# Vershik's intermediate level standardness criterion and the scale of an automorphism

Stéphane Laurent *

**Abstract**

Vershik's standardness criterion takes a particular form of combinatorial nature in the case of $r_n$-adic filtrations, which we call Vershik's intermediate level criterion in this paper. This criterion has been intensively used in the ergodic-theoretic literature, but it is not easily applicable by probabilists because it is stated in a language proper to the theory of measurable partitions and has not been translated in probabilistic terms. We aim to provide an easily applicable probabilistic statement of this criterion. Finally, Vershik's intermediate level criterion is illustrated by revisiting Vershik's definition of the scale of an invertible measure-preserving transformation.

# Contents



*Université de Liège

# 1 Introduction

Although there has been many efforts ([4], [5], [6], [7]) oriented towards the translation of Vershik's theory on decreasing sequence of measurable partitions in a theory of filtrations written in the language of stochastic processes, many papers in the ergodic theory literature dealing with standard filtrations still remain difficult to read for probabilists outside the class of experts in this topic. Difficulties do not lie in basic concepts of ergodic theory such as the ones presented in introductory books on measure-preserving systems, but rather in the language of the theory of measurable partitions initiated by Rokhlin (see [15]). Rokhlin's correspondance (see [3]) between measurable partitions and complete $\sigma$- fields is not a complicated thing, but the approach to filtrations is somewhat geometrical in the language of partitions, whereas probabilists are more comfortable with considering a filtration as the history of a stochastic process whose dynamic is clearly described.

Particularly, many ergodic theory papers (such as [HH], [Ho], [HR]) dealing with standardness are concerned by *$r_n$-adic filtrations*: those filtrations $\mathcal{F} = (\mathcal{F}_n)_{n\leqslant 0}$ for which $\mathcal{F}_n = \mathcal{F}_{n-1} \vee \sigma(\varepsilon_n)$ for each $n \leqslant 0$, where the *innovation* $\varepsilon_n$ is a random variable independent of $\mathcal{F}_{n-1}$ and uniformly distributed on $r_n$ possible values, for some sequence $(r_n)_{n\leqslant 0}$ of positive integers. For such filtrations, Vershik's general standardness criterion, which has received some attention in the probability theory literature ([4], [7]), and which we call *Vershik's second level (standardness) criterion* for more clarity, takes a particular form, which is in fact the original form of Vershik's standardness criterion who focused on $r_n$-adic filtrations, and which we call *Vershik's intermediate level (standardness) criterion*. Its statement involves tree automorphisms and characterizes standardness in terms of a problem of combinatorial nature. Ergodicians directly apply Vershik's intermediate level criterion in their works, but it has not been translated in the probability theory literature (an attempt has been done in the PhD. thesis [5] of the author), thereby causing difficulties for the probabilist reader. The present paper provides a probabilistic statement of Vershik's intermediate level criterion. This statement is not as brief as Vershik's analogous statement in the language of measurable partitions, but it is directly applicable to investigate standardness of $r_n$-adic filtrations without making call to notions unfamiliar to probabilists, except eventually the notion of tree automorphisms. Roughly speaking, tree automorphisms lie in the heart of Vershik's intermediate level criterion because all possible local innovations $(\varepsilon_n, \ldots, \varepsilon_0)$ of a $r_n$-adic filtration $\mathcal{F}$ differ by the action of a $\mathcal{F}_{n-1}$-measurable random tree automorphism.

We shall show that Vershik's intermediate level criterion is equivalent to Vershik's second level criterion in the context of $r_n$-adic filtrations. All results in the present paper are self-contained except for the proofs of rather elementary statements for which we will refer to [7]. Section 2 aims to provide the non-specialist reader with some motivations for Vershik's two critera by recalling their relations with the notions of productness and standardness. In section 3 we state Vershik's second level criterion, similarly to [4] and [7], and its elementary properties. Vershik's intermediate level criterion is then the purpose of section 4. In section 5, we shall illustrate Vershik's intermediate level criterion by formulating Vershik's definition of the *scale of an automorphism* ([13]) in terms of this criterion. Vershik did not consider this approach of the scale of an automorphism, but rather focused on a definition that visibly concerns the orbits of this automorphism. The equivalence of the two definitions was announced by the author in [7], without proof, and then [7] shows how to derive the scale of Bernoulli automorphisms from the theorem on productness of the split-word process presented in [6]. With our definition, many properties of the scale of an automorphism stated by Vershik in [13] appears as direct applications of elementary properties of Vershik's intermediate level criterion or more general results of the theory of filtrations. We will see that the scale of a completely ergodic automorphism is nonempty as a consequence of Vershik's theorem on lacunary isomorphism, whereas Vershik proved this proposition by a direct construction.

# 2 Standardness and productness

We briefly present the meaningful notions of productness and standardness and their relations with Vershik's intermediate level criterion and Vershik's second level criterion. These notions manifestly motivate the development of these criteria.

The pioneering works of Vershik mainly deal with $r_n$-adic filtrations. For a given sequence $(r_n)_{n\leqslant 0}$ of integers $r_n \geqslant 2$, a filtration $\mathcal{F}$ is said to be $r_n$*-adic* if $\mathcal{F}_n = \mathcal{F}_{n-1} \vee \sigma(\varepsilon_n)$ for every $n \leqslant 0$ where $\varepsilon_n$ is a random variable independent of $\mathcal{F}_{n-1}$ and uniformly distributed on a finite set consisting of $r_n$ elements. Such random variables $\varepsilon_n$ are called *innovations* of $\mathcal{F}$. The process $(\varepsilon_n)_{n\leqslant 0}$ is then a sequence of independent random variables and it is itself called an innovation of $\mathcal{F}$, and we also say that $(\varepsilon_n, \ldots, \varepsilon_0)$ is a *local innovation* of $\mathcal{F}$. In other words, the innovation $\varepsilon_n$ is a random variable generating an independent complement of $\mathcal{F}_{n-1}$ in $\mathcal{F}_n$, by saying that a $\sigma$- field $\mathcal{C}$ is an *independent complement* of a $\sigma$- field $\mathcal{B}$ in a $\sigma$- field $\mathcal{A} \supset \mathcal{B}$ if it is independent of $\mathcal{B}$ and $\mathcal{A} = \mathcal{B} \vee \mathcal{C}$. Independent complements are not unique in general, as testified by the following lemma whose proof is left as an easy exercise.

**Lemma 2.1.** *Let $(\mathcal{B}, \mathcal{A})$ be an increasing pair of $\sigma$- fields and $V$ be a random variable generating an independent complement of $\mathcal{B}$ in $\mathcal{A}$. If $V$ is uniformly distributed on a finite set $F$, then $\Phi(V)$ also generates an independent complement of $\mathcal{B}$ in $\mathcal{A}$ for all $\mathcal{B}$-measurable random permutations $\Phi$ of $F$.*

Actually one can prove that any independent complement of $\mathcal{B}$ in $\mathcal{A}$ is generated by $\Phi(V)$ for some $\mathcal{B}$-measurable random permutation $\Phi$, but we shall not need this result.

One of the main achievements of Vershik in his pioneering works is a criterion which characterizes *productness* of $r_n$-adic filtrations. A filtration is said to be of *product type* if it is the filtration generated by a sequence of independent random variables. In particular, an $r_n$-adic filtration is of product type if it is the $r_n$-adic filtration generated by a uniformly $r_n$-ary independent random drawings, that is, a sequence $(\varepsilon_n)$ of independent random variables with $\varepsilon_n$ uniformly distributed on $r_n$ distinct values for each $n$. With our terminology, Vershik's above mentioned result is stated in the following theorem. Filtrations $\mathcal{F} = (\mathcal{F}_n)_{n \leqslant 0}$ whose final $\sigma$- fields $\mathcal{F}_0$ are essentially separable are simply called *essentially separable*.

**Theorem 2.2.** *An essentially separable $r_n$-adic filtration satisfies Vershik's intermediate level standardness criterion if and only if it is of product type.*

The statement of Vershik's intermediate (standardness) level criterion does not make sense for arbitrary filtrations but only $r_n$-adic ones. But Vershik also provided in his pionneering works an equivalent statement of this criterion which is applicable to arbitrary filtrations. In the present paper, according to [7], we call it *Vershik's second level (standardness) criterion*, or, shortly, the *second Vershik property*, and we also say that a filtration is *Vershikian* when it satisfies this property. In [14], Vershik states the following theorem.

**Theorem 2.3.** *An essentially separable filtration satisfies Vershik's second level standardness criterion if and only if it is standard.*

*Standardness* for an arbitrary filtration is defined with the help of the notion of immersion. A filtration $\mathcal{F}$ is said to be *immersed* in a filtration $\mathcal{G}$

if every $\mathcal{F}$-martingale is a $\mathcal{G}$-martingale (that implies $\mathcal{F} \subset \mathcal{G}$). We refer to [4] and [6] for more details on the immersion property. Then a filtration $\mathcal{F}$ is said to be *standard* if, up to isomorphism, it is immersed in the filtration generated by some sequence of independent random variables each having a diffuse law, or, equivalently (see [6]), in the filtration generated by any sequence of independent random variables.

# 3 Vershik's second level standardness criterion

## 3.1 The Kantorovich metric and Vershik's progressive predictions

The *Kantorovich* distance plays a major role in the statement of the second level Vershik property. Given a separable metric space $(E, \rho)$, the Kantorovich distance $\rho'$ on the set $E'$ of probabilities on $E$ is defined by

$$\rho'(\mu, \nu) = \inf_{\Lambda \in \mathcal{J}(\mu,\nu)} \iint \rho(x, y) \, \mathrm{d}\Lambda(x, y),$$

where $\mathcal{J}(\mu, \nu)$ is the set of joinings of $\mu$ and $\nu$, that is, the set of probabilities on $E \times E$ whose first and second marginal measures are $\mu$ and $\nu$ respectively.

In general, the topology induced by $\rho'$ on the set $E'$ of probability on $E$ is finer than the topology of weak convergence. These two topologies coincide when $(E, \rho)$ is compact, in particular $(E', \rho')$ is compact in this case. The metric space $(E', \rho')$ is complete and separable whenever $(E, \rho)$ is (see e.g. [1]).

The following lemma will be used to prove the equivalence between Vershik's intermediate and second level properties.

**Lemma 3.1.** *Let $r \geqslant 2$ be an integer and let $f$ and $g$ be functions from $\{1, \ldots, r\}$ to a Polish metric space $(E, \rho)$. Denote by $\nu$ the uniform probability on $\{1, \ldots, r\}$. Then the infimum in the Kantorovich distance $\rho'\big(f(\nu), g(\nu)\big)$ is attained for the joint law of a random pair $\big(f(\varepsilon), g(\varepsilon')\big)$ where $\varepsilon$ is a random variable distributed according to $\nu$ and $\varepsilon' = \sigma(\varepsilon)$ for some permutation $\sigma$ of $\{1, \ldots, r\}$.*

*Proof.* Any joining of $f(\nu)$ and $g(\nu)$ is the law of a random pair $\big(f(\varepsilon), g(\varepsilon')\big)$ where $\varepsilon \sim \nu$ and $\varepsilon' \sim \nu$, and the expectation $\mathbb{E}\Big[\rho\Big(f(\varepsilon), g(\varepsilon')\Big)\Big]$ is a linear form of the joint law of $\varepsilon$ and $\varepsilon'$. Therefore, there exists at least an extremal point in the set of all joinings where the minimal possible value of this expectation is attained, and a joining is an extremal point when $\varepsilon' = \sigma(\varepsilon)$ for some permutation $\sigma$ (Birkhoff's theorem). $\square$

Let $\mathcal{F}$ be a filtration, $E$ a Polish metric space and $X \in L^1(\mathcal{F}_0; E)$. The Vershik's second level property of $X$ involves *Vershik's progressive predictions* $\pi_n X$ of $X$, which correspond to the so-called *universal projectors* in [10] and [14]. They are recursively defined as follows: we put $\pi_0 X = X$, and $\pi_{n-1} X = \mathcal{L}[\pi_n X \,|\, \mathcal{F}_{n-1}]$ (the conditional law of $\pi_n X$ given $\mathcal{F}_{n-1}$) ; thus, the $n$-th progressive prediction $\pi_n X$ of $X$ with respect to $\mathcal{F}$ is a random variable taking its values in the Polish space $E^{(n)}$, which is recursively defined by $E^{(0)} = E$ and $E^{(n-1)} = \left(E^{(n)}\right)'$, denoting as before by $E'$ the space of probability measures on any separable metric space $E$. The state space $E^{(n)}$ of $\pi_n X$ is Polish when endowed with the distance $\rho_n$ obtained by iterating $|n|$ times the construction of the Kantorovich distance starting with $\rho$: we recursively define $\rho_n$ by putting $\rho_0 = \rho$ and by defining $\rho_{n-1} = (\rho_n)'$ as the Kantorovich distance issued from $\rho_n$.

Finally, in order to state Vershik's second level criterion, we introduce the *dispersion* $\operatorname{disp} X$ of (the law of) an integrable random variable $X$ in a Polish metric space. It is defined as the expectation of $\rho(X', X'')$ where $X'$ and $X''$ are two independent copies of $X$, that is, two independent random variables having the same law as $X$.

## 3.2 Vershik's second level criterion

Let $\mathcal{F}$ be a filtration, let $E$ be a Polish metric space and $X \in L^1(\mathcal{F}_0; E)$. We say that the random variable $X$ satisfies *Vershik's second level (standardness) criterion*, or the *second Vershik property*, or, for short, that $X$ is *Vershikian* (with respect to $\mathcal{F}$) if $\operatorname{disp} \pi_n X \longrightarrow 0$ as $n$ goes to $-\infty$. Then we extend this definition to $\sigma$-fields $\mathcal{E}_0 \subset \mathcal{F}_0$ and to the whole filtration as follows: we say that a $\sigma$-field $\mathcal{E}_0 \subset \mathcal{F}_0$ is *Vershikian* if each random variable $X \in L^1(\mathcal{E}_0; [0,1])$ is Vershikian, and we say that that the filtration $\mathcal{F}$ is *Vershikian* if the final $\sigma$-field $\mathcal{F}_0$ is Vershikian.

The following proposition is proved in [7] when $(E, \rho)$ is a compact metric space, but it is easy to check that the proof remains valid for a Polish metric space.

**Proposition 3.2.** *For any Polish space $(E, \rho)$, a random variable $X \in L^1(\mathcal{F}_0, E)$ is Vershikian if and only if the $\sigma$-field $\sigma(X)$ is Vershikian.*

Below we state Vershik's theorem on lacunary isomorphism which will be applied in section 5 to prove the nonemptiness of the scale of a completely ergodic automorphism. The first version of this theorem was stated and proved by Vershik in the context of $r_n$-adic filtrations, but was proved without using any standardness criterion: Vershik showed that from every Kolmogorovian $r_n$-adic filtration, it is possible to extract a filtration of product type. The analogous proof for conditionally non-atomic filtrations (that is, filtrations admitting innovations with diffuse law) is provided by Émery and Schachermayer [4]. This proof is somewhat constructive but quite technical, whereas a short proof of the general version of the theorem on lacunary isomorphism stated below, based on Vershik's second level criterion, is given in [7]. A filtration is said to be *Kolmogorovian* if $\mathcal{F}_{-\infty} := \cap_{n \leqslant 0} \mathcal{F}_n$ is the degenerate $\sigma$-field.

**Theorem 3.3.** *Let $\mathcal{F} = (\mathcal{F}_n)_{n \leqslant 0}$ be an essentially separable filtration. If $\mathcal{F}$ is Kolmogorovian, there exists a strictly increasing map $\phi : -\mathbb{N} \to -\mathbb{N}$ such that the extracted filtration $(\mathcal{F}_{\phi(n)})_{n \leqslant 0}$ is Vershikian.*

It is not straightforward to see from the definition of the Vershik property that every filtration extracted from a Vershikian filtration is itself Vershikian, whereas this is very easy to see from the definition of the I-cosiness criterion which is known to be equivalent to the Vershik property (see [4], [7]). It is also easy to see that this property holds for Vershik's intermediate level criterion, but this criterion only concerns $r_n$-adic filtration. Note that it is always possible to take $\phi(0) = 0$ in the theorem above since the Vershik property is an asymptotic one (see [7]).

We shall also apply the following lemma in section 5. It is proven in [7].

**Lemma 3.4.** *For any Polish metric space $E$, a random variable $X \in L^1(\mathcal{F}_0; E)$ is Vershikian if and only if the filtration generated by the stochastic process $(\pi_n X)_{n \leqslant 0}$ is Vershikian.*

The next lemma says that the second level Vershik property is hereditary for immersion; we refer to [7] for its proof.

**Lemma 3.5.** *Let $\mathcal{F}$ be a filtration, $\mathcal{E}$ a filtration immersed in $\mathcal{F}$, and $E$ a Polish metric space. A random variable $X \in L^1(\mathcal{E}_0; E)$ is Vershikian with respect to $\mathcal{F}$ if and only it is Vershikian with respect to $\mathcal{E}$. Consequently, if the filtration $\mathcal{F}$ is Vershikian, then so is also $\mathcal{E}$.*

# 4 Vershik's intermediate level criterion

Vershik's intermediate level criterion is the object of section 4.3. In section 4.1 we mainly fix some notations about tree automorphisms which will be needed to prove the equivalence between Vershik's two criteria, and in section 4.2 we introduce the split-word processes which will be needed to state Vershik's intermediate level criterion.

Throughout this section we shall speak of *words* on a set $A$ called the *alphabet*. A *word* $w$ on $A$ is an element of $A^\ell$, or equivalently an application from $\{1, \ldots, \ell\}$ to $A$, for some integer $\ell \geqslant 1$ called the *length* of $w$. A word of length $\ell$ is shortly termed as an $\ell$*-word*. The *letters* of an $\ell$-word $w$ are $w(1)$, …, $w(\ell)$. When $A$ is treated as a Polish space it is understood that the set $A^\ell$ of $\ell$-words on $A$ is treated as the corresponding product Polish space.

## 4.1 Tree automorphisms

All notions defined below are relative to a given sequence $(r_n)_{n\leqslant 0}$ consisting of integers $r_n \geqslant 2$, from which we define the sequence $(\ell_n)_{n\leqslant 0}$ by $\ell_n = \prod_{k=n+1}^{0} r_k$ for all $n \leqslant 0$.

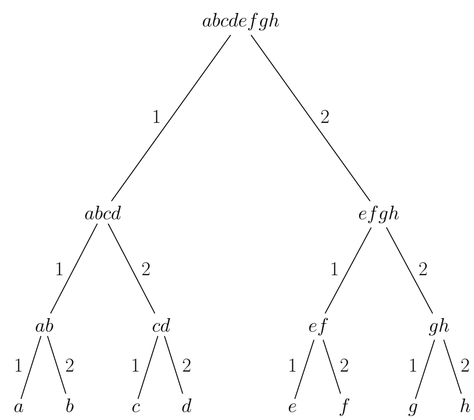


Figure 1: A labeled tree.

Define the sets $B_n = \prod_{k=n+1}^{0}\{1, \ldots, r_k\}$ for $n \leqslant -1$. The group $G_n$ of *tree automorphisms* of $B_n$ is a subgroup of the group of permutations of $B_n$ recursively defined as follows. The group $G_{-1}$ is the whole group of permutations of $\{1, \ldots, r_0\}$, and a permutation $\tau \in G_n$ maps an element $b_n = (c_n, b_{n+1}) \in B_n = \{1, \ldots, r_{n+1}\} \times B_{n+1}$ to $\tau(b_n) = \big(\sigma(c_n), \psi(\sigma(c_n))(b_{n+1})\big)$

where $\sigma$ is a permutation of $\{1, \ldots, r_{n+1}\}$ and $\psi$ is a map from $\{1, \ldots, r_{n+1}\}$ to $G_{n+1}$.

**Lemma 4.1.** *Let $\mathcal{F}$ be an $r_n$-adic filtration and $(\varepsilon_n)_{n \leqslant 0}$ an innovation of $\mathcal{F}$. If $\tau$ is a random $\mathcal{F}_n$-measurable tree automorphism then $\tau(\varepsilon_{n+1}, \ldots, \varepsilon_0)$ is a local innovation of $\mathcal{F}$.*

*Proof.* This is easily proved by recursion with the help of lemma 2.1. □

For any set $A$, there is a natural action of $G_n$ on the set of $\ell_n$-words on $A$. First introduce the lexicographic order on $B_n$, which is made visual by drawing a tree as such in figure 1 and then by enumerating the branches of this tree from the left to the right; for this order, the position $p(b)$ of $b = (b_{n+1}, \ldots, b_0) \in B_n$ is given by $p(b) = 1 + \sum_{k=n+1}^{0}(b_k - 1)\ell_k$. Now introduce the following notation.

**Notation 4.2.** Given a word $w$ of length $\ell_n$, $n \leqslant -1$, and a branch $b \in B_n$, we denote by $t_n(w, b) = w(p(b))$ the letter of $w$ whose index is the position of $b$ for the lexicographic order on $B_n$.

In other words, the $i$-th letter of $w$ is $t_n(w, b)$ for the branch $b = p^{-1}(i)$. The application $t_n$ is clearly made visual on figure 1: the letter $t_n(w, b)$ of $w$ is the label at the leaf of the branch $b$. Then the action of $G_n$ on $A^{\ell_n}$ is defined as follows. For a tree automorphism $\tau \in G_n$ and a word $w \in A^{\ell_n}$ we define $\tau.w$ as the word satisfying $t_n(\tau.w, b) = t_n\big(w, \tau(b)\big)$ for every branch $b \in B_n$, that is, the $p(b)$-th letter of $\tau.w$ is the $p(\tau(b))$-letter of $w$.

Now we introduce additional notations that we will use later.

**Notation 4.3.** Given an underlying sequence $(r_n)_{n \leqslant 0}$ and defining the sequence $(\ell_n)_{n \leqslant 0}$ as above, then, for any $n \leqslant 0$ and any word $w$ of length $\ell_{n-1} = r_n \ell_n$ on an alphabet $A$, we denote by $\tilde{w}$ the word of length $r_n$ on the alphabet $A^{\ell_n}$ obtained from $w$, that is, the $j$-th letter of $\tilde{w}$ is $\tilde{w}(j) = w_{(j-1)\ell_n+1} \ldots w_{j\ell_n}$.

Then note that the action of a tree automorphism in $G_{n-1}$ on a word $w \in A^{\ell_{n-1}}$ consists of $r_n$ tree automorphisms in $G_n$ which respectively act on the subwords $\tilde{w}(1), \ldots, \tilde{w}(r_n)$ and of a permutation of these subwords. This yields relation (4.1) below.

**Notation 4.4.** For a fixed sequence $(r_n)_{n \leqslant 0}$ and given two words $w, w'$ of length $\ell_n$ on a Polish metric space $(A, \theta)$, we define the Hamming distance between $w$ and $w$' by

$$\delta_n^{\theta}(w, w') = \frac{1}{\ell_n} \sum_{i=1}^{\ell_n} \theta\big(w(i), w'(i)\big)$$

and the associated distance between the orbits of $w$ and $w'$ under the action of $G_n$ by

$$d_n^\theta(w, w') = \min_{\tau \in G_n} \delta_n^\theta(w, \tau.w').$$

When this causes no ambiguity we write $d_n$ and $\delta_n$ instead of $d_n^\theta$ and $\delta_n^\theta$.

It can be easily checked that, with notation 4.3, we have the recurrence relation

$$d_{n-1}^\theta(w, w') = \min_{\sigma \in \mathcal{S}_{r_n}} \frac{1}{r_n} \sum_{i=1}^{r_n} d_n^\theta\Big(\tilde{w}(i), \tilde{w}'(\sigma(i))\Big). \tag{4.1}$$

for any words $w, w'$ of length $\ell_{n-1}$ on the alphabet $A$ and where $\mathcal{S}_{r_n}$ denotes the group of permutations of $\{1, \ldots, r_n\}$.

## 4.2 Split-word processes

Throughout this section, we consider a Polish metric space $(A, \theta)$. The set $A$ is termed as *alphabet*.

Given a sequence $(r_n)_{n \leqslant 0}$ of integers $r_n \geqslant 2$, called the *splitting sequence*, we will soon define an *$r_n$-adic split-word process on $A$*. We firstly define the *length sequence* $(\ell_n)_{n \leqslant 0}$ by $\ell_n = \prod_{k=n+1}^0 r_k$ for all $n \leqslant 0$. Next, according to notation 4.3, every word $w$ of length $\ell_{n-1} = r_n \ell_n$ on the alphabet $A$ is naturally identified as a word $\tilde{w}$ of length $r_n$ on the alphabet $A^{\ell_n}$, and we define the *splitting map* $s_n \colon A^{\ell_{n-1}} \times \{1, 2, \ldots, r_n\} \longrightarrow A^{\ell_n}$ for each $n \leqslant 0$ by $s_n(w, j) = \tilde{w}(j)$. That is, to each word $w$ of length $\ell_{n-1}$ on $A$ and each integer $j \in [1, r_n]$, the splitting map $s_n$ assigns the $j$-th letter of $w$ treated as a $r_n$-word on $A^{\ell_n}$.

Then we say that, with respect to some filtration $\mathcal{F}$, a process $(W_n, \varepsilon_n)_{n \leqslant 0}$ is a *split-word process on the alphabet $A$ with splitting sequence* $(r_n)_{n \leqslant 0}$, or an *$r_n$-adic split-word process on $A$*, if for each $n \leqslant 0$,

- $W_n$ is a random word on $A$ of length $\ell_n$;
- $\varepsilon_n$ is a random variable uniformly distributed on $\{1, 2, \ldots, r_n\}$ and is independent of $\mathcal{F}_{n-1}$, and $W_n = s_n(W_{n-1}, \varepsilon_n) = \widetilde{W}_n(\varepsilon_n)$, that is, the word $W_n$ is the $\varepsilon_n$-th letter of of $W_{n-1}$ treated as an $r_n$-word on $A^{\ell_n}$;
- $W_n$ and $\varepsilon_n$ are $\mathcal{F}_n$-measurable.

An $r_n$-adic split-word process $(W_n, \varepsilon_n)_{n \leqslant 0}$ generates an $r_n$-adic filtration for which $(\varepsilon_n)_{n \leqslant 0}$ is an innovation. Note that one has $W_0 = t_n(W_n, \varepsilon_{n+1}, \ldots, \varepsilon_0)$

with notation 4.2, and note also that $(W_n, \varepsilon_n)_{n\leqslant 0}$ is Markovian with respect to the filtration $\mathcal{F}$ hence the filtration it generates is immersed in $\mathcal{F}$.

Given a sequence $(\gamma_n)_{n\leqslant 0}$ of probability measures $\gamma_n$ on $A^{\ell_n}$, the existence of such a process $(W_n, \varepsilon_n)_{n\leqslant 0}$ with $W_n \sim \gamma_n$ occurs whenever each $\gamma_n$ is the image under the splitting map $s_n$ of the independent product of $\gamma_{n-1}$ with the uniform probability on $\{1, 2, \ldots, r_n\}$. For example, $\gamma_n$ can be taken as the projection on $\ell_n$ consecutive coordinates of a stationary probability measure on $A^{\mathbb{Z}}$.

**Example 4.5.** [*The "ordinary" split-word processes*] The ordinary split-word process with splitting sequence $(r_n)_{n\leqslant 0}$ is the process $(W_n, \varepsilon_n)_{n\leqslant 0}$ defined above when the probability $\gamma_n$ on $A^{\ell_n}$ is the product probability of some probability measure $\mu$ on $A$. Standardness of the filtration generated by an ordinary split-word process is known to be characterized by a certain asymptotic condition on the splitting sequence (see [6]).

In the next two lemmas, we consider a Polish metric alphabet $(A, \theta)$.

**Lemma 4.6.** *Let $\mathcal{F}$ be an $r_n$-adic filtration and $(\varepsilon_n)_{n\leqslant 0}$ an innovation of $\mathcal{F}$. For every $\mathcal{F}_0$-measurable random variable $W_0$ in $A$, there exists an $r_n$-adic $\mathcal{F}$-split-word process $(W_n, \varepsilon_n)_{n\leqslant 0}$ with final letter $W_0$.*

*Proof.* We firstly construct $W_{-1}$. Since $W_0$ is measurable with respect to $\mathcal{F}_{-1} \vee \sigma(\varepsilon_0)$ there exist a $\mathcal{F}_{-1}$-measurable random variable $F_{-1}$ and a Borel function $f$ such that $W_0 = f(F_{-1}, \varepsilon_0)$. Define $W_{-1}$ as the $r_0$-word whose $j$-th letter is $f(F_{-1}, j)$. Now, assuming that $W_n$, …, $W_0$ are constructed, we construct $W_{n-1}$ in the same way: we write $W_n = g(F_{n-1}, \varepsilon_n)$ for some Borel function $g$ and some $\mathcal{F}_{n-1}$-measurable random variable $F_{n-1}$, and we define $W_{n-1}$ as the $\ell_{n-1}$-word such that, with notation 4.3, $g(F_{n-1}, j)$ is the $j$-th letter of $\widetilde{W}_{n-1}(j)$ for every integer $j \in [1, r_n]$. ☐

The equivalence between Vershik's intermediate level and second level criteria will be proved as a consequence of the following lemma.

**Lemma 4.7.** *Let $(r_n)_{n\leqslant 0}$ be a splitting sequence and $(\ell_n)_{n\leqslant 0}$ the corresponding length sequence. There exist some maps $\iota_n \colon A^{\ell_n} \to A^{(n)}$, $n \leqslant 0$, satisfying the following properties :*

- *the map $\iota_n$ induces an isometry from the quotient space $\frac{A^{\ell_n}}{G_n}$ to $A^{(n)}$, when $\frac{A^{\ell_n}}{G_n}$ is equipped with the distance $d_n^\theta$ (notation 4.4) and $A^{(n)}$ is equipped with the iterated Kantorovich distance $\theta_n$ (section 3.1);*

- *for any $r_n$-adic split-word process $(W_n, \varepsilon_n)_{n\leqslant 0}$ on $A$, one has $\pi_n W_0 = \iota_n(W_n)$.*

*Proof.* Firstly, it is not difficult to check that $\pi_n W_0 = \iota_n(W_n)$ where the maps $\iota_n \colon A^{\ell_n} \to A^{(n)}$ are recursively defined as follows. Given an integer $k \geqslant 2$ and a Polish space $(E, \rho)$, denote by $D_k \colon E^k \to E'$ the map defined by $D_k(x_1, \ldots, x_k) = \frac{1}{k}(\delta_{x_1} + \cdots + \delta_{x_k})$. Then define $\iota_0(w) = w$ and, using notation 4.3, define $\iota_{n-1}(w) = D_{r_n}\Big(\iota_n(\tilde{w}(1)), \ldots, \iota_n(\tilde{w}(r_n))\Big)$. From this construction it is easy to see that the map $\iota_n$ is invariant under the action of $G_n$, and then defines a map from $\frac{A^{\ell_n}}{G_n}$ to $A^{(n)}$. By lemma 3.1, the Kantorovich distance between $D_k(x)$ and $D_k(x')$ for any $x, x' \in E^k$ is given by $\rho'\Big(D_k(x), D_k(x')\Big) = \min_{\sigma \in \mathcal{S}_k} \frac{1}{k} \sum_{i=1}^{k} \rho(x_i, x'_{\sigma(i)})$. Using this fact and the recurrence relation (4.1) on $d_n^\theta$, it is easy to check by recursion that, for any words $w, w' \in A^{\ell_n}$ the Kantorovich distance between $\iota_n(w)$ and $\iota_n(w')$ equals $d_n^\theta(w, w')$. □

## 4.3 Vershik's intermediate level criterion

The statement of Vershik's intermediate level criterion, as well as its equivalence with Vershik's second level criterion, are based on the following lemma. Recall that the pseudo-distance $d_n^\theta$ is defined in notation 4.4.

**Lemma 4.8.** *With respect to an underlying filtration $\mathcal{F}$, let $(W_n, \varepsilon_n)_{n\leqslant 0}$ be a split-word process on a Polish metric alphabet $(A, \theta)$. Then* $\operatorname{disp} \pi_n W_0 = \widetilde{\mathbb{E}}\Big[d_n^\theta(W_n^*, W_n^{**})\Big]$ *where $(W_n^*)$ and $(W_n^{**})$ are independent copies of the process $(W_n)$ on some probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$. In other words,* $\operatorname{disp} \pi_n W_0 = \operatorname{disp}(G_n \cdot W_n)$ *denoting by $G_n \cdot W_n$ the orbit of $W_n$ under the action of $G_n$ (section 4.1).*

*Proof.* This straightforwardly results from lemma 4.7. □

Then, for any Polish metric space $(A, \theta)$, we say that, with respect to an $r_n$-adic filtration $\mathcal{F}$, a random variable $W_0 \in L^1(\mathcal{F}_0; A)$ satisfies *Vershik's intermediate level criterion* if, with the notation of the lemma above, $\widetilde{\mathbb{E}}\Big[d_n^\theta(W_n^*, W_n^{**})\Big]$ goes to 0 as $n$ goes to $-\infty$. We shortly say that $W_0$ satisfies the *intermediate Vershik property.* This definition makes sense in view of the lemma above, which shows that any property on the sequence of expectations $\mathbb{E}\Big[d_n^\theta(W_n^*, W_n^{**})\Big]$ only depends of $W_0$ (and of the underlying filtration), and in view of lemma 4.6, which guarantees the existence of a split-word process with final letter $W_0$.

We also extend the definition of the intermediate Vershik property to $\sigma$-fields $\mathcal{E}_0 \subset \mathcal{F}_0$ and to the whole filtration $\mathcal{F}$ as for the second Vershik property. We then immediately get the following theorem from lemma 4.8.

**Theorem 4.9.** *With respect to some $r_n$-adic filtration, the intermediate Vershik property and the second Vershik property are equivalent (for a random variable, a $\sigma$-field or the whole filtration).*

**Corollary 4.10.** *The analogue of proposition 3.2 for the intermediate Vershik property holds true.*

The following lemma which will be used in section 5.

**Lemma 4.11.** *Let $\mathcal{F}$ be an ambient $r_n$-adic filtration and $(A, \theta)$ be a Polish metric space. A random variable $W_0 \in L^1(\mathcal{F}_0; A)$ satisfies the intermediate Vershik property if and only if there exists a sequence $(w^{(n)})_{n\leqslant 0}$ consisting of words $w^{(n)}$ of length $\ell_n$ such that $\mathbb{E}\Big[d_n^\theta(W_n, w^{(n)})\Big]$ goes to $0$ for any split-word process $(W_n, \varepsilon_n)_{n\leqslant 0}$ with final letter $W_0$.*

*Proof.* This obviously results from the inequalities

$$\inf_w \mathbb{E}\Big[d_n^\theta(W_n, w)\Big] \leqslant \widetilde{\mathbb{E}}\Big[d_n^\theta(W_n^*, W_n^{**})\Big] \leqslant 2 \inf_w \mathbb{E}\Big[d_n^\theta(W_n, w)\Big],$$

which are easy to prove. □

# 5 The scale of an automorphism

Vershik defined the scale of an automorphism in [13]. We shall see that his definition can be rephrased in terms of Vershik's intermediate level criterion.

Let $T$ be an invertible measure-preserving transformation (in other words, an automorphism) of a Lebesgue space $(E, \nu)$. Vershik's definition of the scale of $T$ is the following one. With the same terminology used to describe the split-word processes, consider a splitting sequence $(r_n)_{n\leqslant 0}$ and the corresponding length sequence $(\ell_n)_{n\leqslant 0}$. Given $x \in E$, one defines a word $w_n(x)$ of length $\ell_n$ on $E$ for each $n \leqslant 0$ by $w_n(x) = (x, Tx, \ldots, T^{\ell_n - 1}x)$. Thus $w_n$ is a random word on $E$. The scale of $T$ is the set of splitting sequences $(r_n)_{n\leqslant 0}$ (consisting of integers $r_n \geqslant 2$) satisfying the following property: for every $f \in L^1(\nu)$ there exists a sequence $(c^{(n)})_{n\leqslant 0}$ consisting of vectors

$c^{(n)} \in \mathbb{R}^{\ell_n}$ such that the sequence of random variables $d_n\Big(f(w_n), c^{(n)}\Big)$ goes to 0 in probability, where $d_n$ is the pseudo-metric $d_n^\theta$ defined in notation 4.4 with $(A, \theta) = (\mathbb{R}, |\cdot|)$. The scale of $T$ is denoted by $\mathfrak{S}(T)$.

Denote by $\gamma_n$ the law of the random word $w_n$ on $E$. In view of lemma 4.11 and corollary 4.10, the sequence $(r_n)_{n\leqslant 0}$ belongs to the scale of $T$ if and only if, with respect to some $r_n$-adic filtration, Vershik's intermediate level criterion holds for the final letter $W_0$ of some split-word process $(W_n, \varepsilon_n)_{n\leqslant 0}$ with $W_n \sim \gamma_n$. The $\gamma_n$ obviously satisfy the consistency condition required for the existence of this process.

It is clear, owing to corollary 4.10, that this property defines an invariant of $T$: if $S = \phi \circ T \circ \phi^{-1}$ is an invertible measure-preserving transformation conjugate to $T$ then the property is equivalent to Vershik's intermediate level criterion holding for the final letter $\phi(W_0)$ of the split-word process $\Big(\phi(W_n), \varepsilon_n\Big)_{n\leqslant 0}$. Some other basic properties of the scale given by Vershik in [13] can be derived as direct applications of basic properties of Vershik's intermediate level criterion, such as corollary 4.10.

**Remark 5.1.** Actually, as pointed out by Vershik in [13], his definition above only concerns completely ergodic transformations $T$ (i.e. all powers of $T$ are ergodic), and it may be extended to arbitrary transformations $T$ by replacing each vector $c^{(n)}$ by a vector-valued function $c^{(n)}(x)$ constant on the ergodic components of $T^{\ell_n}$. We shall not investigate this extension, except for our concluding remarks at the end of this section; the reader has to be aware that some results in [13] are not valid with the definition given above.

In the sequel we denote by $(W_n, \varepsilon_n)_{n\leqslant 0}$ the (unique in law) split-word process associated to $T$ for a given sequence $(r_n)_{n\leqslant 0}$, and $\mathcal{F}$ the $r_n$-adic filtration it generates (unique up to isomorphism). In particular, $W_0$ is an $E$-valued random variable with law $\nu$.

The following proposition shows in particular that the filtration of the ordinary split-word process (example 4.5) is the filtration $\mathcal{F}$ when $T$ is the Bernoulli shift on $(A, \mu)^{\mathbb{Z}}$. Thus the scale of a Bernoulli shift coincides with the set of sequences $(r_n)_{n\leqslant 0}$ for which the corresponding ordinary split-word process generates a standard filtration. A *generator* of an automorphism $T$ is a measurable function $f$ from $(E, \nu)$ to a Lebesgue space $(A, \mu)$ such that $\sigma(X) = \vee_{i=-\infty}^{\infty} \sigma\Big(f(T^i X)\Big)$ for some ( $\iff$ for every) random variable $X$ distributed on $E$ according to $\nu$. For a (possibly non Bernoulli) shift on $A^{\mathbb{Z}}$, a natural generator is the function $f\colon A^{\mathbb{Z}} \to A$ which sends a sequence in $A^{\mathbb{Z}}$

to its coordinate at index 0.

**Proposition 5.2.** *If $f$ is a generator of $T$, then $\mathcal{F}$ is the filtration generated by the split-word process $\Big(f(W_n), \varepsilon_n\Big)_{n \leqslant 0}$.*

*Proof.* It suffices to show that the first letter $W_n(1)$ of $W_n$ is measurable with respect to $\vee_{m=-\infty}^{n} \sigma\Big(f(W_n), \varepsilon_n\Big)$ for every $n \leqslant 0$. For notational convenience, we only treat the case $n = 0$, and it will be clear how to similarly treat the case of any $n \leqslant 0$. One has $f(W_n) = \Big(f(T^{P_n}W_0), \ldots, f(T^{Q_n}W_0)\Big)$ where $P_n \leqslant 0$ and $Q_n \geqslant 0$ are random integers measurable with respect to $\sigma(\varepsilon_{n+1}, \ldots, \varepsilon_0)$, and satisfy $P_n \to -\infty$ and $Q_n \to +\infty$ since they obviously satisfy $P_n \leqslant -\sum_{i=n+1}^{0} \ell_i \mathbb{1}_{\varepsilon_i \neq 1}$ and $Q_n \geqslant \sum_{i=n+1}^{0} \ell_i \mathbb{1}_{\varepsilon_i \neq r_i}$, thereby showing that $W_0$ is measurable with respect to $\vee_{m=-\infty}^{0} \sigma\Big(f(W_n), \varepsilon_n\Big)$. □

A famous theorem by Rokhlin says that any aperiodic transformation $T$ has a countable generator $f$, that is, $f$ takes its values in a countable space. Recall that $T$ is said to be *aperiodic* when $\mathbb{P}(W_0 = T^i W_0 \text{ for some } i \geqslant 1) = 0$. In particular, $T$ is aperiodic whenever it is ergodic.

In the sequel we denote by $D_n = W_n(1)$ the first letter of $W_n$ for every $n \leqslant 0$. Obviously, $\mathcal{F}$ is also generated by the process $(D_n, \varepsilon_n)_{n \leqslant 0}$ and $D_n = (T^{\ell_n})^{\varepsilon_n - 1}(D_{n-1})$. We state the proposition below only by way of remark.

**Proposition 5.3.** *If $T$ is aperiodic then $\mathcal{F}$ is generated by the process $(W_n)_{n \leqslant 0}$.*

*Proof.* It suffices to show that $\sigma(\varepsilon_n) \subset \sigma(D_{n-1}, D_n)$ when assuming aperiodicity of $T$. Let $K_n \leqslant J_n := (\varepsilon_n - 1)\ell_n$ be the smallest integer such that $D_n = T^{K_n}(D_{n-1})$, hence the equality $T^{K_n}(D_{n-1}) = T^{J_n}(D_{n-1})$ almost surely holds. Therefore $\mathbb{P}(K_n \neq J_n) \leqslant \sum_{k \neq j} \mathbb{P}\left(T^k(D_{n-1}) = T^j(D_{n-1})\right) = 0$. □

Now, for an aperiodic $T$, we shall prove that $(r_n)_{n \leqslant 0} \in \mathfrak{S}(T)$ means that the whole filtration $\mathcal{F}$ is Vershikian (theorem 5.5).

**Lemma 5.4.** *If $T$ is aperiodic then $\sigma(\pi_n D_0) = \sigma(D_n)$.*

*Proof.* It suffices to show that $\mathcal{L}(D_n \,|\, \mathcal{F}_{n-1})$ generates the same $\sigma$-field as $D_{n-1}$ for every $n \leqslant 0$ since the $\sigma$-field generated by the conditional law of a random variable $X$ given any $\sigma$-field depends on $X$ only through the $\sigma$-field $\sigma(X)$. To do so, put $S = T^{\ell_n}$. Conditionally on $\mathcal{F}_{n-1}$, the random variable $D_n$ is uniformly chosen among $D_{n-1}$, $S(D_{n-1})$, …, $S^{r_n - 1}(D_{n-1})$. Hence, the conditional law $\mathcal{L}(D_n \,|\, \mathcal{F}_{n-1})$ determines the set $\{D_{n-1}, S(D_{n-1}), \ldots, S^{r_n-1}(D_{n-1})\}$. Let $\sigma$ be a random permutation of $I := \{0, 1, \ldots, r_n - 1\}$ such that $S^{\sigma(j)}(D_{n-1}) =$

$S^j\left(S^{\sigma(0)}(D_{n-1})\right)$ for every $j \in I$. We shall show that $\sigma$ almost surely equals the identity map of $I$; the lemma will obviously follow. Let $K = 0$ if $\sigma$ is the identity permutation and $K$ be a strictly positive integer such that $\{S^K(D_{n-1}) = D_{n-1}\}$ otherwise. Then $\mathbb{P}(K \neq 0) \leqslant \sum_{k>0} \mathbb{P}\left(S^k(D_{n-1}) = D_{n-1}\right) = 0$. □

**Theorem 5.5.** *For an aperiodic $T$, the sequence $(r_n)_{n\leqslant 0}$ belongs to the scale of $T$ if and only if the $r_n$-adic filtration $\mathcal{F}$ is Vershikian.*

*Proof.* This stems from lemma 5.4 and lemma 3.4. □

Now we will give a proof of the following result based on the theorem on lacunary isomorphism (theorem 3.3). Vershik proved it in [13] by a direct construction.

**Proposition 5.6.** *The scale of a completely ergodic invertible measure-preserving transformation is not empty.*

Our proof is an application of the theorem on lacunary isomorphism based on corollary 5.8 which is derived from proposition 5.7 below. The following notations are used in this proposition. We consider a splitting sequence $(r_n)_{n\leqslant 0}$ and for each $n \leqslant 0$, we put $\mathcal{J}_n = D_n^{-1}(\mathcal{I}_n)$ where $\mathcal{I}_n$ is the $\sigma$-field of $T^{\ell_n}$-invariant events. A random variable is $\mathcal{J}_n$-measurable if and only if it is of the form $f(D_n)$ where $f$ is a $T^{\ell_n}$-invariant Borel function. Therefore $(\mathcal{J}_n)_{n\leqslant 0}$ is a decreasing sequence of $\sigma$-fields. Note also that for any $\mathcal{J}_n$-measurable random variable $J_n$ and any integer $i$ which is multiple of $\ell_n$, the random pair $(J_n, D_n)$ has the same distribution as $(J_n, T^i D_n)$. Recall that the ergodic theorem says that

$$\frac{1}{k}\sum_{i=0}^{k-1} f(T^i D_n) \xrightarrow{L^1} \mathbb{E}\Big[f(D_n) \,|\, \mathcal{J}_n\Big] \quad \text{as } k \to +\infty$$

for all $f \in L^1(\nu)$.

**Proposition 5.7.** *Let $T$ be an invertible measure-preserving transformation and $\mathcal{F}$ the $r_n$-adic filtration associated to $T$ as above. Then $\mathcal{F}_{-\infty} = \lim \nearrow \mathcal{J}_n$.*

*Proof.* Recall that we denote by $D_n = W_n(1)$ the first letter of $W_n$. It is clear that $\{D_n \in A\} \in \mathcal{F}_{-\infty}$ for every Borel set $A$ which is invariant by $T^{\ell_n}$, thereby yielding the inclusion $\lim \nearrow \mathcal{J}_n \subset \mathcal{F}_{-\infty}$.

Conversely, putting $\mathcal{J}_{-\infty} = \lim \nearrow \mathcal{J}_n$, it suffices to show that $\mathbb{E}[Z \,|\, \mathcal{F}_n]$ tends in $L^1$ to a $\mathcal{J}_{-\infty}$-measurable random variable for each $Z \in L^1(\mathcal{F}_0)$. It is not difficult to check that this property holds whenever it holds for all random variables $Z = f(D_{n_0})$ where $f \in L^1(\nu)$. Then, given $f \in L^1(\nu)$, we shall show that $\mathbb{E}\left[f(D_{n_0}) \,|\, \mathcal{F}_n\right] \xrightarrow{L^1} \mathbb{E}\left[f(D_{n_0}) \,|\, \mathcal{J}_{n_0}\right]$ for every $n_0 \leqslant 0$. Conditionally on $\mathcal{F}_n$ (with $n \leqslant n_0$), $D_{n_0}$ is uniformly chosen among $D_n$, $T^{\ell_{n_0}}(D_n)$, …, $T^{\left(\ell_n/\ell_{n_0}-1\right)\ell_{n_0}}(D_n)$, therefore

$$\begin{aligned}\mathbb{E}\left[f(D_{n_0}) \,|\, \mathcal{F}_n\right] &= \frac{1}{\ell_n/\ell_{n_0}-1} \sum_{j=0}^{\ell_n/\ell_{n_0}-1} f\Big(\left(T^{\ell_{n_0}}\right)^j(D_n)\Big) \\ &= \frac{1}{\ell_n/\ell_{n_0}-1} \sum_{j=0}^{\ell_n/\ell_{n_0}-1} f\Big(\left(T^{\ell_{n_0}}\right)^j(T^X D_{n_0})\Big)\end{aligned}$$

where $X$ is a random integer which is multiple of $\ell_{n_0}$ and is independent of $D_{n_0}$, and then the result follows from the ergodic theorem. □

**Corollary 5.8.** *The filtration $\mathcal{F}$ is Kolmogorovian if and only if $T^{\ell_n}$ is ergodic for all $n \leqslant 0$.*

Proposition 5.6 is then straightforwardly shown by applying the theorem on lacunary isomorphism (theorem 3.3), by noting that any filtration $(\mathcal{F}_{\phi(n)})_{n\leqslant 0}$ extracted from $\mathcal{F}$ with $\phi(0) = 0$ is the filtration of the split-word process associated to $T$ with another splitting sequence.

Now we prove the proposition below as another illustration of our definition of the scale. We do not know whether the converse inclusion holds.

**Proposition 5.9.** *Let $S$ and $T$ be invertible measure-preserving transformations. Then $\mathfrak{S}(S \times T) \subset \mathfrak{S}(S) \cap \mathfrak{S}(T)$.*

*Proof.* Let $(W_n, \varepsilon_n)_{n\leqslant 0}$ be the $r_n$-adic split-word process associated to $S \times T$ and $\mathcal{F}$ the filtration it generates. For each $n \leqslant 0$, one has $W_n = (Y_n, Z_n)$ and $(Y_n, \varepsilon_n)_{n\leqslant 0}$ and $(Z_n, \varepsilon_n)_{n\leqslant 0}$ are the $r_n$-adic split-word processes associated to $S$ and $T$ respectively, each of them generating a filtration immersed in $\mathcal{F}$. Therefore the result follows from the fact that $\sigma(W_0)$ is Vershikian if $(r_n)_{n\leqslant 0} \in \mathfrak{S}(S \times T)$ (proposition 3.2) and from the hereditariness of the Vershik property for immersion (lemma 3.5). □

We close this section by translating Vershik's general definition of the scale (remark 5.1) into a probabilistic statement generalizing the preceding definition. Vershik's general definition says that $(r_n)$ belongs to $\mathfrak{S}(T)$ if $d_n(W_n) \to 0$ where $d_n(W_n) = \inf_w \mathbb{E}[d_n(W_n, w) \,|\, \mathcal{J}_n]$ (where $d_n$ is the pseudometric introduced in notation 4.4 and the $\mathcal{J}_n$ are $\sigma$- fields introduced before proposition 5.7). In addition one also has $d_n(W_n) = \inf_w \mathbb{E}[d_n(W_n, w) \mid \mathcal{F}_{-\infty}]$, owing to the following lemma.

**Lemma 5.10.** *The random variable $D_n$ is conditionally independent of $\mathcal{F}_{-\infty}$ given $\mathcal{J}_n$.*

*Proof.* Recall that $\mathcal{F}_{-\infty} = \lim \nearrow \mathcal{J}_n$ (proposition 5.7). The statement of the lemma amounts to saying that $\mathbb{E}[g(D_n) \mid \mathcal{J}_n] = \mathbb{E}[g(D_n) \mid \mathcal{F}_{-\infty}]$ for $g \in L^1(\nu)$, which is equivalent to $\mathbb{E}[g(D_n) \mid \mathcal{J}_n] = \mathbb{E}[g(D_n) \mid \mathcal{J}_m]$ for every $m \leqslant n$. To prove this equality, we take a random variable $J_m \in L^\infty(\mathcal{J}_m)$, hence $J_m = f(D_m)$ where the function $f \in L^\infty(\nu)$ is $T^{\ell_m}$-invariant. Since $D_m = {(T^{\ell_n})}^X D_n$ where $X$ is a random integer independent of $D_0$ and uniformly distributed on $\{0, -1, \ldots, -\ell_m/\ell_n + 1\}$,

$$\mathbb{E}\left[J_m g(D_n)\right] = \frac{1}{\ell_m/\ell_n} \sum_{i=-\ell_m/\ell_n+1}^{0} \mathbb{E}\left[f({(T^{\ell_n})}^i D_n) g(D_n)\right] = \mathbb{E}\left[J_n g(D_n)\right]$$

where $J_n = \frac{1}{\ell_m/\ell_n} \sum_{i=-\ell_m/\ell_n+1}^{0} f\Big({(T^{\ell_n})}^i D_n\Big)$. But $J_n$ is measurable with respect to $\mathcal{J}_n$ since

$$\sum_{i=-\ell_m/\ell_n+1}^{0} f\Big({(T^{\ell_n})}^i T^{\ell_n} D_n\Big) = f(T^{\ell_n} D_n) + \sum_{i=-\ell_m/\ell_n+2}^{0} f\Big({(T^{\ell_n})}^i D_n\Big)$$

and $f(T^{\ell_n} D_n) = f(T^{-\ell_m+\ell_n} D_n)$. Therefore we get

$$\mathbb{E}\left[J_m g(D_n)\right] = \mathbb{E}\Big[J_n \mathbb{E}[g(D_n) \mid \mathcal{J}_n]\Big]$$

but we similarly prove that $\mathbb{E}\Big[J_m \mathbb{E}[g(D_n) \mid \mathcal{J}_n]\Big] = \mathbb{E}\Big[J_n \mathbb{E}[g(D_n) \mid \mathcal{J}_n]\Big]$. ☐

Thus, the general definition of "$r_n \in \mathfrak{S}(T)$" sounds like the intermediate Vershik property of $W_0$ conditionally on $\mathcal{F}_{-\infty}$. It is expected the criteria for filtrations (such as Vershik's criteria and the I-cosiness criterion) can be more generally stated conditionally on $\mathcal{F}_{-\infty}$ so that most results (such as the theorem on lacunary isomorphism) remain to be true. But, currently, we do not feel the motivation to develop this generalization.

# References


[1] F. Bolley, 2008. *Separability and completeness for the Wasserstein distance.* Séminaire de Probabilités XLI Lecture Notes in Mathematics, 2008, Volume 1934/2008, 371–377

[2] G. Ceillier, 2010. *Filtrations à temps discret.* PhD dissertation. Institut Fourier de Grenoble.

[3] Y. Coudène, *Une version mesurable du théorème de Stone-Weierstrass*, Gazette des mathématiciens **91**, SMF, janvier 2002.

[4] M. Émery, W. Schachermayer: On Vershik's standardness criterion and Tsirelson's notion of cosiness. Séminaire de Probabilités XXXV (Springer Lectures Notes in Math. 1755, Berlin 2001) pp 265–305

[FS00] J. Feldman, M. Smorodinsky, *Decreasing sequences of measurable partitions: product type, standard and prestandard*, Ergodic Theory and Dynamical Systems **20**, 1079–1090, 2000.

[HH] D. Heicklen, C. Hoffmann, $[T, T^{-1}]$ *is not standard*, Ergodic Theory and Dynamical Systems **18**, 875–878, 1998.

[Ho] C. Hoffman, *A zero entropy $T$ such that the $[T, \mathrm{Id}]$ endomorphism is nonstandard*, Proc. A.M.S. **128**, no. 1, 183-188, 1999.

[HR] C. Hoffman, D. Rudolph, *If the $[T, \mathrm{Id}]$ automorphism is Bernoulli then the $[T, \mathrm{Id}]$ endomorphism is standard*, Proc. A.M.S. **128**, no. 1, 183-188, 1999.

[5] S. Laurent: Filtrations à temps discret négatif. PhD Thesis, Université de Strasbourg, Strasbourg (2004)

[6] S.Laurent: On standardness and I-cosiness. Séminaire de Probabilités XLIII (2010), 127–186.

[7] S. Laurent: On Vershikian and I-cosy random variables and filtrations. Teoriya Veroyatnostei i ee Primeneniya 55 (2010), 104–132. Also published in: Theory Probab. Appl. 55 (2011), 54–76

[8] M. Smorodinsky, 1998. Processes with no standard extension. Israel Journal of Mathematics 107, 327–331.

[9] A.M. Vershik, *Theorem on lacunary isomorphisms of monotonic sequences of partitions*. Funktsional'nyi Analiz i Ego Prilozheniya **2**:3, 17–21, 1968. English translation: Functional Analysis and Its Applications **2**:3, 200–203, 1968.

[10] A.M. Vershik, *Decreasing sequences of measurable partitions, and their applications*. Dokl. Akad. Nauk SSSR, **193**, 748–751, 1970. English translation: Soviet Math. Dokl. **11**, 1007–1011, 1970.

[11] A.M. Vershik, *Continuum of pairwise nonisomorphic diadic sequences*. Funktsional'nyi Analiz i Ego Prilozheniya **5**:3, 16–18, 1971. English translation: Functional Analysis and Its Applications **5**:3, 182–184, 1971.

[12] A.M. Vershik, *Approximation in measure theory* (in Russian). PhD Dissertation, Leningrad University, 1973. (expanded and updated version: [14])

[13] A.M. Vershik: Four definitions of the scale of an automorphism. Funktsional'nyi Analiz i Ego Prilozheniya, **7**:3, 1–17 (1973). English translation: Functional Analysis and Its Applications, **7**:3, 169–181 (1973).

[14] A.M. Vershik: The theory of decreasing sequences of measurable partitions (in Russian). Algebra i Analiz, **6**:4, 1–68 (1994). English translation: St. Petersburg Mathematical Journal, **6**:4, 705–761 (1995)

[15] A. M. Vershik, V. A. Rokhlin and the modern theory of measurable partitions, Amer. Math. Soc. Transl. (2) Vol. 202, 11–20, 2001.