.*

*Proof of the theorem.*

Note first that symbols of the initial string which are not in $W$ do not count because, by transience of finite string dynamics starting with one symbol $w$, their density decreases in time. So, we can assume that the initial string is $x_i \equiv w$.

It is useful to start with particular cases where we can find explicitely the invariant measure.

**Right linear case.** Here only productions $w \to \alpha w$ are possible.

The invariant measure can be constructed as follows. Take normalized probabilities

$$p(\alpha) = \frac{q(w \to \alpha w)}{\sum_\alpha q(w \to \alpha w)}$$

Then replace in the initial string $x_i \equiv w$ each symbol (independently of all others) by $\alpha \neq e$ with probability $p(\alpha)$. The resulting random infinite string has the distribution coinciding with that of invariant measure for right linear random grammar with rates $q(.)$. Note that the number of $w$-symbols is conserved but the invariant correlation functions are zero for any string containing at least one symbol $w$.

The proof of this statement is obtained by a simple remark that each symbol of the initial string gives rise to a $d$-Markov chain and these chains are independent of each other. Correlation functions of this chain are equal to those of the invariant measure and to the correlation functions of the random string we obtained by one step substitution.

**Linear case.** Here only productions $w \to w\alpha w$ are possible. Also we can get the invariant measure using one step replacement of each $w$ independently by $\alpha w$ with probabilities

$$p(\alpha) = \frac{q(w \to w\alpha w)}{\sum_\alpha q(w \to w\alpha w)}$$

Note that substring $ww$ become rarer as $t \to \infty$ as we assumed that there is no production $w \to ww$.

**All symbols are nonterminal.** This case is a deviation from the proof of the main theorem but it demonstrates sufficiently well what can occur in more general cases.

Let us consider first the case where the only productions are of the type $s \to xy, x, y, s \in S$ and assume that all of them are positive. Define probabilities by normalization

$$p(s, xy) = \frac{q(s, xy)}{\sum q(s, .)}$$

Let us introduce two stochastic $(\mid S \mid \times \mid S \mid)$-matrices, left and right:

$$Q_l(s_1, s_2) = \sum_{u \in S} p(s_1, s_2 u)$$

$$Q_r(s_1, s_2) = \sum_{u \in S} p(s_1, u s_2)$$

**Lemma 6** *Invariant one point correlation function $\pi(s)$ is equal to the stationary distribution of the finite Markov chain with $\mid S \mid$ states and transition matrix*

$$Q = \frac{1}{2}(Q_l + Q_r)$$

*if this Markov chain is irreducible aperiodic.*

*Proof.* Each symbol $s$ of the initial string produces a binary planar tree of descendants. There are $2^n$ symbols on the level $n$ of this tree (discrete time $n$) and a unique path to each of these symbols. Summation of the one symbol probabilities at each vertex of the tree at level $n$ with weight $2^{-n}$ (empirical correlation functions) is equivalent to consider binomial expansion of $Q^n$. Taking $n \to \infty$ we have the result.

We shall use algebraic formalism for calculation of other correlation functions. This will prove also uniqueness of limiting correlation functions. Let $F$ be an algebra of all real functions on $S$, $\delta$-functions $\delta_s(.)$ form its basis. Define a comultiplication $\Delta F \to F \otimes F$ by

$$(\Delta f)(x, y) = \sum_s f(s) p(s, xy), f \in F$$

or

$$\Delta \delta_s = \sum_s p(s, xy) \delta_x \otimes \delta_y$$

**Remark 3** *This comultiplication is not coassociative in general.*

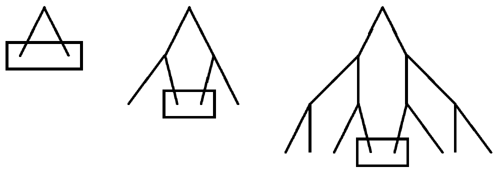

Figure 1: Two point correlations

**Proposition 3** *If $\pi(s)$ is the one-particle correlation function then two particle nearest neighbours correlation function $\pi_2$ is given by*

$$\pi_2 = \frac{1}{2}(\Delta\pi)\frac{1}{2}Q_l \otimes Q_r(1 - \frac{1}{2}Q_l \otimes Q_r)^{-1}$$

To prove this one should just note we sum contributions of different neighbours of the level $n$ of the tree and then put $n \to \infty$. Case $n = 3$ is shown on the following Figure.

We have

$$\pi_2 = \frac{1}{2}\Delta\pi + \frac{1}{4}(\Delta\pi)Q_l \otimes Q_r + ... = \sum_{k=1}^{\infty} \frac{1}{2^k}(\Delta\pi)Q_l^{k-1} \otimes Q_r^{k-1}$$

If the matrix $Q$ is aperiodic but reducible then the random grammar can have several invariant measures. We shall see it in the next paragraph.

**General case.** If $q(w \to e) = 0$ then one can use similar arguments with trees as in the previous paragraph. If there are, for example, productions $w \to \alpha_1 w \alpha_2 w \alpha_3$ then we consider trees with 5 lines emanating from each vertex. Three of them correspond to terminal symbols and two other - to nonterminal. The same arguments obviously work to prove the existence of the limit $n \to \infty$.

### 4.2 $L$-systems.

In two following subsections we deal mostly with discrete time case (OL-systems). It has some notational advantages. Instead of rates we have here probabilities $p(i \to \alpha)$ to substitute string $\alpha$ instead of symbol $i \in S$. We give here a different approach to context free case based on branching processes. All definitions which we have introduced for random grammars can be easily reformulated for random L-systems as well.

Let us consider a finite directed graph, we call it one-particle graph. Its set of vertices is $S$, there is a line $s \to x$ iff there is a production $s \to \alpha$ with $\alpha$ containing symbol $x$. As in finite Markov chains, we introduce (maximal) closed classes - subsets $S' \subset S$ such that for all $x, y \in S'$ there is a directed path in the one-particle graph from $x$ to $y$ and back. We will write for closed classes $S_i$ that $S_2 < S_1$ if there is a production $x \to \alpha$ with $x \in S_1$ and $\alpha$ having at least one symbol from $S_2$. Equivalently one can say that we define the directed graph of closed classes drawing a line from $S_1$ to $S_2$. There are no cycles in this graph. Thus the set of closed classes is partially ordered. By transitivity we shall say that $S_1 < S_3$ if there exists $S_2$ such that $S_1 < S_2$ and $S_2 < S_3$.

Let $M$ be the matrix $m_{ij}, i, j \in W$, of mean production rates where

$$m_{ij} = \sum_{\alpha} N_j(\alpha) p(i \to \alpha)$$

where $N_j(\alpha)$ is the number of symbols $j$ in $\alpha$.

For each closed class $S_a$ there is its own matrix $M(a) = M(a, a)$ of production rates. All of them we assume to be positively regular and thus nonperiodic. We use here the terminology of [13] and numerate the closed classes so that the matrix $M$ were lower triangular composed of matrices $M(a, b), a \geq b$ with the elements $m_{ij}, i \in S_a, j \in S_b$.

Let $\rho_a$ be the maximal positive eigenvalue of $M(a)$. Denote $\rho_{max} = \max \rho_a$.

Let $n$ be the number of closed classes $S_a$ such that $\rho_a > 1$ and there are no closed classes $S_b$ such that $S_b < S_a$ and $\rho_b > \rho_a$.

**Theorem 6** *There are at least $n$ extreme invariant measures for the OL-system.*

*Proof.* We can calculate one-point correlation function as the limiting density of particles of a particular type in the corresponding branching process. Take such $S_a$ and consider the initial string with only symbols in the class $S_a$. Then with positive probability symbols of $S_a$ will not die out. But one can claim even more: the number

$N_t(S_i)$ of symbols from classes $S_i$ will be $O(N_t(S_a))$ if $S_i < S_a$. This follows from results of [12], [13].

Thus the limiting one point correlation function will be nonzero for this class but zero for classes above $S_a$. By compactness we get at least $n$ different limiting distributions.

## 4.3 Fractal Correlation Functions

Invariant measures do not provide sufficient characterization of Random Grammars or L-systems. One of the reasons for this is that there are strong connections between Grammars and L-systems on one side and fractals on another side (see [5]). We shall see it immediately from the example below.

Let us remind these connections (see [5]) starting with a random context free grammar, which will bring us to the famous Cantor set. Here $S = \{0, 1\}$ and the only productions are

$$0 \to 000, 1 \to 101$$

The rates, corresponding to the productions are both equal to 1. If we take DOL-system with the same productions we get exactly the recurrent procedure to obtain the Cantor set, After time $t = 0, 1, ...,$ we assign to each symbol $x_1, ..., x_n, n = 3^t$, consecutive subintervals $[(k-1)3^{-t}, k3^{-t}]$ correspondingly of black (in case $x_k = 1$) and white (in case $x_k = 0$) colour. Let $C_t$ be the union of black intervals. Then the Cantor set is $\bigcap_t C_t$. Random grammar, which we introduced, only makes recurrent procedure of Cantor set nonparallel.

In this example the limiting (= invariant) measure is atomic concentrated on the configuration $x_i \equiv 0$. Using geometric language it is equivalent to say that Cantor set has Lebesgue measure zero. If we want to know the asymptotics of one particle correlation function $p_t(1)$ it is equivalent to the question about fractal dimension of the Cantor set. Thus, to study the asymptotics of correlation functions we have to introduce "fractal language" for random grammars and L-systems.

**Lemma 7** *Consider a transient OL-system. Assume some power of the matrix of mean production rates to be positive and let* $\rho$ *be maximal eigenvalue of* $M$, $\vec{v}$ *- corresponding positive left eigenvector. Start from finite string not identically zero and denote* $N_t$ *the vector of numbers of symbols from* $S$ *in the string after time* $t$. *Then* $\frac{N_t}{\rho^t}$ *tends in distribution to* $\xi v$ *where* $\xi$ *is a random variable on* $R_+$, *which is positive with positive probability.*

This is known from the theory of branching processes with several particle types, see [7], [8].

Thus after time $t$ we shall have approximately $c\rho^t$ symbols in the string. Consider now some string $\gamma$ and let $n_t(\gamma) = n_t(\gamma;\alpha)$ be (random) number of substrings $\gamma$ in the string at time $t$, if the initial string is $\alpha$.

We shall call random L-system weakly degenerate if all matrices $M_a$ are positive regular and their maximal positive eigenvalues are different from 1 and different from each other.

**Theorem 7** *Assume OL-system to be weakly degenerate. Then for each $\gamma$ and each initial finite string $\alpha$, under the condition that the length of the string goes to infinity when $t \to \infty$, a.s. either there exists $1 < a = a(\gamma,\alpha) \leq m$ such that*

$$\frac{\log n_t(\gamma)}{\log N_t} \to h(\gamma) \doteq \frac{\log a}{\log \rho}$$

*as $t \to \infty$ or $n_t(\gamma)$ grows slower then any exponent. In the latter case we put $h(\gamma) = 0$. Otherwise speaking, $n_t(\gamma) \approx m^{th(\gamma)}$. The limit is called fractal exponent (or critical exponent) $h(\gamma)$ of $\gamma$ (or of the corresponding correlation function).*

*If $h(\gamma) = 1$ we call $\gamma$ (or the corresponding correlation function) normal type, if $0 < h(\gamma) < 1$ we call it of fractal type. If it is zero we call it of zero type. If at least one string is of fractal type we say that there is fractal behaviour in the random gramma or L-system.*

We again start with particular cases.

**Linear case.** In the right linear case there is no exponential growth but it is of some interest as a model example. In the linear case all normal type strings are substrings of the following strings

$$\omega\alpha_1\omega\alpha_2...\omega\alpha_n$$

All other have zero type. There is no fractal type strings.

**Cantor grammar.**

**Proposition 4** *In this case the only normal type strings are $\gamma = 00...0$. The only fractal exponent is $h = \frac{\log 2}{\log 3}$. Possible fractal strings are all finite substrings, except $00...0$, which can appear in subsequent substitutions. All other strings cannot appear at all.*

*Proof.* We use a method which we call killing the invariant measure. We consider a modified Cantor system where $S$ is potentially infinite: $S = \{0, 1, 3, 9, ..., 3^k, ...\}$. Substitutions are

$$0 \to 3, 3^k \to 3^{k+1}, 1 \to 101$$

and also have rate 1. Otherwise speaking, we encode long substrings of zeros. After this rescaling all possible (which appear in the process) substrings have normal type. That is for the original system (after decoding $3^k$ to zeros) they have the same critical exponent.

**One point correlations.** One can consider the system of all one point correlation functions as particles of a branching process. Assume that the branching process is nondegenerate and positive regular (some power of the matrix of mean productions is positive). Then all one-particle correlation functions are of normal type. This follows from well known results in the theory of branching processes (see [7], [8], [9]).

For degenerate cases theorem 7 follows from the results of [12], [13].

**Two point correlations.**

**Proposition 5** *Assume that only productions of the type $s \to xy$ are possible and all the matrix $Q$ of Lemma 6. Then there is no fractal behaviour. All normal strings are exactly those which can appear with positive probability during the process.*

This proposition was in fact proved above.

So, to get fractal behaviour we should consider cases with degenerate one-point correlations.

**Decomposable cases.**

**Theorem 8** *Assume that there are only two closed classes $S_1 < S_2$. Then*

1. *If $\rho_{S_1} > \rho_{S_2} > 1$ then there is only one invariant measure and not more than one critical exponent, the same for each initial string containing at least one $S_2$-symbol;*

2. *If $\rho_{S_1} > 1 > \rho_{S_2}$ then there is only one invariant measure and no fractal behaviour;*

3. *If $1 < \rho_{S_1} < \rho_{S_2}$ then there are two extreme invariant measures and one critical exponent;*

*4. If $\rho_{S_1} < 1 < \rho_{S_2}$ then there is one invariant measure and no fractal behaviour.*

We now can propose the following conjecture for any context free grammar.

**Conjecture 2** *The set of extremal invariant measures and the set of critical exponents are finite.*

For each class $S_a$ consider a new OL-system $\mathcal{L}(S_a)$. Its productions and probabilities $p_{S_a}(.) = p(.)$ are

$$p_{S'}(s \to \beta) = \sum p(s \to \alpha)$$

where the sum is over all $\alpha$ such that after deleting all symbols from other classes we get $\beta$.

Consider case 3 and start with initial infinite string containing only symbols from $S_2$. Then symbols from $S_2$ will dominate and the the fractal exponent for one-particle correlation function of symbols from $S_1$ will be $\frac{\log \rho_{S_1}}{\log \rho_{S_2}}$.

Other cases can be treated similarly.

## 4.4 Measures on Languages

**Sentences** Consider a context free grammar with $\mid W \mid = 1$. Let $L = L(P, \alpha)$ be the set of sentences in the given language which can be obtained by substitution process starting from finite string $\alpha$. Let $\mu$ be the hitting distribution on $L$, i.e. $\mu(\beta)$ equals the probability of hitting $\beta$ starting from $\alpha$. Let us take all sentences of length $N$ and enumerate each sentence by the integer points of the intervals $[-\frac{N-1}{2}, \frac{N-1}{2}]$ for $N$ odd and $[-\frac{N}{2}+1, \frac{N}{2}]$ for $N$ even. Consider conditional distribution $\mu^N$

$$\mu^N(\beta) = \frac{\mu(\beta)}{\sum_{\beta: l(\beta)=N} \mu(\beta)}$$

We want to consider thermodynamic limit of this random field $\xi_i^N$.

To give a more explicit characterization of this limiting field consider the distribution $\nu_w$ on the set of sentences obtained when starting from the symbol $w \in W$.

**Theorem 9** *The random field $\xi_i^N$ weakly converges, as $N$ tends to infinity, to a translation invariant random field on $Z$ with values in $V$.*

*This limiting field can be obtained also in the following way: consider the translation invariant limiting measure for the process with $W$-symbols and substitute instead each $w$-symbol independently a sentence randomly chosen from distribution $\nu$.*

At the end we want to add some remarks. It is quite obvious that each random grammar with fractal exponents should have some geometrical fractal interpretation (many examples and references see in [5]). Geometrical incarnations of random grammars (like Cantor) are tightly connected with exit boundaries. In another place we shall come to this point with more details.

In [1] there is a review of recent results about more general right-linear random grammars - we do not assume them to be context free. Linear case corresponds to two-sided evolution of finite string (see also [1]) . Context free linear case is quite trivial compared to non context free case.

# References

[1] V.A.Malyshev. Interacting Chains of Characters. INRIA report, No. 3057, 1996. Appeared in Uspehi Mat. nauk, 1997, No. 2, pp.59-86.

[2] V.A.Malyshev, R.A. Minlos. Gibbs Random Fields. Kluwer. 1990.

[3] P. Eichhorst., W. Savitch. Growth Functions of Stochastic Lindenmayer Systems. Information and Control, 1980, v. 45, 217-228.

[4] T. Yokomori. Stochastic Characterizations of EOL Languages. Information and Control, 1980, v. 45, 26-33.

[5] P. Prusinkiewicz, J. Hanan. Lindenmayer Systems, Fractals, and Plants. Lecture Notes in Biomathematic, v. 79, Springer. 1989.

[6] Th. Liggett. Interacting Particle Systems. Springer. 1985.

[7] Th. Harris. The theory of branching processes.Springer. 1963.

[8] K.B. Athreya, P.E. Ney. Branching Processes. Springer. 1972.

[9] B. A. Sevastianov. Branching processes. Moscow. Nauka. 1971.

[10] S. Madjid. Foundations of Quantum Group Theory. Cambridge Univ. Press. 1995.

[11] G. Fayolle, V. Malyshev, M. Menshikov. Constructive methods in countable Markov Chains. Cambridge Univ. Press. 1995.

[12] H. Kesten, B. P. Stigum. Limit Theorems for Decomposable Multi-Dimensional Galton-Watson Processes.. J. Math. Anal. Appl., v. 17 (1967), 309-338.

[13] Ch. Mode. Multitype Branching Processes. Theory and Applications. American Elsevier. NY. 1971.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399