AN ABSTRACT OF THE THESIS OF

David C. Plaehn for the degree of Doctor of Philosophy in Mathematics

presented on May 9, 1997.

Title: One-Dimensional Kohonen Maps are Super-stable with Exponential Rate.

Abstract approved: Signature redacted for privacy.

Robert M. Burton

We consider the one-dimensional Kohonen algorithm with iid stimuli and two different kernel functions: (1) the standard (indicator) kernel but with two or more neighbors and (2) a strictly decreasing kernel. We show that case (2) is self-organizing with probability one and that a.s. self-organization in case (1) depends in part on neighborhood size relative to the total number of 'particles' ('neurons', 'units'). For both cases we show that if the stimulus distribution is absolutely continuous with respect to Lebesgue measure and the density is bounded above and away from zero, then the Kohonen algorithm produces a super-stable Markov chain with exponential rate, i.e. given two starting maps, X_0, Y_0, and the same stimuli, then for the resulting Markov chains, $\{X_n\}, \{Y_n\}$ (i) $D(X_n, Y_n) \leq e^{-\beta(n-W)}$ (ii) there exists a finite invariant measure for the chains and (iii) $\boldsymbol{P}(W = n) \leq AB^n$ where $\beta, A > 0$, $0 < B < 1$, $D(\cdot,\cdot)$ is a metric on the space of maps and W is a positive, integer-valued random variable depending on X_0 and Y_0.

One-Dimensional Kohonen Maps are Super-stable with Exponential Rate

by

David C. Plaehn

A dissertation

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed May 9, 1997
Commencement June 1997

Doctor of Philosophy thesis of David C. Plaehn presented on May 9, 1997

APPROVED:

Signature redacted for privacy.

Major Professor, representing Mathematics

Signature redacted for privacy.

Chair of the Department of Mathematics

Signature redacted for privacy.

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Signature redacted for privacy.

David C. Plaehn, Author

ACKNOWLEDGMENT

I wish to thank Dr. Robert M. Burton for his guidance and good humor, and my parents for all their encouragement.

CONTRIBUTION OF AUTHORS

Dr. Robert M. Burton was involved in the outline of the approach [9] as well as the writing of this thesis, especially, part of the Introduction the latter half of Chapter 4.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

DEDICATION

To my mother,

Marilyn Rose Plaehn

and in memory of my father,

Louis Carsen Plaehn.

ONE-DIMENSIONAL KOHONEN MAPS ARE SUPER-STABLE WITH EXPONENTIAL RATE

1. INTRODUCTION

In articles published in the early 80's ([27], [24], [25], [26]) the Finnish electrical engineer, Tuevo Kohonen, introduced a type of neural network now called a self-organizing map (SOM) or self-organizing feature map. The model worked remarkably well in simulations in duplicating two phenomena observed or suspected in biological systems:

(1) Self-organization—the ability of a system to organize itself according to a certain sensory or receptor feature.

(2) Quantization—the ability to mimic "measures" on the sensory feature or receptors.

We consider a generalization of the model presented by Kohonen in [27]. The key feature of this model is the algorithm for updating the weights. The process is an iterative one, generating a sequence of weights $\{X_n\}$. The latter is a Markov chain.

1.1. Biological Background

With regards to self-organization ((1) above) and biological systems, there is a correspondence between areas of the mammalian cerebral cortex and the senses. Within each area there is further organization relative to features of the stimuli.

For example, the somatosensory region is roughly a map of the the human body (the areas of the body sensitive to touch). The auditory cortex corresponding to pitch is organized in a linear, monotonic fashion according to frequency. In general, similar inputs are mapped onto neighboring areas.

To illustrate (2), note that the area of the human somatosensory cortex associated with the (richly innervated) fingers is relatively large when compared with the area associated with less innervated regions, e.g., the legs or back.

Apparently, genetics accounts for the broad outlines of this organization. However, there is a great degree of plasticity to this arrangement, especially in the early stages of development of the nervous system.

1.2. Mathematical Description

Let $V \subset \boldsymbol{Z}^{d^*}$ be a finite vertex set and let $\widehat{\Upsilon}^d = \{X : V \to [0,1]^d | X$ is one-to-one$\}$ where $d^* \leq d$. Let

$$D(X,Y) = \min\{\max_{i \in V}\|X(i) - Y(i)\|, \max_{i \in V}\|X(i) - Y(\ell - i + 1)\|\} \qquad (1.1)$$

where $\|\cdot\|$ is the euclidean norm. Thus D is a semimetric on $\widehat{\Upsilon}^d$. Define Υ^d to be the set of equivalence classes on $\widehat{\Upsilon}^d$ so as to make (Υ^d, D) a metric space. This will be the state space of our Markov dynamical system. When $d = 1$ let $\Upsilon^1 = \Upsilon$. Also, let Υ_{m} be the set of elements of Υ that are monotonic increasing or decreasing.

We have an underlying probability distribution μ which we call the *environment.* The environment is expressed by iid random variables $\{\omega_n\}_{n=1}^{\infty} \subset [0,1]^d$ which have the distribution μ. There is also a kernel or neighborhood function $K : \{0, 1, \ldots, \ell - 1\} \to [0,1]$ or $K : \{0, 1, \ldots, \ell - 1\} \times \boldsymbol{N} \to [0,1]$ which is assumed to be non-increasing and, in the second case, to be dependent on time.

Let $X_0 \in \Upsilon^d$ be random with a distribution absolutely continuous with respect to Lebesgue measure. If X_n is the state of the process at time n then the state of the process at time $n+1$, X_{n+1}, is defined as follows. Let

$$i'_{n+1} := \arg\min_{i \in V} \|X_n(i) - \omega_{n+1}\|.$$

Let

$$X_{n+1} = F(X_n, \omega_{n+1}) \tag{1.2}$$

and define

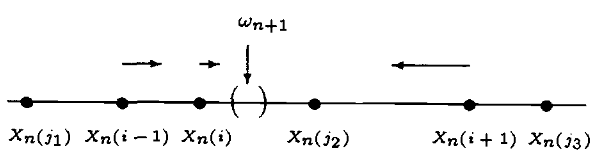
$$F(X_n, \omega_{n+1})(i) = K(|(i'_{n+1} - i|)\omega_{n+1} + (1 - K(|(i'_{n+1} - i|))X_n(i) \tag{1.3}$$

$$= X_n(i) + K(|(i'_{n+1} - i|)(\omega_{n+1} - X_n(i)) \tag{1.4}$$

We restrict our attention the so-called *Kohonen string* where $d^* = 1$ and $V = \{1, 2, \ldots, \ell\}$. Other oft-cited models are the Kohonen 0-neighbor model (no graph structure) and the Kohonen grid ($d^* = 2$). Often the environmental distribution will be Lebesgue measure. It is also common in practice to take the kernel to be

$$K(i) = \alpha \mathbf{1}_{\{i \leq r\}}.$$

In this case we call $\alpha \in (0, 1)$ the *shrinking parameter* and r the *range parameter* and the process the *standard Kohonen process.*



$$X_{n+1}(j) = X_n(j) + \alpha \mathbf{1}_{\{|j-i| \leq 1\}}(j)(\omega_{n+1} - X_n(j))$$

FIGURE 1.1. The standard Kohonen process, $r = 1$ (nearest neighbor model).

1.3. Neural Networks

Abstractly, a neural network is a directed graph [8] whose vertices are called **neural units** or **processing units**. The edges of the graph show relations between the neural units, representing synaptic connections. Typically, with each edge is associated a weight. The arrows of the graph indicate the flow of information or activity. A neural units output is a function of its input. The problem we consider corresponds to a network with one input or receptor unit connected to each of ℓ output or sensory units (Figure 1.2). The outputs are linearly ordered from 1 to ℓ. To the connection or neural link between the input and the output unit i, a weight $X(i)$ is assigned. The input receives iid stimuli $\{\omega_n\}$ distributed as μ. Typically, ω_k and $X(i)$ reside in the same space $S \subseteq \boldsymbol{R}^d$. In our case, $S = [0,1]$. Which output unit(s) is/are triggered or activated by a given stimuli ω is determined by the argument i that minimizes $\|\omega - X(i)\|$ where $\|\cdot\|$ is the Euclidean norm. If i' is the 'selected' argument or 'vertex' then output unit i' and perhaps some of its neighbors are activated. All output units receive the input data but activation is typically local.

The algorithm was originally devised in a two-dimensional setting by Kohonen as a model of a biological phenomenon called retinotopy—the self-organization of the visual cortex according to the 'grid' of receptors in the retina. The algorithm was extended to other dimensions and extensively studied [34], [28], especially using non-rigorous simulations. Although the above models are not realistic in a biological sense, they have found numerous applications, in various forms, in speech processing [28], image pre-processing and compression [30], insurance scoring, statistical data analysis [36] and numerical integration [35]. (Some of the latter applications come under the heading, vector quantization [30], [21], [6].)

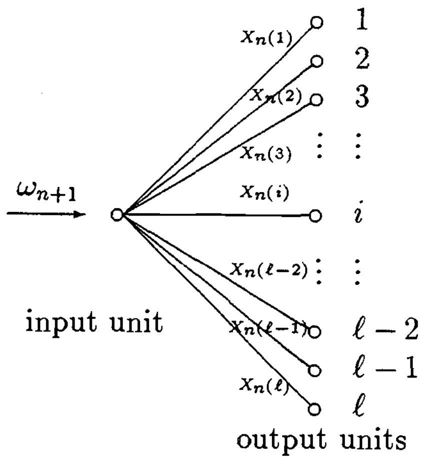


FIGURE 1.2. The one-dimensional Kohonen neural network.

1.4. Results

Our contribution to this theory is to show that the underlying Markov process may be viewed as a composition of iid maps and that this process has a strong stability property we call super-stable. In particular the Markov evolution quickly forgets its initial state. Thus the long range future of the process depends only on the (random) environment. We show that in the so-called *standard Kohonen process*, the size of the range parameter r is critical with respect to the certainty of self-organization. We also show self-organization and super-stability for strictly decreasing kernel functions. Such kernels have been used in some models, for example, algorithms to handle the travelling salesman problem [34]. Our approach is constructive and geometric, differing from the methods used in the literature.

2. REVIEW OF THE LITERATURE

Beyond Kohonen's 1982 paper "Analysis of a Simple Self-Organizing Process,' most of what was known about the algorithm was via simulations (see, for example, [26]). The first fundamental rigourous paper was that of Cottrell and Fort in 1987 [13]. The first results were for the standard Kohonen process in the nearest neighbor ($r = 1$) and the zero neighbor ($r = 0$) cases. Although, we are restricting our attention to the Kohonen string (so in the standard Kohonen process, this means $r \geq 1$), we mention that the zero neighbor case has been shown to be both super-stable [9] and Doeblin [6] in d dimensions. Also, see [20] for other results.

In the following, we will denote Lebesgue measure (on the unit interval) by λ, the support of a measure m by $\mathrm{supp}(m)$, and the continuous or diffuse part of a measure m by m_c (i.e. $m_c(x) = 0$ for all x).

2.1. The Nearest Neighbor Model

Rigourous results on this topic are mainly restricted to the following: Cottrell and Fort (C&F/87) [13] and Bouton and Pagès (B&P/93) [4], (B&P/94) [5].

2.1.1. *Self-Organization*

In the one-dimensional model, self-organization refers to the tendency of the maps to become monotonic (increasing or decreasing). Kohonen showed that Υ_{m} is an absorbing class for the process. For a.s. self-organization of the process, we have

Author/Date	Conditions on μ
C&F/87	$\mu = \lambda$
B&P/93	supp(μ_c) contains a non-empty open set

If $\tau_{\mathrm{m}} := \inf_n\{X_n \in \Upsilon_{\mathrm{m}}\}$—the hitting time for Υ_{m}—then in both of the above articles it is proved that $\forall\ X_0\ \exists\ T, \eta > 0$ such that $\boldsymbol{P}(\tau_{\mathrm{m}} < T) > \eta$ which implies $\boldsymbol{P}(\tau_{\mathrm{m}} < \infty) = 1$.

Bouton and Pagès give an example of a measure μ with no diffuse part where self-organization does not occur.

2.1.2. *Existence of a Unique Invariant Measure*

The following results were found concerning the existence of a unique invariant measure for the Markov chain. See section 3.4.3 for the definitions of 'invariant measure,' 'Doeblin' and 'Feller.'

Author/Date	Conditions on μ	Comment
C&F/87	$\mu = \lambda$	Proved Doeblin recurrence
B&P/94	$\exists\ O \subset [0,1]$, open set, $\exists\ \eta$ such that $\mu_{\mid O} \geq \eta\lambda_{\mid O}$	Proved Doeblin recurrence
B&P/94	μ is continuous	Modified chain is Feller

2.1.3. *Stability*

In our approach to the stability of the process we begin with two starting positions X_0 and Y_0 and look at $D(X_n, Y_n)$ where our neighborhood function K is independent of time. Other researchers have used neighborhood functions of the form $K(i,n) = \alpha_n \sigma(i)$ where $\alpha_n \in (0,1)$ for all n and $\sigma : \{0, 1, \ldots, \ell\} \to [0,1]$ with $\sigma(0) = 1$. In addition, the following conditions are imposed

$$\sum_{n=1}^{\infty} \alpha_n = \infty \qquad \sum_{n=1}^{\infty} \alpha_n^2 < \infty. \tag{2.1}$$

Applying the algorithm with this neighborhood function, the stability of an associated ordinary differential equation (ODE) is investigated. To describe this in more detail we need

Definition 1 *The* Voronoï tessellation $C_i(x)_{i \in V}$ *of* $x \in \Upsilon$ *is defined by*

$$C_i(x) = \{z \in [0,1]^d | ||x_i - y|| < ||x_k - y||, \ k \neq i\}, \quad \forall i \in V$$

Now let

$$h_i^\sigma := \sum_{k \in V} \sigma(|k-i|) \int_{C_k(x)} (x_i - \omega)\mu(d\omega) \tag{2.2}$$

So h_i^σ is an averaging function. The associated ODE is given by

$$\dot{x} = -h^\sigma(x_0), \quad x_0 \in \Upsilon \tag{2.3}$$

Conditional convergence, in the sense of Kushner and Clark, to an (asymptotically) stable equilibrium x^* of the ODE is then studied. (Recall that x^* is an equilibrium point of the given ODE if $h^\sigma(x^*) = 0$.) An equilibrium x^* is said to be stable if it has a stable attracting area Γ_{x^*}. Let $x(x^0, \cdot)$ denote any solution of the ODE starting at x^0; Γ_{x^*} is defined as a neighborhood of x^* satisfying the following:

(i) $\forall x^0 \in \Gamma_{x^*},\ \forall u \in \boldsymbol{R}_+,\ x(x^0, u) \in \Gamma_{x^*}$

(ii) $\forall x^0 \in \Gamma_{x^*},\ \lim_{u\to\infty} x(x^0, u) = x^*$

(iii) $\forall C \subset \Gamma_{x^*}$, C compact, $\forall \epsilon > 0, \exists\ \eta_{\epsilon,C} > 0$ such that

$$\forall x^0 \in C,\ ||x^0 - x^*|| \leq \eta_{\epsilon,C} \Rightarrow \sup_{u\in R_+} ||x(x^0, u) - x^*|| \leq \epsilon$$

The Kushner and Clark theorem is as follows:

Theorem 1 *Assume that* $\sum_{t\geq 0} \alpha_t = \infty$ *and* $\sum_{t\geq 0} \alpha_t^2 < \infty$. *Let* x^* *be a zero of* h *and let* C *be a compact subset of its stable attracting area* Γ_{x^*}. *Then the sequence* $(X_t)_{t\geq 0}$ *"conditionally " (a.s.) converges to* x^**; that is,*

$$\lim_{t\to\infty} X_t = x^*$$

on the event $A_C^{x^*} := \{(X_t)_{t\geq 0}$ *is bounded and* $X_t \in C$ *infinitely often*$\}$.

This leads to the following

Definition 2 *If* $(X_t)_{t\in N}$ *satisfies Theorem 1 and* x^* *is a stable equilibrium,* $(X_t)_{t\in N}$ *"conditionally " a.s. converges to* x^*.

Let F^+ be the set of increasing maps in $\Upsilon_{\mathbf{m}}$ and F^- be the set of decreasing maps in $\Upsilon_{\mathbf{m}}$ (so $\Upsilon_{\mathbf{m}} = F^+ \cup F^-$). With the above in mind, assuming the condition in equation 2.1, the following results were obtained:

Author/Date	Conditions on μ	Result
C&F/87	$\mu = \lambda$	h^σ has a unique equilibrium x^* in F^+ and $X_t \to x^*$ a.s.
B&P/93	$\text{supp}(\mu) = [0,1]$ and diffuse	h^σ has at least one equilibrium point x^* in F^+
B&P/93	μ has a strictly log-concave density $f > 0$ on $(0,1)$	any equilibrium x^* is stable, so $X_t \to x^*$ conditionally a.s.

2.2. More General One-Dimensional Models

These models assume σ to be non-increasing and $r \geq 1$.

2.2.1. *Self-Organization*

Apparently, Flanagan [17] has established self-organization for the $2r$-neighbor models with a strictly decreasing neighborhood function.

2.2.2. *Stability*

With the notation above, let $(\mathcal{S})$ denote the following assumption:

$$(\sigma(1) < 1 \text{ and } \ell \geq 2) \quad \text{or} \quad (\sigma(2) < 1 \text{ and } \ell \geq 3) \quad \text{or} \quad (\sigma(3) < 1 \text{ and } \ell \geq 5). \tag{2.4}$$

In two papers in 1993 and 1995 ([18], [20]), Jean-Claude Fort and Gilles Pagés found that assuming equation 2.1, condition $(\mathcal{S})$ and that $X_0 \in F^+$, then

Conditions on μ	Result
$\mu = \lambda$	h^σ has a unique equilibrium x^* in F^+ and $X_t \to x^*$ a.s.
$\text{supp}(\mu) = [0,1]$ and diffuse	h^σ has at least one equilibrium point x^* in F^+
μ has a continuous density f on [0,1] and either (i) $\log f$ is concave on [0,1] and $f(0) + f(1) > 0$ or (ii) $\log f$ is strictly concave on [0,1]	if x^* is an equilibrium point then $x^* \in F^+$ and $X_t \to x^*$ conditionally a.s.

2.3. Multidimensional Models

2.3.1. *Self-Organization*

In [19], Fort and Pagés discuss the self-organization of the Kohonen string (and grid), showing it has no 'strong self-organization' property, i.e., there does not exist an absorbing set such as Υ_m in the one-dimensional case. They also give a definition of 'weak self-organization' as well as mention Flanagan's [17] 'temporary self-organizing property.'

2.3.2. *Stability*

Let $K(I^d, \mu, \sigma)$ represent the Kohonen string on the d-dimensional unit cube with environmental measure μ on I^d and neighborhood function σ. In [20] there is the following

Theorem 2 *Let $K(I^2, \mu, \sigma)$ be the Kohonen string on I^2 where $\mu = \mu_1 \otimes \mu_2$, $\sigma = \sigma_1 \otimes \sigma_2$, and μ_1 and μ_2 have continuous densities on $[0,1]$. Assume that x^{1*} is a stable equilibrium for $K(I, \mu_1, \sigma_1)$. Let $x^{2*} = \int_0^1 \omega\mu(d\omega) \in [0,1]$. then there is some $\eta > 0$ such that $\text{var}(\mu_2) < \eta$ implies $x^* = x^{1*} \otimes x^{2*}$ is a stable equilibrium of $K(I^2, \mu, \sigma)$.*

3. BACKGROUND

As an aid to understanding the following sections we mention some definitions and results from probability. We begin with a probability space $(\Omega, \mathcal{F}, \boldsymbol{P})$, where Ω is a space, $\mathcal{F}$ is a σ-algebra on Ω and $\boldsymbol{P}$ is a finite, positive measure, $\boldsymbol{P} : \mathcal{F} \rightarrow [0,1]$ such that $\boldsymbol{P}(\Omega) = 1$. In probability, a σ-algebra is called a **σ-field** and a set $A \in \mathcal{F}$ is called an **event.** We will assume that all events are in $\mathcal{F}$ unless otherwise stated. A **random variable** Y is an S-valued function on Ω measurable with respect to F i.e. $Y : \Omega \rightarrow S$ and $Y^{-1}(E) \in \mathcal{F}$ for $E \in \mathcal{B}(S)$ where $\mathcal{B}(S)$ is the Borel σ-algebra on S and S is assumed to be Polish i.e. a complete, separable metric space. Typically, $S = \boldsymbol{R}^k$ for some k. The measure μ **induced by** Y or, more commonly, the **distribution** of Y, is defined by $\mu(E) = \boldsymbol{P}(\{\omega : Y(\omega) \in E\}) = \boldsymbol{P}(Y \in E)$ where in the last term we have followed the standard practice of suppressing the independent variable ω. The notation $\boldsymbol{P} \circ Y^{-1}$ is also used for the distribution of Y. Note that μ is a probability measure on the range of Y. We denote the σ-field determined by Y by $\sigma(Y) = \{Y^{-1}(E) : E \in \mathcal{B}(S)\}$. Similarly, $\sigma(Y_1, \ldots, Y_2)$ is the smallest σ-field containing $\sigma(Y_1), \ldots, \sigma(Y_n)$.

The probabilists' term for a characteristic function of a set A, χ_A, is the **indicator function** for the set A, $\mathbf{1}_A$. Probabilists reserve **characteristic function** (of a distribution) for the Fourier transform (of a measure). Almost every(where) (a.e.) translates to **almost sure(ly) (a.s.)** or " **with probability one**".

The initials '**i.o.**' mean 'infinitely often.' In the definition,

$$\{A_n \ i.o.\} := \limsup_n A_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k,$$

they refer to the fact that if $\omega \in \{A_n\ i.o.\}$ then ω is in infinitely many of the A_k.

$\mathcal{L}$ is a **λ-system** if (i) $\Omega \in \mathcal{L}$. (ii) If $A, B \in \mathcal{L}$ and $A \subset B$, then $B - A \in \mathcal{L}$. (iii) If $A_n \in \mathcal{L}$ and $A_n \uparrow A$, then $A \in \mathcal{L}$. $\mathcal{P}$ is a **π-system** if it is closed under intersections and contains Ω.

Theorem 3 (Dynkin's π–λ Theorem) *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system containing $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Proof: See [3].

3.1. Expectation and Independence

Independence is one of the fundamental concepts in probability. Two sets $A,\ B \in \mathcal{F}$ are independent if $\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\boldsymbol{P}(B)$. Two σ-fields, $\mathcal{F}$ and $\mathcal{G}$, are independent if $\boldsymbol{P}(F \cap G) = \boldsymbol{P}(F)\boldsymbol{P}(G)$ for all $F \in \mathcal{F},\ G \in \mathcal{G}$. Two random variables, Y and Z, are independent if $\sigma(Y)$ and $\sigma(Z)$ are. A finite collection of σ-fields, $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$, is independent if $\boldsymbol{P}(\cap_{i=1}^n F_i) = \prod_{i=1}^n \boldsymbol{P}(F_i)$ for all $F_i \in \mathcal{F}_i$. An infinite collection of σ-fields is independent if every finite sub-collection is. Similarly, an infinite collection of random variables is independent if their associated σ-fields are.

The **expectation** or **expected value** of a random variable X is denoted by $\boldsymbol{E}[X]$ and defined as $\boldsymbol{E}[X] = \int X(\omega)\boldsymbol{P}(d\omega) = \int X d\boldsymbol{P}$, where the integral is taken over the space Ω. An important consequence of independence is

$$\boldsymbol{E}[XY] = \boldsymbol{E}[X]\boldsymbol{E}[Y]. \tag{3.1}$$

for independent X and Y. Note that if $X = \mathbf{1}_A$ and $Y = \mathbf{1}_B$ where A and B are independent then

$$\boldsymbol{E}[XY] = \boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\boldsymbol{P}(B) = \boldsymbol{E}[X]\boldsymbol{E}[Y]$$

If X and Y are independent simple functions then, by the latter, 3.1 again holds. By the standard arguments this result can be extended to nonnegative and then to integrable independent random variables X and Y.

If $\mu = \boldsymbol{P} \circ X^{-1}$ then by applying the arguments just used one can prove a change of variables formula

$$\boldsymbol{E}[f(X)] = \int_S f(x)\mu(dx) \tag{3.2}$$

where $f : S \to S$ is measurable.

As an application of the π–λ theorem, we have

Theorem 4 *If $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are independent and each $\mathcal{A}_i$ is a π-system, then $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), \ldots, \sigma(\mathcal{A}_n)$ are independent.*

Proof: See [16].

Corollary 1 *In order for $X_1, X_2, \ldots, X_n$ to be independent, it is sufficient that*

$$\boldsymbol{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} \boldsymbol{P}(X_i \leq x_i)$$

for all $x_1, \ldots, x_n \in (-\infty, \infty]$.

Proof: Let $\mathcal{A}_i = \{\{X_i \leq x_i\} : x_i \in (-\infty, \infty]\}$. Then $\mathcal{A}_i$ is a π-system and $\sigma(\mathcal{A}_i) = \sigma(X_i)$ and the result follows from Theorem 4.■

The **rth moment**, m_k, of a random variable X is defined by $m_k = \boldsymbol{E}[X^k]$.

3.2. Conditional Expectation and Conditional Probability

Definition 3 *Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-field and X be an integrable random variable, then the* **conditional expectation of X given $\mathcal{G}$.**, *$\boldsymbol{E}[X|\mathcal{G}]$, is a random variable such that (i) $\boldsymbol{E}[X|\mathcal{G}]$ is measurable $\mathcal{G}$. (ii) $\int_G \boldsymbol{E}[X|\mathcal{G}]d\boldsymbol{P} = \int_G X d\boldsymbol{P}$. Any random variable satisfying (i) and (ii) is called* version *of $\boldsymbol{E}[X|\mathcal{G}]$.*

Note that $\boldsymbol{E}[X|\mathcal{G}]$ exists by the Radon-Nikodym Theorem and is unique up to sets of measure zero. If we think of $\mathcal{G}$ as information then $\boldsymbol{E}[X|\mathcal{G}]$ is the best approximation to X given $\mathcal{G}$. If $\mathcal{G}$ is 'coarser' than $\mathcal{F}$ then $\boldsymbol{E}[X|\mathcal{G}]$ has possibly less 'detail' than X, is a 'smoothing' of X.

Examples: (1) If $\mathcal{G} = \{\Omega, \emptyset\}$ then the only measurable functions with respect to $\mathcal{G}$ are constants. And $\int_\Omega \boldsymbol{E}[X|\mathcal{G}]d\boldsymbol{P} = \int_\Omega X d\boldsymbol{P} = \boldsymbol{E}[X] = \int_\Omega \boldsymbol{E}[X]d\boldsymbol{P}$. Thus, $\boldsymbol{E}[X|\mathcal{G}] = \boldsymbol{E}[X]$.

(2) If $\mathcal{G} = \mathcal{F}$ then $\boldsymbol{E}[X|\mathcal{G}] = X$.

(3) If $\mathcal{G}$ is generated by a countable partition of Ω, $G_1, G_2, \ldots$, then $\boldsymbol{E}[X|\mathcal{G}]$ must be constant on each G_i.

$$\int_{G_i} \boldsymbol{E}[X|\mathcal{G}]d\boldsymbol{P} = \boldsymbol{E}[X|\mathcal{G}](\omega)\boldsymbol{P}(G_i) = \int_{G_i} X d\boldsymbol{P} \qquad \text{for} \quad \omega \in G_i$$

Thus,

$$\boldsymbol{E}[X|\mathcal{G}](\omega) = \sum_i \left(\frac{1}{\boldsymbol{P}(G_i)} \int_{G_i} X d\boldsymbol{P} \right) \mathbf{1}_{G_i}(\omega)$$

Conditional expectations enjoy many nice properties:

Theorem 5 *Let X, Y, $\{X_n\}$ be integrable.*

(i) *If $X = a$ a.s. then $\boldsymbol{E}[X|\mathcal{G}] = a$.*

(ii) (Linearity) *For constants a and b, $\boldsymbol{E}[aX + bY|\mathcal{G}] = a\boldsymbol{E}[X|\mathcal{G}] + b\boldsymbol{E}[Y|\mathcal{G}]$.*

(iii) *If $X \leq Y$ a.s., then $\boldsymbol{E}[X|\mathcal{G}] \leq \boldsymbol{E}[Y|\mathcal{G}]$.*

(iv) $|\boldsymbol{E}[X|\mathcal{G}]| \leq \boldsymbol{E}[|X||\mathcal{G}]$.

(v) (Dominated Convergence Theorem for Conditional Expectations)

If $\lim_{n\to\infty} X_n = X$ a.s., $|X_n| \leq Y$ for all n and Y is integrable then $\lim_{n\to\infty} \boldsymbol{E}[X_n|\mathcal{G}] = \boldsymbol{E}[X|\mathcal{G}]$.

Proof: See [3], p. 468.

Remark: The above proofs are straightforward using the definition and standard results from analysis.

Two additional properties that are used frequently follow.

Theorem 6 *If X is measurable $\mathcal{G}$ and X and XY are integrable, then*

$$\boldsymbol{E}[XY|\mathcal{G}] = X\boldsymbol{E}[Y|\mathcal{G}] \tag{3.3}$$

almost surely.

Proof: See [3], p. 469.

Remark: A key step in the latter proof is to show that 3.3 holds for indicator functions. Let G, $G' \in \mathcal{G}$ and $X = \mathbf{1}_G$, then

$$\begin{aligned}
\int_{G'} \boldsymbol{E}[XY|\mathcal{G}]d\boldsymbol{P} &= \int_{G'} \mathbf{1}_G Y \, d\boldsymbol{P} && \text{by definition} \\
&= \int_{G' \cap G} Y \, d\boldsymbol{P} \\
&= \int_{G' \cap G} \boldsymbol{E}[Y|\mathcal{G}]d\boldsymbol{P} && \text{by definition} \\
&= \int_{G'} \mathbf{1}_G \boldsymbol{E}[Y|\mathcal{G}]d\boldsymbol{P}
\end{aligned}$$

Since the above is true for all $G' \in \mathcal{G}$, $X\boldsymbol{E}[Y|\mathcal{G}]$ must be a version of $\boldsymbol{E}[XY|\mathcal{G}]$.

Theorem 7 *If X is integrable and the σ-fields $\mathcal{G}_1$ and $\mathcal{G}_2$ satisfy $\mathcal{G}_1 \subset \mathcal{G}_2$ then*

$$\boldsymbol{E}[\boldsymbol{E}[X|\mathcal{G}_1]|\mathcal{G}_2] = \boldsymbol{E}[X|\mathcal{G}_1] = \boldsymbol{E}[\boldsymbol{E}[X|\mathcal{G}_2]|\mathcal{G}_1] \tag{3.4}$$

almost surely.

Proof: The first equation follows from 3.3 and Theorem 5 (i) since $\boldsymbol{E}[X|\mathcal{G}_1]$ is measurable $\mathcal{G}_2$. For the second equation, note that for $G_1 \in \mathcal{G}_1$,

$$\begin{aligned}
\int_{G_1} \boldsymbol{E}[\boldsymbol{E}[X|\mathcal{G}_2]|\mathcal{G}_1] \, d\boldsymbol{P} &= \int_{G_1} \boldsymbol{E}[X|\mathcal{G}_2] \, d\boldsymbol{P} && \text{by definition} \\
&= \int_{G_1} X d\boldsymbol{P} && \text{by definition} \\
&= \int_{G_1} \boldsymbol{E}[X|\mathcal{G}_1] \, d\boldsymbol{P} && \text{by definition}
\end{aligned}$$

Since the above is true for all $G_1 \in \mathcal{G}_1$, $\boldsymbol{E}[X|\mathcal{G}_1]$ must be a version of $\boldsymbol{E}[\boldsymbol{E}[X|\mathcal{G}_2]|\mathcal{G}_1]$. ∎

The **conditional probability of an event A given an event B** is denoted by $\boldsymbol{P}(A|B)$ and is defined by $\boldsymbol{P}(A|B) = \boldsymbol{P}(A \cap B)/\boldsymbol{P}(B)$ where $\boldsymbol{P}(B) \neq 0$. If $\mathcal{G} \subseteq \mathcal{F}$ then the **conditional probability of an event A given $\mathcal{G}$** is defined by $\boldsymbol{P}(A|\mathcal{G}) = \boldsymbol{E}[\mathbf{1}_A|\mathcal{G}]$. Note that $\boldsymbol{P}_B(A) := \boldsymbol{P}(A|B)$ defines a probability measure on Ω whose support is B. It is not always the case that $\boldsymbol{P}(A|\mathcal{G})(\omega)$ is a probability measure for all ω. However, in most cases of interest, it is. First we need a

Definition 4 *Let $(\Omega, \mathcal{F}, \boldsymbol{P})$ be a probability space, $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ be a measurable map, and $\mathcal{G}$ be a σ-field, $\mathcal{G} \subseteq \mathcal{F}$. Then $\mu : \Omega \times \mathcal{S} \to [0,1]$ is said to be a* **regular conditional distribution** *for X given $\mathcal{G}$ if: (i) for each $A \in \mathcal{F}$, $\mu(\cdot, A)$ is a version of $\boldsymbol{P}(X \in A|\mathcal{G})$ (ii) for almost every ω, $\mu(\omega, \cdot)$ is a probability measure on $(S, \mathcal{S})$. If $S = \Omega$ and X is the identity map, then μ is called a* **regular conditional probability**.

Theorem 8 *With the notation above, if S is Polish and $\mathcal{S} = \mathcal{B}(S)$ then regular conditional probabilities exist.*

Proof: See [16], p. 27 and p. 199.

We will follow the convention of letting $\boldsymbol{E}(Y|X_n, X_{n-1}, \ldots, X_0)$ denote $\boldsymbol{E}(Y|\sigma(X_0, X_1, \ldots, X_n))$.

3.3. Weak Convergence

Associated with a sequence of random variables $\{X_n\}$ on a probability space $(\Omega, \mathcal{F}, \boldsymbol{P})$ is the sequence of their distributions $\{\boldsymbol{P} \circ X_n^{-1}\}$. We will be interested in the convergence properties of the latter. Assume throughout this section, as before, that S is a complete, separable metric space. Let $C_b(S)$ be the set of continuous, bounded real-valued functions on S and let $\mathcal{B}(S)$ be the Borel σ-field on S. Then the dual of $C_b(S)$, $C_b^*(S)$, is (isometrically isomorphic to) the space of signed measures. The weak * topology on $C_b^*(S)$ is the topology of pointwise convergence, i.e., a net $\{\mu_\alpha\}_{\alpha \in I}$ in $C_b^*(S)$ converges to μ if $\int f d\mu_\alpha \to \int f d\mu$ for all $f \in C_b(S)$. Basis elements are of the form

$$V_\mu(f_1, f_2, \ldots, f_n; \epsilon_1, \epsilon_2, \ldots, \epsilon_n) =$$
$$\{\nu \in C_b^*(S) : \left| \int f_i d\nu - \int f_i d\mu \right| < \epsilon_i, i = 1, 2, \ldots, n\} \tag{3.5}$$

Let $\mathcal{M}(S)$ the set of probability measures on S. If all the measures in 3.5 are probability measures then these sets form the basis elements for what probabilists term the **weak topology** on $\mathcal{M}(S)$.

Definition 5 *Let $\{\mu_n\}$ and μ be probability measures on $\mathcal{B}(S)$. Then $\{\mu_n\}$* **converges weakly** *to μ—denoted by $\mu_n \Rightarrow \mu$—if*

$$\lim_{n \to \infty} \int f d\mu_n = \int f d\mu \tag{3.6}$$

for all $f \in C_b(S)$.

By Alaoglu's Theorem the unit ball in $C_b^*(S)$ is weak * compact and hence sequentially compact. But if $\mathcal{M}(S)$ is not weak * closed in the unit ball, a convergent subsequence may not converge to a probability measure. Probabilists say a sequence of probability measures $\{\mu_n\}$ **converges vaguely** to a measure μ if 3.6 holds for all continuous f such that $\lim_{|x| \to \infty} f(x) = 0$ and the restriction that μ is a probability measure is dropped. If μ is a probability measure then $\{\mu_n\}$ is called **tight**. If μ is not a probability measure then $\mu(S) < 1$ and mass is said to 'escape to infinity.'

A set A whose boundary ∂A satisfies $\mu(\partial A) = 0$ is called a **μ-continuity set**. We have the following

Theorem 9 *The following are equivalent*

(i) $\mu_n \Rightarrow \mu$.

(ii) *Equation 3.6 holds for holds for all infinitely differentiable functions vanishing outside a bounded set.*

(iii) $\limsup_n \mu_n(F) \leq \mu(F)$ *for all closed* F.

(iv) $\liminf_n \mu_n(G) \geq \mu(G)$ *for all open* G.

(v) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for all* μ*-continuity sets* A.

Proof: see [2] and [1].

3.4. Markov Chains

3.4.1. *Introduction*

A **stochastic process** is a set of indexed random variables $\{Y_t\}_{t\in I}$. Typically, $I = 0, 1, 2, \ldots$ or $I = [0, T]$ or $I = [0, \infty)$. A **Markov chain** $\{Y_n\}$ is a stochastic process with $I = 0, 1, 2, \ldots$ and

$$\boldsymbol{P}(Y_{n+1} \in B | X_n, \ldots, X_1) = \boldsymbol{P}(Y_{n+1} \in B | X_n).$$

for all $B \in \mathcal{B}(S)$. In the following, let $\mathcal{F}_{[0,n]} := \sigma(X_k, 0 \leq k \leq n)$ and $\mathcal{F}_{(n,\infty)} := \sigma(X_k, n < k < \infty)$.

Note that for Y measurable $\sigma(X_{n+1})$ this is equivalent to

$$\boldsymbol{E}[Y | X_0, X_1, \ldots, X_n] = \boldsymbol{E}[Y | X_n]. \tag{3.7}$$

Clearly, 3.7 is true for $Y = \mathbf{1}_A$ where $A \in \sigma(X_{n+1})$. By linearity of conditional expectations it is true for all simple functions that are measurable $\sigma(X_{n+1})$, etc..

In fact, more is true.

Theorem 10 *The following are equivalent*

(i) $\{X_n\}$ *is Markov.*

(ii) *Let* $A \in \mathcal{F}_{[0,n]}$, $B \in \mathcal{F}_{(n,\infty)}$. *Then, for all* n,

$$\boldsymbol{P}(A \cap B|X_n) = \boldsymbol{P}(A|X_n)\boldsymbol{P}(B|X_n) \tag{3.8}$$

(iii) *For all* $B \in \mathcal{F}_{(n,\infty)}$,

$$\boldsymbol{P}(B|\mathcal{F}_{[0,n]}) = \boldsymbol{P}(B|X_n) \tag{3.9}$$

(iv) *For all* $A \in \mathcal{F}_{[0,n]}$,

$$\boldsymbol{P}(A|\mathcal{F}_{(n,\infty)}) = \boldsymbol{P}(A|X_n) \tag{3.10}$$

Proof: To get a flavor of the reasoning involved, part of the proof is given below. For the entire proof, see [12]. Let A and B be as defined above.

(i) $\Rightarrow$ (iii): The statement is true for $B \in \sigma(X_{n+1})$. Suppose the statement is true for $B \in \sigma(X_{n+i})$, $1 \le i \le k$ and for all integrable functions that are measurable $\sigma(X_{n+i})$, $1 \le i \le k$. Let $B \in \sigma(X_{n+k+1})$, then

$$\begin{aligned}
\boldsymbol{E}[\mathbf{1}_B|\mathcal{F}_{[0,n]}] &= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_B|\mathcal{F}_{[0,n+k]}]|\mathcal{F}_{[0,n]}] && \text{by 3.4}\\
&= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_B|X_{n+k}]|\mathcal{F}_{[0,n]}] && \text{by (i)}\\
&= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_B|X_{n+k}]|X_n] && \text{since } \boldsymbol{E}[\mathbf{1}_B|X_{n+k}] \text{ is measurable } \sigma(X_{n+k})\\
&= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_B|\mathcal{F}_{[0,n+k]}]|X_n] && \text{by (i)}\\
&= \boldsymbol{E}[\mathbf{1}_B|X_n] && \text{by 3.4}
\end{aligned}$$

Now, suppose $n_2 > n_1 > n$ and $B_{n_i} \in \mathcal{F}_{n_i}$. Then

$$\begin{aligned}
\boldsymbol{P}(B_{n_1} \cap B_{n_2} | \mathcal{F}_{[0,n]}) &= \boldsymbol{E}[\mathbf{1}_{B_{n_1}} \mathbf{1}_{B_{n_2}} | \mathcal{F}_{[0,n]}] \\
&= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_{B_{n_1}} \mathbf{1}_{B_{n_2}} | \mathcal{F}_{[0,n_1]}] | \mathcal{F}_{[0,n]}] && \text{by 3.4} \\
&= \boldsymbol{E}[\mathbf{1}_{B_{n_1}} \boldsymbol{E}[\mathbf{1}_{B_{n_2}} | \mathcal{F}_{[0,n_1]}] | \mathcal{F}_{[0,n]}] && \text{by 3.3} \\
&= \boldsymbol{E}[\mathbf{1}_{B_{n_1}} \boldsymbol{E}[\mathbf{1}_{B_{n_2}} | X_{n_1}] | \mathcal{F}_{[0,n]}] && \text{by hypothesis} \\
&= \boldsymbol{E}[\mathbf{1}_{B_{n_1}} \boldsymbol{E}[\mathbf{1}_{B_{n_2}} | X_{n_1}] | X_n] && \text{since } \mathbf{1}_{B_{n_1}} \boldsymbol{E}[\mathbf{1}_{B_{n_2}} | X_{n_1}] \\
& && \text{is msble } \sigma(X_{n_1}) \\
&= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_{B_{n_1}} \mathbf{1}_{B_{n_2}} | \mathcal{F}_{[0,n_1]}] | X_n] && \text{by hypothesis} \\
&= \boldsymbol{E}[\mathbf{1}_{B_{n_1}} \mathbf{1}_{B_{n_2}} | X_n] && \text{by 3.4}
\end{aligned}$$

This can be extended to arbitrary finite intersections. Thus, the statement is true on a π-system. Since the statement holds on a λ-system, the result follows.

(ii) $\Rightarrow$ (iii): We need to show that

$$\boldsymbol{E}[\mathbf{1}_A \mathbf{1}_B] = \int_A \boldsymbol{E}[\mathbf{1}_B | X_n] d\boldsymbol{P} = \boldsymbol{E}[\mathbf{1}_A \boldsymbol{E}[\mathbf{1}_B | X_n]]$$

for all $A \in \mathcal{F}_{[0,n]}$. Starting with the last term

$$\begin{aligned}
\boldsymbol{E}[\mathbf{1}_A \boldsymbol{E}[\mathbf{1}_B | X_n]] &= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_A \boldsymbol{E}[\mathbf{1}_B | X_n] | X_n]] && \text{by 3.4} \\
&= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_A | X_n] \boldsymbol{E}[\mathbf{1}_B | X_n]] && \text{by 3.3} \\
&= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_A \mathbf{1}_B | X_n]] && \text{by (ii)} \\
&= \boldsymbol{E}[\mathbf{1}_A \mathbf{1}_B]
\end{aligned}$$

(iii) $\Rightarrow$ (ii):

$$\begin{aligned}
\boldsymbol{P}(A \cap B | X_n) &= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_A \cdot \mathbf{1}_B | \mathcal{F}_{[0,n]}] | X_n] && \text{by definition and 3.4} \\
&= \boldsymbol{E}[[\mathbf{1}_A \boldsymbol{E}[\mathbf{1}_B | \mathcal{F}_{[0,n]}] | X_n] && \text{by 3.3} \\
&= \boldsymbol{E}[[\mathbf{1}_A \boldsymbol{E}[\mathbf{1}_B | X_n]] | X_n] && \text{by hypothesis} \\
&= \boldsymbol{E}[\mathbf{1}_B | X_n] \boldsymbol{E}[\mathbf{1}_A | X_n] && \text{by 3.3} \\
&= \boldsymbol{P}(B | X_n) \boldsymbol{P}(A | X_n) && \blacksquare
\end{aligned}$$

We define the **transition probability** $\boldsymbol{p}_n(x, A)$ through

$$\boldsymbol{P}(\{X_{n+1} \in A\} \cap \{X_n \in F\}) = \int_{\{X_n \in F\}} \boldsymbol{P}(\{X_{n+1} \in A\} | \sigma(X_n)) d\boldsymbol{P}$$
$$= \int_F p_n(x, A) \boldsymbol{P} \circ X_n^{-1}(dx)$$

for all $A,\ F \in \mathcal{B}(S)$.

If π is the initial distribution then π and $\{p_n\}$ completely determine the distribution of the process. So

$$\boldsymbol{P}(X_n \in B_n, \ldots, X_0 \in B_0) = \int_{B_0} \pi(dx_0) \int_{B_1} p_0(x_0, dx_1) \ldots \int_{B_{n-1}} p_{n-1}(x_{n-1}, B_n) \tag{3.11}$$

Note that a given Markov chain determines a mapping $\Phi : \Omega \to S^\infty (S^\infty = \prod_{n=0}^\infty S_n,\ S_n = S)$ by $\Phi(\omega) = (X_0(\omega), X_1(\omega), \ldots)$ on the σ-field $\mathcal{B}(S^\infty)$. If we let $\Omega = S^\infty$ then Ω is called the **canonical probability space**. If $\omega \in \Omega$ then for $\omega = \{x_0, x_1, \ldots\}$ we let $X_n,\ n \geq 0$ be the projection mappings, i.e. $X_n(\omega) = x_n$. A point $\omega \in \Omega$ is called a **trajectory** or **path**.

For an initial distribution π, we will denote the distribution of Φ by $\boldsymbol{P}_\pi$. If $\delta_x(A) = 1$ for $x \in A$ and 0 otherwise, then we will use $\boldsymbol{P}_x$ instead of $\boldsymbol{P}_{\delta_x}$. Similarly, let $\boldsymbol{P}_Y := \boldsymbol{P}_{P \circ Y^{-1}}$.

A Markov chain determines a sequence of transition probabilities and an initial distribution. Conversely,

Proposition 1 *Let $\{p_n\}$ be a sequence of regular conditional probabilities, $p_n : S \times \mathcal{B}(S) \to [0,1]$, and let π be a probability measure on S. Then there exists a Markov chain $\{X_n\}$ with transition probabilities $\{p_n\}$ and initial distribution π.*

Proof: See [16], p. 240.

If $p_n(x, A) = p(x, A)$ for all n, that is, there is no dependence on n, then the Markov chain is called **stationary** or **homogeneous**.

From here on, we will consider only homogeneous chains. Some examples of homogeneous Markov chains are

(1) $\{X_n\}$ where the X_k are iid.

(2) Let $\{Z_n\}$ be iid and $X_n = \sum_{k=0}^{n} Z_k$ or, in general,

(3) Let $\{Z_n\}$ be iid and let $F : S \times S \to S$. Define $X_{n+1} = F(X_n, Z_{n+1})$. The latter is how the Markov chain we consider is defined.

For $A \in \mathcal{B}(S)$, we define the **n-step transition probabilities $p^{(n)}$** by

$$p^{(1)}(x, A) = p(x, A)$$
$$p^{(n+1)}(x, A) = \int p^{(n)}(y, A) p(x, dy)$$

One can show, using induction, that

$$\boldsymbol{P}_x(X_n \in A) = p^{(n)}(x, A).$$

Similarly, one can show that

$$p^{(n+m)}(x, a) = \int p^{(n)}(y, A) p^{(m)}(x, dy)$$

We will also consider so-called **skeleton** chains i.e., if $\{X_n\}$ is a Markov chain then if $n^* \in \boldsymbol{Z}^+$, $\{Y_n\} := \{X_{n^* n}\}$ is a skeleton chain. If $p(\cdot,\cdot)$ is the transition probability for $\{X_n\}$ then $p^{(n^*)}(\cdot,\cdot)$ is the transition probability for $\{Y_n\}$ i.e. $\{Y_n\}$ is also a Markov chain.

3.4.2. *The Strong Markov Property*

Suppose $\{X_n\}$ is a Markov chain. Let $\widetilde{X}_n = X_{m+n}$ for some fixed $m, n \geq 0$. Then the Markov property implies that the distribution of $\widetilde{X}_n$ is $\boldsymbol{P}_{X_m}$. There is a useful generalization of this. We will need the following

Definition 6 *A sequence of σ-fields,* $\{\mathcal{F}_n\}$, *is a* **filtration** *if* $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$.

A typical example of a filtration is given by $\mathcal{F}_n := \sigma(X_0, X_1, \dots, X_n)$. We will call this the **standard filtration.**

Definition 7 *Let* $\{X_n\}$ *be a stochastic process such that* $\sigma(X_n) \subseteq \mathcal{F}_n$ *where* $\{\mathcal{F}_n\}$ *is a filtration. A random variable* $\tau : \Omega \to \boldsymbol{N} \cup \{\infty\}$ *is a* **stopping time with respect to** $\{\mathcal{F}_n\}$ *if* $\{\tau \leq n\} \in \mathcal{F}_n$.

Theorem 11 *Let* $\{X_n\}$ *be a Markov chain and* τ *a stopping time with respect the standard filtration. Then*

$$\boldsymbol{P}((X_\tau, X_{\tau+1}, X_{\tau+2}, \dots) \in B | \mathcal{F}_\tau) = \boldsymbol{P}_{X_\tau}((X_0, X_1, X_2, \dots) \in B) \tag{3.12}$$

for $B \in \mathcal{B}(S^\infty)$.

Proof: See [7], pp. 131-2.

3.4.3. *Invariant Measures and the Doeblin Condition*

If $\{X_n\}$ is a Markov chain with transition probability $p(\cdot, \cdot)$, then ν is an **invariant probability measure** for $\{X_n\}$ if

$$\nu(A) = \int p(x, A)\nu(dx) = \boldsymbol{P}_\nu(X_1 \in A), \quad A \in \mathcal{B}(S).$$

Thus, by induction,

$$\boldsymbol{P}_\nu(X_n \in A) = \nu(A).$$

Definition 8 *A stationary Markov chain,* $\{X_n\} \subset \Lambda$, *is* (weak) **Feller** *if the map* $x \mapsto \boldsymbol{E}[f(x)]$ *is continuous for all bounded, continuous* $f : \Lambda \to \boldsymbol{R}$.

Meyn and Tweedie in *Markov Chains and Stochastic Stability* [29] develop a number of conditions to guarantee the existence of a finite or σ-finite invariant measure. We use the following

Theorem 12 *If the homogeneous Markov chain $\{X_n\}$ is Feller and there exists a compact set A such that*

$$\boldsymbol{P}_x(\{X_n\} \text{ enters } A) = 1 \quad \forall x \in S \tag{3.13}$$

then the chain has at least one invariant measure which is finite on compact sets.

Proof: See [29] pp. 294-6.

Another way to prove a Markov chain has an invariant probability measure is to show that it satisfies the Doeblin condition.

Definition 9 *Let $\{X_n\}$ be a $(S, \mathcal{B}(S))$-valued homogeneous Markov chain with transition probability $p(\cdot, \cdot)$. If there exists a non-negative measure χ on $(S, \mathcal{B}(S))$, $n_0 \geq 1$, $c > 0$, $C \in \mathcal{B}(S)$ such that (i) $\chi(C) > 0$ and (ii) $\forall x \in S$, $\forall A \in \mathcal{B}(S)$, $A \subset C$ $\boldsymbol{P}_x(X_{n_0} \in A) \geq c\chi(A)$. then $\{X_n\}$ is said to satisfy the* **Doeblin condition** *or to be* **Doeblin recurrent**.

Theorem 13 *Let $\{X_n\}$ be as before. If $\{X_n\}$ satisfies the Doeblin condition then it admits a unique invariant probability measure ν such that (i) $\nu(A) \geq c\chi(A)$* for every *$A \in \mathcal{B}(S) \cap C$, and (ii) $\forall n \geq 1$, $\forall x \in S$, $\forall A \in \mathcal{B}(S)$ $|\boldsymbol{P}_x(X_n \in A) - \nu(A)| \leq (1 - c\chi(C))^{(n/n_0)-1}$.*

Proof: See [14].

3.4.4. *Construction of an Invariant Probability Measure*

We give the details of the construction of the probability measure given in section 5.3.2.

Suppose that the Markov chain $\{X_n\}$ satisfies the conditions of Theorem 12, with π is the invariant measure and A the compact set in equation 3.13. Let

$$S_A = \inf\{n \geq 0 : X_n \in A\}$$

$$T_A = \inf\{n > S_A : X_n \in A\}$$

where $T_A := T_A^1$ and assume

$$\boldsymbol{E}_x T_A < \infty \quad \forall\, x \in S.$$

Note that

$$\boldsymbol{E}_x T_A = \sum_{n=1}^{\infty} n \boldsymbol{P}_x(T_A = n) \tag{3.14}$$

$$= \boldsymbol{P}_x(T_A = 1)$$

$$+\boldsymbol{P}_x(T_A = 2) + \boldsymbol{P}_x(T_A = 2) \tag{3.15}$$

$$+\boldsymbol{P}_x(T_A = 3) + \boldsymbol{P}_x(T_A = 3) + \boldsymbol{P}_x(T_A = 3)$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \ddots \tag{3.16}$$

$$= \sum_{n=0}^{\infty} \boldsymbol{P}_x(T_A > n)$$

Define π_A by

$$\pi_A(C) := \frac{\pi(A \cap C)}{\pi(A)}$$

for measurable C.

Suppose $C \in \mathcal{B}(S)$, then since π is invariant for $\{X_n\}$,

$$\boldsymbol{P}_\pi(X_1 \in C) = \int p(x, C)\pi(dx) = \pi(C)$$

where p is the transition probability for $\{X_n\}$. By induction then, since $\{X_n\}$ is homogeneous,

$$\boldsymbol{P}_\pi(X_n \in C) = \pi(C).$$

Consequently, if T is a stopping time with respect to the standard filtration, then

$$\boldsymbol{P}_{\pi}(X_T \in C) = \pi(C).$$

If $C \subseteq A$, then the above gives

$$\boldsymbol{P}_{\pi_A}(X_{T_A} \in C) = \pi_A(C). \tag{3.17}$$

Define a measure,

$$\bar{\mu}(C) = \sum_{n=1}^{\infty} \boldsymbol{P}_{\pi_A}(X_n \in C, T_A > n), \qquad C \subseteq S \setminus A.$$

Observe the similarity between the latter and equation 3.14. Apparently, $\bar{\mu}$ is finite.

Now, for measurable C, let

$$\bar{\nu}(C) = \pi_A(C) + \bar{\mu}(C \cap S \setminus A)$$

To show that $\bar{\nu}$ is invariant for $\{X_n\}$ we consider two cases. First let $C \subseteq A$, then

$$\begin{aligned}
&\int p(x,C)\bar{\nu}(dx) = \\
&\quad \int_A p(x,C)\pi(dx) + \int_{A^c} p(x,C) \sum_{n=1}^{\infty} \int_{A^c} \cdots \int_{A^c} \int_A \pi(dx_0)p(x_0,dx_1)\cdots p(x_{n-1},dx) \\
&= \int_A p(x,C)\pi(dx) + \sum_{n=1}^{\infty} \int_{A^c} \int_{A^c} \cdots \int_{A^c} \int_A \pi(dx_0)p(x_0,dx_1)\cdots p(x_{n-1},dx)p(x,C) \\
&= \boldsymbol{P}_{\pi_A}(X_1 \in C, T_A = 1) + \sum_{n=1}^{\infty} \boldsymbol{P}_{\pi_A}(X_{n+1} \in C, T_A = n+1) \\
&= \sum_{n=1}^{\infty} \boldsymbol{P}_{\pi_A}(X_n \in C, T_A = n) \\
&= \boldsymbol{P}_{\pi_A}(X_{T_A} \in C) \\
&= \pi_A(C) \\
&= \bar{\nu}(C)
\end{aligned}$$

using Tonelli's theorem, equation 3.17 and the definition of $\bar{\nu}(C)$.

For $C \subseteq S \setminus A$,

$$\int p(x,C)\bar{\nu}(dx) =$$

$$\int_A p(x,C)\pi(dx) + \int_{A^c} p(x,C) \sum_{n=1}^{\infty} \int_{A^c} \cdots \int_{A^c} \int_A \pi(dx_0)p(x_0,dx_1)\cdots p(x_{n-1},dx)$$

$$= \boldsymbol{P}_{\pi_A}(X_1 \in C, T_A > 1) + \sum_{n=1}^{\infty} \boldsymbol{P}_{\pi_A}(X_{n+1} \in C, T_A > n+1)$$

$$= \sum_{n=1}^{\infty} \boldsymbol{P}_{\pi_A}(X_n \in C, T_A > n)$$

$$= \bar{\mu}(C)$$

$$= \bar{\nu}(C).$$

Thus, $\bar{\nu}$ is invariant for $\{X_n\}$. Finally, note that

$$\begin{aligned}
\bar{\nu}(S) &= \pi_A(S) + \bar{\mu}(S) \\
&= \pi_A(A) + \bar{\mu}(S \setminus A) \\
&= 1 + \sum_{n=1}^{\infty} \boldsymbol{P}_{\pi_A}(X_n \in S \setminus A, T_A > n) \\
&= \boldsymbol{P}_{\pi_A}(T_A > 0) + \sum_{n=1}^{\infty} \boldsymbol{P}_{\pi_A}(T_A > n) \\
&= \sum_{n=0}^{\infty} \boldsymbol{P}_{\pi_A}(T_A > n) \\
&= \boldsymbol{E}T_A.
\end{aligned}$$

So let

$$\nu = \frac{\bar{\nu}}{\boldsymbol{E}_{\pi_A} T_A}$$

be the probability measure.

3.5. Moment Generating Functions

What analysts term the Laplace transform of a measure μ, probabilists call the **moment generating function (m.g.f.)** M_X of the random variable X whose distribution is μ, i.e., $M_X(s) = \boldsymbol{E}[e^{sX}]$. If $S = \boldsymbol{R}$ then

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx}\mu(dx)$$

when it exists. If X is discrete, for example, $X = 1, 2, \ldots$, then

$$M_X(s) = \sum_{n=1}^{\infty} e^{sn} p_n$$

where $p_n = \boldsymbol{P}(X = n)$.

If $s \leq 0$ then $M_X(s)$ exists. If $M_X(s_0)$ is defined for $s_0 > 0$ then it is defined for $s < s_0$. If the latter is the case then M_X is defined on a neighborhood of 0 and for s in that neighborhood

$$M_X(s) = \sum_{k=0}^{\infty} \frac{s^k}{k!}\boldsymbol{E}[X^k].$$

(applying the Dominated Convergence Theorem to the fact that

$$\sum_{k=0}^{n} \frac{s^k}{k!}\boldsymbol{E}[X^k] < e^{|sX|} < e^{-sX} + e^{sX}$$

and the latter is integrable $\boldsymbol{P}$ for the given s (see [3], pp. 285-6)).

Thus from Taylor theory

$$M_X^{(k)}(0) = \boldsymbol{E}[X^k] = m_k.$$

Part of the usefulness of moment generating functions derives from

Theorem 14 *Let μ be a probability measure on* $\mathbf{R}$ *having finite moments $m_k = \int_{-\infty}^{\infty} x^k \mu(dx)$ of all orders. If the power series $\sum_{k=0}^{\infty} m_k r^k/k!$ has positive radius of convergence, then μ is the only probability measure with the moments $m_1, m_2, \ldots$.*

Proof: See [3], pp. 406-407.

Thus, if M_X is defined on a neighborhood of 0 then it corresponds to one and only one distribution (up to sets of measure zero).

3.6. Geometric Boundedness and Stochastic Domination

Definition 10 *A random variable* $Z : \Omega \rightarrow \boldsymbol{Z}^+$ *is called* **geometric with parameter** $\boldsymbol{p}$ *if* $\boldsymbol{P}(Z = n) = (1-p)^{n-1}p$.

We think of a geometric random variable as giving the number of the trial when the first 'success' occurs where all the trials are independent (so the probability of a 'failure' is $1 - p$). For example, Z could represent the number of the coin toss in which the first head occurs where the coin tosses are independent and the probability of getting a head is p.

Definition 11 *A random variable* Y *is* **lag geometric** *if* $Y = C + Z$ *where* C *is a positive constant and* Z *is a geometric random variable.*

Definition 12 *A random variable* $Y : \Omega \rightarrow \boldsymbol{Z}^+$ *is* **geometrically bounded** *if there exists constants* $a > 0,\ 0 < b < 1$ *such that* $\boldsymbol{P}(Y = n) \leq ab^n$.

Let $\mathcal{M}(S)$ be the set of probability measures on S and, in addition to S being Polish, let S be endowed with a closed partial ordering ($\leq$). As usual we assume all sets are in $\mathcal{B}(S)$ and all functions are measurable $\mathcal{B}(S)$. Let $\mathcal{I}^*(S)$ be the set of all bounded, increasing (i.e. $x \leq y \Rightarrow f(x) \leq f(y)$) real-valued functions on S and $\mathcal{I}(S)$ be the family of sets $A \in S$ such that $\mathbf{1}_A$ is increasing. Equivalently,

$$A \in \mathcal{I}(S) \text{ iff } x \in A,\ x \leq y \Rightarrow y \in A.$$

Definition 13 *Let* $\mu_1,\ \mu_2 \in \mathcal{M}(S)$. *We say* μ_1 *is* **stochastically smaller** *than* μ_2 *and write* $\mu_1 \prec \mu_2$ *iff* $\int f d\mu_1 \leq \int f d\mu_2$ *for all* $f \in \mathcal{I}^*(S)$. *If, for an underlying* *probability space* $(\Omega, \mathcal{F}, \boldsymbol{P})$, $\mu_i = \boldsymbol{P} \circ X_i^{-1}$, $i = 1, 2$ *where* X_1, X_2 *are random variables, then we say* X_1 *is* **stochastically dominated** *by* X_2.

A simple approximation argument shows that

$$\mu_1 \prec \mu_2 \Leftrightarrow \mu_1(A) \leq \mu_2(A) \quad \forall A \in \mathcal{I}(S).$$

Observe the following

Lemma 1 *If* X *is a random variable then the following are equivalent:*

1.) X is stochastically dominated by Y where Y is lag geometric.

2.) $\boldsymbol{P}(X > n) \leq ke^{-n\beta}$

3.) $\boldsymbol{P}(X = n) \leq k'e^{-n\beta'}$

4.) $\boldsymbol{P}(X > nM) \leq k''e^{-n\beta''}$

for positive M *and appropriate, positive* k, k', k'', β, β', β''.

Proof: $1 \Rightarrow 2$: Let $Y = C + Z$ where $C > 0$ and Z is geometric with parameter p. Then, by hypothesis,

$$\begin{aligned}
\boldsymbol{P}(X > n) &\leq \boldsymbol{P}(Y > n) \\
&= \boldsymbol{P}(C + Z > n) \\
&= \boldsymbol{P}(Z > n - C) \\
&= \sum_{i=\lceil n-C \rceil}^{\infty} (1-p)^{i-1} p \\
&= (1-p)^{\lceil n-C \rceil - 1} \\
&\leq (1-p)^{n-C}
\end{aligned}$$

So let $\beta = -\ln(1-p)$ and $k = \max\{1, (1-p)^{-C}\}$.

$2 \Rightarrow 1$: Defining Y as above, then $\boldsymbol{P}(Y > n) \leq (1-p)^{n-C}$. So choose p such that

$0 < p < 1 - e^{-\beta}$ and C such that $k \le (1-p)^{-C}$, i.e., $C \ge -\ln k/\ln(1-p)$.

$3 \Rightarrow 2$:

$$\boldsymbol{P}(X > n) \le \sum_{i=n+1}^{\infty} k' e^{-\beta' i} = \frac{k e^{-\beta'}}{1 - e^{-\beta'}} e^{-\beta' n}$$

$2 \Rightarrow 3$: $\boldsymbol{P}(X = n) \le \boldsymbol{P}(X > n-1) = k e^{\beta} e^{-n\beta}$

$3 \Rightarrow 4$: Clear.

$4 \Rightarrow 3$:

$$\boldsymbol{P}(X > n) = \boldsymbol{P}(X = \frac{n}{M} M) \le k'' e^{-n \frac{\beta}{M}''}$$

∎

A useful fact about geometrically bounded random variables is

Lemma 2 *If Y is geometrically bounded then Y is almost surely finite.*

Proof: We look at the complement—the event that Y is infinite. Assume $a > 0$, $0 < b < 1$ and $\boldsymbol{P}(Y > n) \le ab^n$.

$$\begin{aligned} \boldsymbol{P}(Y = \infty) &= \boldsymbol{P}(\cap_{n=0}^{\infty}\{Y > n\}) \\ &= \lim_{n\to\infty} \boldsymbol{P}(\{Y > n\}) \\ &\le \lim_{n\to\infty} ab^n = 0 \end{aligned}$$

where we have used continuity of measures from above (since $\boldsymbol{P}$ is finite.) Thus, $\boldsymbol{P}(Y < \infty) = 1$. ∎

In what follows we will have occasion to use some facts about geometrically bounded random variables. Clearly, a geometric random variable is geometrically bounded. Certain sums of geometric random variables are also geometrically bounded.

Lemma 3 *Let $\{Z_n\}$ be an iid sequence of geometric random variables with parameter p and L be a geometric random variable with parameter q where L is independent of the $\{Z_n\}$. Then $W := \sum_{n=1}^{L} Z_n$ is geometric with parameter pq.*

Proof: Let M be the moment generating function for Z_k then

$$\begin{aligned} M(s) = \boldsymbol{E}[e^{sZ_k}] &= \sum_{n=1}^{\infty} e^{sn}\boldsymbol{P}(X = n) \\ &= \sum_{n=1}^{\infty} e^{sn}(1-p)^{n-1}p \\ &= \frac{pe^s}{1-(1-p)e^s} \end{aligned}$$

So, if M_W is the moment generating function for W, then

$$\begin{aligned} M_W(s) = \boldsymbol{E}[e^{s\sum_{k=1}^{L} Z_k}] &= \boldsymbol{E}[\boldsymbol{E}[e^{s\sum_{k=1}^{n} Z_k}|L]] \\ &= \sum_{k=1}^{\infty}(\frac{pe^s}{1-(1-p)e^s})^k \boldsymbol{P}(L = k) \\ &= \sum_{k=1}^{\infty}(\frac{pe^s}{1-(1-p)e^s})^k (1-q)^{k-1}q \\ &= \frac{pqe^s}{1-(1-p)e^s} \cdot \frac{1}{1-\frac{pe^s(1-q)}{1-(1-p)e^s}} \\ &= \frac{pqe^s}{1-(1-pq)e^s} \end{aligned}$$

Hence W is a geometric random variable with parameter pq by Theorem 14 and the results of Section 3.5. ∎

Also, we have

Lemma 4 *If X and Y are independent and geometrically bounded, then $X + Y$ is geometrically bounded.*

Proof: Since X and Y are geometrically bounded, there exist constants $a_i > 0$, $0 < b_i < 1$, $i = 1, 2$ such that $\boldsymbol{P}(X = n) \leq a_1 b_1^n$ and $\boldsymbol{P}(Y = n) \leq a_2 b_2^n$. Let $b = \max\{b_1, b_2\}$. Calculating,

$$
\begin{aligned}
\boldsymbol{P}(X+Y=n) &= \sum_{k=0}^{n} \boldsymbol{P}(X+Y=n|X=k)\boldsymbol{P}(X=n) \\
&= \sum_{k=0}^{n} \boldsymbol{P}(Y=n-k|X=k)\boldsymbol{P}(X=n) \\
&= \sum_{k=0}^{n} \boldsymbol{P}(Y=n-k)\boldsymbol{P}(X=n) \quad \text{by independence} \\
&= \sum_{k=0}^{n} a_2 b_2^{n-k} a_1 b_1^k \\
&\leq a_1 a_2 \sum_{k=0}^{n} b^n = a_1 a_2 (n+1) b^n = a_1 a_2 (n+1) b^{\frac{n}{2}} b^{\frac{n}{2}} \\
&\leq \frac{a_1 a_2 2 b^{\frac{1}{2}}}{-e \ln b} b^{\frac{n}{2}}
\end{aligned}
$$

■

Clearly, then, any finite sum of independent, geometrically bounded random variables is geometrically bounded.

From the above, we can prove

Lemma 5 *Let $\{Y_n\}$ and L' be iid, lag geometric random variables and let L' be lag geometric independent of the $\{Y_n\}$. If $W = \sum_{k=0}^{L'} Y_k$, then W is geometrically bounded.*

Proof: Let $Y_n = C + Z_n$ where $C \in \boldsymbol{Z}^+$ and $\{Z_n\}$ are iid geometric. Let $L' = C' + L$ where $C' \in \boldsymbol{Z}^+$ and L is geometric. Then

$$
\begin{aligned}
W = \sum_{k=0}^{L'} Y_k &= \sum_{k=0}^{C'+L} (C + Z_k) \\
&= CC' + CL + \sum_{k=0}^{C'} Z_k + \sum_{k=0}^{L} Z_k
\end{aligned}
$$

By Lemma 3 the last term is geometrically bounded. By Lemma 4 the third term is geometrically bounded. The second term is geometrically bounded by noting that $\boldsymbol{P}(CL > n) = \boldsymbol{P}(L > n/C)$. Consequently, by Lemma 4 and Lemma 1, W is geometrically bounded. ■

If S_1 and S_2 are Polish spaces with their respective Borel σ-fields, then a **stochastic kernel** in $S_1 \times S_2$ is a function $k : S_1 \times \mathcal{B}(S_2) \to [0,1]$ such that $k(\cdot, A)$ is measurable for each $A \in \mathcal{B}(S_2)$ and $k(x, \cdot) \in \mathcal{M}(S_2)$ for each $x \in S_1$. For such a kernel k and $\mu_1 \in \mathcal{M}(S_1)$, denote $\mu_1 * k$ as the element of $\mathcal{M}(S_1 \times S_2)$ determined by

$$(\mu_1 * k)(A_1 \times A_2) = \int_{A_1} k(x, A_2)\mu_1(dx).$$

Denote the second marginal distribution of $\mu_1 * k$ by μ_1^k. A stochastic kernel k on $S_1 \times S_2$ is said to be **stochastically monotonic** if $k(x, \cdot) \prec k(y, \cdot)$ for all $x \le y$. A stochastic kernel is called "upward" if for all x, $k(x, \cdot)$ is a measure with support in $\{y \in S : y | y \ge x\}$.

We have the following fundamental

Theorem 15 *The following conditions are equivalent for* $\mu_1, \mu_2 \in \mathcal{M}(S)$:

(i) $\mu_1 \prec \mu_2$;

(ii) *there exists a* $\lambda \in \mathcal{M}(S \times S)$ *with support in* $K = \{(x,y) \in S \times S : x \le y\}$ *with first marginal* μ_1 *and second marginal* μ_2;

(iii) *there exists a real-valued random variable* Z *and two measurable functions* f *and* $g : \boldsymbol{R} \to S$ *with* $f \le g$ *such that the distribution of* $f(Z)$ *is* μ_1 *and the distribution of* $g(Z)$ *is* μ_2;

(iv) *there exist two* S*-valued random variables* X_1, X_2 *such that* $X_1 \le X_2$ *a.s. and the distribution of* X_i *is* μ_i *for* $i = 1, 2$;

(v) *there exists an upward kernel* k *on* $S \times S$ *such that* $\mu_2 = \mu_1^k$;

(vi) $\mu_1(B) \le \mu_2(B)$ *for all closed* $B \in \mathcal{I}(S)$.

Proof: See [22].

4. SUPER-STABILITY OF MARKOV DYNAMICAL SYSTEMS

4.1. Introduction

Typically of central concern to an investigator of a process is the question of stability. The latter can have variety of meanings. These, in turn, can be applied to different aspects of the process.

In that, the Kohonen process is stochastic, there is the associated sequence of induced measures $\{\boldsymbol{P} \circ X_n^{-1}\}$ in addition to the sequence $\{X_n\}$ itself. The convergence properties of the former is usually of prime importance to a probabilist. We focus however on the dynamical stability of $\{X_n\}$.

As mentioned in Chapter 2, others have considered a related process when the shrinking (also called 'gain') parameter α is a function of time and subject to the conditions

$$\sum_{n=1}^{\infty} \alpha_n = \infty \qquad \sum_{n=1}^{\infty} \alpha_n^2 < \infty.$$

As noted they consider the stability of the ODE

$$\dot{x} = -h^{\sigma}(x_0), \quad x_0 \in \Upsilon$$

where

$$h_i^{\sigma} := \sum_{k \in V} \sigma(|k-i|) \int_{C_k(x)} (x_i - \omega)\mu(d\omega)$$

the neighborhood function is $K(i,n) = \alpha_n \sigma(i)$ and $C_i(x)_{i \in V}$ is the Voronoï tessellation.

For each α_n above there is an associated invariant measure ν_{α_n}. The behavior of $\{\nu_{\alpha_n}\}$ has been studied in [5].

4.2. Traditional Views of Stability

As stated, stability can take different forms. From a dynamical systems perspective, one is concerned with the triple (T, S, D) or the quadruple (T, S, D, μ), where S is a space, D a metric on the space, μ a measure (often a probability measure) on S, and $T : S \to S$ a (often continuous) transformation on S. The behavior of the orbits or trajectories $\{T^k x\}$ is studied where $x \in S$ and $T^k = T \circ T^{k-1}$, where $\circ$ denotes composition. In our case, the transformations $\{T_k\}$ are random and so the orbits take the form $\{T_k^k x\}$

Some traditional formulations of stability for a dynamical system are

(i) *Lagrange stability*: for each $x \in S$ the orbit starting at x is a precompact subset of S.

(ii) *Lyapunov stability*: for each initial condition $x \in S$,

$$\lim_{y \to x} \sup_{k \geq 0} D(T^k y, T^k x) = 0.$$

(iii) *Asymptotic stability*: there exists a fixed point x^* ($T^k x^* = x^*$ for all k) and a neighborhood N_{x^*} of x^* such that

$$\lim_{k \to \infty} D(T^k x, x^*) = 0 \quad \forall\, x \in N_{x^*}$$

(iv) *Global Asymptotic stability*: the system is stable in the Lyapunov sense and for some fixed $x^* \in S$

$$\lim_{k \to \infty} D(T^k x, x^*) = 0 \quad \forall\, x \in S$$

(See [29] pp. 16-21)

We now define super-stability.

4.3. Super-stability

Suppose that (Υ, d) is a bounded separable metric space. We assume, without loss of generality, that the diameter of Υ is 1. One way to define Markov chains on Υ is as a random dynamical system. Let $F : \Upsilon \times S \rightarrow \Upsilon$ be measurable. On a probability space $(\Omega, \mathcal{F}, \boldsymbol{P})$, let $Z_1, Z_2, \ldots$ be an iid sequence of random variables, $Z_k : \Omega \rightarrow S$, with distribution μ. Denote $F(\cdot, Z_n)$ by either $F_n(\cdot)$ or $F_{Z_n}(\cdot)$. A Markov chain $\{X_n\}$ may be defined by setting X_0 independent of $\{F_n\}$ to have an initial distribution and setting $X_n = F_n \circ F_{n-1} \circ \cdots \circ F_1(X_0)$ where $\circ$ denotes composition.

Subject to mild regularity conditions all Markov chains may be expressed this way (we do not assume that the transformations are continuous). See Kifer's book [23], page 8, for details. Often this is the natural way to represent the process. This is the case for the self-organizing feature maps studied here.

Definition 14 *A Markov chain given as a dynamical system with random transformation F is* super-stable *if (i) there is a constant $\beta > 0$ so that for every pair of initial random variables X_0 and Y_0 there is an $\boldsymbol{N}$-valued random variable W so that*

$$d(F_n \circ F_{n-1} \circ \cdots \circ F_1(X_0), F_n \circ F_{n-1} \circ \cdots \circ F_1(Y_0)) \leq e^{-\beta(n-W)} \tag{4.1}$$

and (ii) there is an invariant probability measure for the chain. We say the chain is super-stable with exponential rate *if in addition (iii) there are constants $A > 0$ and $0 < B < 1$ independent of X_0 and Y_0 so that $\boldsymbol{P}[W = n] \leq AB^n$ for all n.*

Compare with Lyapunov stability above. Note that this definition is in some sense dual to the property of being a chaotic attractor in which all the randomness is in the initial conditions. Here the initial conditions are irrelevant in the long term, the environment alone gives the behavior.

4.4. Super-stability vs. Doeblin

Now we compare super-stability with exponential rate with a more traditional property, the Doeblin condition. Bouton and Pagès [5] proved that a similar class of models to that we study here is Doeblin. This is a strong uniform recurrence condition which in this case is equivalent to uniform ergodicity.

Definition 15 *A Markov chain is* uniformly ergodic *if*

$$\sup_{x\in\Upsilon} \sup_{A \text{ Borel}} |\boldsymbol{P}^n(x,A) - \mu(A)| \to 0$$

Thus, uniformly in x, the distribution of $(X_n|X_0 = x)$ converges in total variation distance to the invariant distribution. On the other hand this says nothing about the dynamical system that defines the Markov chain.

Super-stability with exponential rate is a definition applied to the dynamical system. It says that the initial condition is exponentially irrelevant and that only the environment determines the (approximate) location of X_n. It does not by itself prove that the chain is Doeblin. For this to be the case, we would need an extra condition. For example, if there were an x_0 with the property that

$$\lim_{x\to x_0} ||\boldsymbol{P}(x,.) - \boldsymbol{P}(x_0,.)|| = 0$$

where the norm is the total variation distance on signed measures then super-stability with exponential rate would imply Doeblin, whence uniform ergodicity.

In the class of examples considered in this paper this holds only in the case that there is no graph structure, i.e. if $r = 0$ (see below).

Dr. Robert Burton has provided some examples that will clarify these distinctions.

Example 1: Let the underlying space be [0,1]. Define the dynamical system by $X_{n+1} = (1/2)X_n$ with probability 1/2 and $X_{n+1} = 1/2 + (1/2)X_n$ with probability 1/2. This is super-stable with exponential rate, has Lebesgue measure invariant but is not Doeblin. The rational numbers are invariant so Doeblin is impossible.

Example 2: Let the underlying space be the unit circle and the dynamical system be $X_{n+1} = (X_n + Y_n)(\text{mod } 1)$ where Y_n is iid uniform on [0,1]. Given two starting values X_0 and X_0', the distance on the circle between X_n and X_n' is constant so this example is not super-stable. However it is Doeblin with Lebesgue measure invariant.

Example 3: Let the underlying space and $X_{n+1} = (1/2)X_n + (1/2)Y_n$ where Y_n is iid uniform on [0,1] then this dynamical system is both super-stable with exponential rate and also Doeblin with Lebesgue measure invariant.

4.5. Uniqueness of the Invariant Measure

Let $\mathcal{M}$ be the set of Borel probability measures on Υ and $\mathcal{C}$ be the set of bounded, uniformly continuous real-valued functions on Υ. Define $T : \mathcal{M} \to \mathcal{M}$

by $T(\nu)$ =dist($F_Z(X)$) where ν =dist(X) and Z is distributed as $Z_1, Z_2, \ldots$. Set $\hat{T} : \mathcal{C} \to \mathcal{C}$ by $\hat{T}f(x) = \mathbf{E}(f(F_Z(x)))$. Thus $\int f dT\nu = \int \hat{T} f d\nu$ i.e. T and $\hat{T}$ are adjoints. To see this, let $B \in \mathcal{B}(\Upsilon)$, then

$$\begin{aligned} T\nu(B) = \boldsymbol{P}(F_Z(X) \in B) &= \boldsymbol{P}(X \in F_Z^{-1}(B)) \\ &= \boldsymbol{E}[\mathbf{1}_{X \in F_Z^{-1}(B)}] \\ &= \boldsymbol{E}[\boldsymbol{E}[\mathbf{1}_{X \in F_Z^{-1}(B)} | Z]] \\ &= \boldsymbol{E}[\nu \circ F_Z^{-1}(B)] \\ &= \int_S \nu \circ F_z^{-1}(B) \mu(dz) \\ &= \int_\Omega \nu \circ F_Z(\omega)^{-1}(B) \boldsymbol{P}(d\omega) \end{aligned}$$

Thus,

$$\begin{aligned} \int f dT\nu &= \int_\Omega \int_\Upsilon f(x) \nu \circ F_{Z(\omega)}^{-1}(dx) \boldsymbol{P}(d\omega) \\ &= \int_\Omega \int_\Upsilon f(F_{Z(\omega)}(x)) \nu(dx) \boldsymbol{P}(d\omega) \\ &= \int_\Upsilon \left[\int_\Omega f(F_{Z(\omega)}(x)) \boldsymbol{P}(d\omega) \right] \nu(dx) \\ &= \int \hat{T} f d\nu \end{aligned}$$

We may also put a metric on $\mathcal{M}$, the Wasserstein metric, which we define by $\rho(\nu_1, \nu_2) = \inf\{\int_{\Upsilon \times \Upsilon} d(X_1, X_2) d\eta | \text{dist}(X_i) = \nu_i, i = 1, 2;\ \eta \in \mathcal{P}(\nu_1, \nu_2)\}$ where $\mathcal{P}(\nu_1, \nu_2)$ is the set of probability measures on $\Upsilon \times \Upsilon$ with marginals ν_1 and ν_2 i.e. $\nu_1(A) = \eta(A \times \Upsilon)$, $\nu_2(A) = \eta(\Upsilon \times A)$ for $A \in \mathcal{B}(\Upsilon)$. The Kantorovich-Rubinstein Theorem together with the assumption that the metric d is bounded implies ρ is a metric which characterizes convergence in distribution, i.e.
$\rho(\nu_n, \nu) \to 0$ iff $\nu_n \to \nu$ (dist) [15], p. 329. This is a complete metric if Υ is Polish.

The following proposition appears in [10]

Proposition 2 *Suppose that a Markov chain $\{X_n\}$ defined as a dynamical system on Υ with random transformation distributed as F is super-stable with $\bar{\nu}$ an invariant probability. Then $\bar{\nu}$ is the unique invariant measure, indeed the distribution of X_n converges weakly to $\bar{\nu}$ with respect to the function class $\mathcal{C}$. If the chain is super-stable with exponential rate then this convergence is exponentially fast with respect to the Wasserstein metric.*

Proof: Let $\sup_{\{x\in\Upsilon\}}|f(x)| = M$ and $F_{Z_n(\omega)}\circ\ldots\circ F_{Z_1(\omega)}(x) := X_n(\omega)$ $F_{Z_n(\omega)}\circ\ldots\circ F_{Z_1(\omega)}(y) := Y_n(\omega)$ for $x,\ y \in \Upsilon$. Let $|f(u)-f(v)| < \epsilon$ when $d(u,v) < \delta$ and k be such that $e^{-\beta k} < \delta$. By hypothesis, $d(X_n, Y_n) \le e^{-\beta(n-W)}$ where W is as in the definition of super-stability. Note that $\lim_{n\to\infty}(\hat{T}^n f(x) - \hat{T}^n f(y)) = 0$ for

$$\hat{T}^n f(x) - \hat{T}^n f(y) = \int_\Omega (f(X_n(\omega)) - f(Y_n(\omega)))\boldsymbol{P}(d\omega)$$
$$= \int_{\{n-W\ge k\}} (f(X_n(\omega) - f(Y_n(\omega)))\boldsymbol{P}(d\omega) + \int_{\{n-W<k\}} (f(X_n(\omega)(f(Y_n(\omega)))\boldsymbol{P}(d\omega)$$
$$\le \epsilon + M\boldsymbol{P}(n-k<W)$$

The result follows since $\lim_{n\to\infty}\boldsymbol{P}(n-k<W) = 0$. Integrating with respect to $\nu(dy)$ and using the bounded convergence theorem,

$$\int(\hat{T}^n f(x) - \hat{T}^n f(y))\nu(dy) = \hat{T}^n f(x) - \int \hat{T}^n f(y)\nu(dy) \to 0$$

as $n\to\infty$. Similarly,

$$\int(\hat{T}^n f(x) - \int \hat{T}^n f(y)\nu(dy))\bar{\nu}(dx) = \int \hat{T}^n f(x)\bar{\nu}(dx) - \int \hat{T}^n f(y)\nu(dy) \to 0$$

as $n\to\infty$. Now

$$\int \hat{T}^n f(x)\bar{\nu}(dx) = \int f(x)T^n\bar{\nu}(dx) = \int f(x)\bar{\nu}(dx)$$

since $\bar{\nu}$ is invariant and

$$\int \hat{T}^n f(y)\bar{\nu}(dy) = \int f(y)T^n\bar{\nu}(dy).$$

Thus,

$$\int f(y)T^n\bar{\nu}(dy) - \int f(x)\bar{\nu}(dx) \to 0$$

as $n \to \infty$. If $T^n\nu := \nu_n$ then $\nu_n \to \bar{\nu}$ weakly since by Portmanteau's Theorem $\mathcal{C}$ is a determining class for weak convergence.

To prove the second part of the proposition, let Y be distributed as $\bar{\nu}$ and X as ν. Then,

$$\begin{aligned}
&\rho(T^n\nu, \bar{\nu}) = \rho(T^n\nu, T^n\bar{\nu}) \\
&= \inf\left\{\int_{\Upsilon\times\Upsilon} d(x_n, y_n)d\eta : \eta \in \mathcal{P}(T^n\nu, \bar{\nu})\right\} \\
&\leq \int_\Omega d(X_n, Y_n)d\boldsymbol{P}
\end{aligned}$$

by change of variables and using the fact that $\boldsymbol{P}$ determines a distribution of the random vector (X_n, Y_n) on $\Upsilon \times \Upsilon$ with appropriate marginals. Continuing,

$$\begin{aligned}
\int_\Omega d(X_n, Y_n)d\boldsymbol{P} &\leq \int_\Omega \min\{1, e^{-\beta(n-W)}\}d\boldsymbol{P} \\
&= \int_{\{W>n\}} \min\{1, e^{-\beta(n-W)}\}d\boldsymbol{P} + \int_{\{W\leq n\}} \min\{1, e^{-\beta(n-W)}\}d\boldsymbol{P} \\
&= \boldsymbol{P}(W > n) + \boldsymbol{E}[e^{-\beta(n-W)}\mathbf{1}_{\{W\leq n\}}] \\
&\leq \sum_{k=n+1}^{\infty} AB^k + \boldsymbol{E}[\boldsymbol{E}[e^{-\beta(n-W)}\mathbf{1}_{\{W\leq n\}}|W]] \\
&\leq \frac{AB^{n+1}}{1-B} + \sum_{k=1}^{n} AB^k e^{-\beta(n-k)} \\
&= \frac{AB^{n+1}}{1-B} + Ae^{-\beta n}\sum_{k=1}^{n}(e^\beta B)^k \\
&= \frac{AB^{n+1}}{1-B} + Ae^{-\beta n}\frac{1-(e^{-\beta}B)^{n+1}}{1-(e^{-\beta}B)}
\end{aligned}$$

The result follows since the latter goes to zero as n goes to infinity. ∎

5. THE ONE-DIMENSIONAL KOHONEN PROCESS

5.1. Introduction

We recall some definitions and notation from Chapter 1.

$V = \{1, \ldots, \ell\}$ is a finite vertex set and $\Upsilon' = \{X : V \to [0,1] | X$ is one-to-one$\}$. We let

$$D(X,Y) = \min\{\max_{i \in V}|X(i) - Y(i)|, \max_{i \in V}|X(i) - Y(\ell - i + 1)|\}. \tag{5.1}$$

Define Υ to be the set of equivalence classes on Υ' so as to make (Υ, d) a metric space. Υ_{m} is the set of elements of Υ that are monotonic increasing or decreasing.

The environment is expressed by d random variables $\omega_1, \omega_2, \ldots$ which have the distribution μ on $[0,1]$. There is also a kernel or neighborhood function $K : \{0, 1, \ldots, \ell\} \to [0,1]$ which is assumed to be non-increasing.

Let $X_0 \in \Upsilon$ be random with a distribution absolutely continuous with respect to Lebesgue measure. If X_n is the state of the process at time n then the state of the process at time $n+1$, X_{n+1}, is defined as follows. Let

$$i'_{n+1} := \arg\min_{i \in V}|X_n(i) - \omega_{n+1}|.$$

Let

$$X_{n+1} = F(\omega_{n+1}, X_n) \tag{5.2}$$

and define

$$F(\omega_{n+1}, X_n)(i) = K(|i'_{n+1} - i|)\omega_{n+1} + (1 - K(|i'_{n+1} - i|)X_n(i) \tag{5.3}$$

$$= X_n(i) + K(|i'_{n+1} - i|)(\omega_{n+1} - X_n(i)) \tag{5.4}$$

In the standard Kohonen process,

$$K(i) = \alpha \mathbf{1}_{\{i \leq r\}}$$

where $0 < \alpha < 1$.

5.2. The Main Results

Theorem 16 *Consider the standard Kohonen process with* $1 \leq r \leq \lfloor \frac{\ell-1}{2} \rfloor$. *Let the environmental distribution be absolutely continuous with respect to Lebesgue measure with a density function bounded and bounded away from zero. The number of particles* ℓ *and the shrinking parameter* α *with* $0 < \alpha < 1$ *are fixed. Then the Markov dynamical system* X_n *is super-stable with exponential rate.*

Theorem 17 *In the notation above suppose that the environmental distribution is absolutely continuous with respect to Lebesgue measure with a density function bounded and bounded away from zero and that the kernel function is strictly decreasing with* $0 < K(i) < 1$ *for all* i. *Then the Markov dynamical system* X_n *is super-stable with exponential rate.*

We prove Theorem 16 and use the latter to prove Theorem 17. Until the proof of Theorem 17 we assume the kernel is that of Theorem 16, $K(i) = \alpha$ for $i \leq r$ and $K(i) = 0$ for $i > r$. We will also initially assume μ is Lebesgue measure. Then we will indicate the necessary changes to accommodate more general environments. The proof proceeds through a number of lemmas.

In the following we will often be concerned with pairs of maps X and Y or pairs of sequences $\{X_n\}$ and $\{Y_n\}$. *For convenience, we will assume that*

$$D(X,Y) = \max_{i \in V} |X(i) - Y(i)|$$

or, for every n

$$D(X_n, Y_n) = \max_{i \in V} |X_n(i) - Y_n(i)|.$$

(Recall that each element of Υ is an equivalence class. If '$\sim$' is the equivalence relation, then $X \sim \widetilde{X}$ if $X(i) = \widetilde{X}(\ell - i + 1)$ for all i. The above statement says then that if

$$D(X_n, Y_n) = \max_{i \in V} |X_n(i) - Y_n(\ell - i + 1)|$$

then replace X_n by $\widetilde{X_n}$.)

We introduce some of the language we use.

Sometimes we refer to $i \in V$ as a particle, and think of $X(i)$ as its position. We can think of the process as moving a few particles at each stage.

Definition 16 *For a given map X, we say that a particle i is* hit *if*

$$|X(i) - \omega| = \min_{j \in V} |X(j) - \omega|$$

where ω is random.

Definition 17 *We say that X and Y in Υ* split *if particle i is hit for X and particle j is hit for Y with $i \neq j$. Otherwise, we say X and Y are* joint-hit.

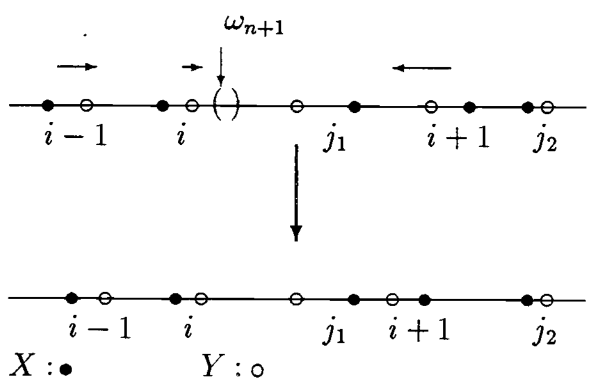


FIGURE 5.1. A joint hit of paired maps

Definition 18 *A* shrinking *is the event that $D(X, Y)$ decreases due to joint-hits.*

This would occur, for example, if all the particles were joint-hit with no splits.

Definition 19 *For two maps X, Y we say there is a* pairing at i *if*

$$|X(i) - Y(i)| < \min_{k \neq i}\{|X(k) - X(i)|, |Y(k) - Y(i)|\}. \tag{5.5}$$

We say X and Y are paired *if there is a pairing at i for all i, $1 \leq i \leq \ell$.*

We say *i is hit on the left* if ω is such that i is hit and $\omega < X(i)$. *Hit on the right* is defined similarly. In general, we say *left of* or *below i* to mean less than $X(i)$ and similarly for *right of* or *above i*. Let $B(i) = \{j \in v \| j - i| \leq r\}$ denote the *neighborhood of i*, where r is the so-called range parameter mentioned in the theorem. We say j is a *neighbor* of i if $j \in B(i)$.

5.3. The Standard Kohonen Process

We make some observations about the process:

Observation 1 *Hitting at i causes a shrinking of the distance between i and its neighbors by a factor of $(1 - \alpha)$.*

If $|X_m(i+p_1) - X_m(i+p_2)| > s$ where $0 \leq |p_1|, |p_2| \leq r$ then an induction argument shows that for n consecutive hits at i

$$(1 - \alpha)^n s < |X_{m+n}(i + p_1) - X_{m+n}(i + p_2)| < (1 - \alpha)^n. \tag{5.6}$$

From the latter we have

Observation 2 *Hits at i do not change the relative order of i and its neighbor(s).*

Consequently, we have

Observation 3 *Once a map is monotonic it remains monotonic.*

Observation 4 *If particle i is hit and $\omega \in$ convex hull of $X(B(i))$ then there is an increase in separation between the neighbors of i and non-neighbor particles above and below the neighborhood of i.*

For example, suppose $X_m(i)$ and $X_m(j)$ are adjacent and $r = 1$, $X_m(i) < X_m(j)$, $|X_m(i) - X_m(j)| > s$
> 0 and $X_m(j_1) < X_m(i-1) < X_m(i) < X_m(i+1) < X_m(j_2)$ where $j_1, j_2 \notin B(i)$. Let $\omega_{m+1} \in (X_m(i) + (s/4), X_m(i) + (s/2))$. Then ω_{m+1} hits i and

$$\begin{aligned} X_{m+1}(i+1) &= \alpha\omega_{m+1} + (1-\alpha)X_m(i+1) \\ &< \alpha(X_m(i+1) - (s/2)) + (1-\alpha)X_m(i+1) \\ &= X_m(i+1) - (\alpha s/2) \end{aligned}$$

So,

$$|X_{m+1}(j_2) - X_{m+1}(i+1)| > \alpha s/2 \tag{5.7}$$

Similarly,

$$|X_{m+1}(j_1) - X_{m+1}(i-1)| > \alpha s/4 \tag{5.8}$$

Observation 5 *If for the maps X_n, $Y_n \in \Upsilon$ there is a joint hit at i then $|X_{n+1}(i+p) - Y_{n+1}(i+p)| = (1-\alpha)|X_n(i+p) - Y_n(i+p)|$ where $|p| \leq r$.*

Observation 6 *The only way to increase the distance between maps is to have a split.*

5.3.1. *Getting to Monotonicity*

In the following, we will be interested in separations between particles and between particles and the ends of the unit interval. To this end, we define $X'_k : \{0,1,\ldots,\ell+1\} \to [0,1]$ by

$$X'_k(i) = \begin{cases} 0 & , \quad i = 0 \\ X_k(i) & , \quad i \in V \\ 1 & , \quad i = \ell+1. \end{cases}$$

Of course, particles 0 and $\ell+1$ are stationary and are not members of any neighborhood. We now prove the *Moving Lemma* in that the event it describes is useful in reordering particles.

Lemma 6 *Let* $\min\{X'_m(k) - X'_m(i) | k \neq i,\ X'_m(k) > X'_m(i)\} > s > 0$ *then there exists an integer* $N(\alpha, s)$ *and a real number* $\theta(\alpha, s) > 0$ *such that*

$$\boldsymbol{P}\{X_{m+N(\alpha,s)}(B(i)) \subset (X_m(i) + \frac{s}{8}, X_m(i) + \frac{3s}{4})\} > \theta(\alpha, s) \tag{5.9}$$

If $\min\{X'_m(i) - X'_m(k) | k \neq i,\ X'_m(i) > X'_m(k)\} > s > 0$ *then for the same* $N(\alpha, s)$ *and* $\theta(\alpha, s)$ *as above*

$$\boldsymbol{P}\{X_{m+N(\alpha,s)}(B(i)) \subset (X_m(i) - \frac{3s}{4}, X_m(i) - \frac{s}{8})\} > \theta(\alpha, s) \tag{5.10}$$

Remark: In the above, $s/8$ and $3s/4$ were chosen somewhat arbitrarily. The key facts are that (a) we can hit at particle with a known separation by it any finite number of times with a corresponding strictly positive probability, (b) $X_{m+N(\alpha,s)}\{B(i)\}$ is contained in either $(X_m(i), X_m(i) + s)$ or $(X_m(i) - s, X_m(i))$, (c) the members of $B(i)$ are adjacent and (d) there is a lower bound on the separations between the members of $B(i)$ and other particles. In particular, in the case of equation 5.9 if $j_1, j_2 \notin B(i)$ and $X_m(j_1) < X_m(i) < X_m(j_2)$ under the conditions of the lemma then

$$|X'_{m+N(\alpha,s)}(j_1) - \min_p X'_{m+N}(i+p)| > s/8,$$

$$|X'_{m+N(\alpha,s)}(j_2) - \max_p X'_{m+N}(i+p)| > s/4,$$

where $1 \le |p| \le r$.

Proof: Without loss of generality assume $m = 0$. It is sufficient to prove the case for equation 5.9. The other case follows similarly. Let $s_0 = s/4$ and

$$s_n = (1-\alpha)s_{n-1} = (1-\alpha)^n s/4, \qquad n \ge 0. \tag{5.11}$$

Let

$$\omega_{n+1} \in I_n := (X_n(i) + s_n, X_n(i) + 2s_n) \tag{5.12}$$

and $1 \le |p| \le r$. Note that

$$\begin{aligned} I_{n+1} &= (X_{n+1}(i) + s_{n+1}, X_{n+1}(i) + 2s_{n+1}) \\ &= (\alpha\omega_{n+1} + (1-\alpha)X_n(i) + (1-\alpha)s_n, \alpha\omega_{n+1} + (1-\alpha)X_n(i) + (1-\alpha)2s_n) \\ &= (\alpha\omega_{n+1} + (1-\alpha)(X_n(i) + s_n), \alpha\omega_{n+1} + (1-\alpha)(X_n(i) + 2s_n)) \\ &= \alpha\omega_{n+1} + (1-\alpha)I_n. \end{aligned} \tag{5.13}$$

Since $\omega_{n+1} \in I_n$ then I_{n+1} is a convex combination of elements of I_n. Consequently,

$$I_{n+1} \subset I_n \tag{5.14}$$

and, thus,

$$I_n \subset I_0, \quad \forall n. \tag{5.15}$$

With the above and the facts that

$$|X_n(i+p) - X_n(i)| \geq (1-\alpha)^n s. \tag{5.16}$$

from equation 5.6 and

$$\omega_{n+1} - X_n(i) < (1-\alpha)^n \frac{s}{2},$$

ω_n will always hit i while maintaining a separation greater than $s/2$ with non-neighbor particles above i. (Of course, if the hits are above i we do not have to worry about hitting particles below i since the particles maintain their relative positions.)

To get $B(i)$ adjacent and such that $X_N(i+p) \in (X_0(i) + (s/8), X_0(i) + (3s/4))$. choose N such that

$$(1-\alpha)^N < \frac{s}{16}. \tag{5.17}$$

We have then

$$N > \frac{\log(s/16)}{\log(1-\alpha)},$$

and

$$\theta = \prod_{k=0}^{N} s_k = \left(\frac{s}{4}\right)^N \prod_{k=0}^{N} (1-\alpha)^k$$

So define

$$N(\alpha, s) := \lceil \frac{\log(s/16)}{\log(1-\alpha)} \rceil + 1 \tag{5.18}$$

and

$$\theta(\alpha, s) := \prod_{k=0}^{N(s,\alpha)} s_k = \left(\frac{s}{4}\right)^{N(s,\alpha)} \prod_{k=0}^{N(s,\alpha)} (1-\alpha)^k . \blacksquare \tag{5.19}$$

Since we will be referring to the event in the Moving Lemma frequently, we denote such an event as a *moving event* with the understanding that s may vary over a sequence of such events.

Definition 20 *Let* $\{X_k\} \subseteq \Upsilon$ *be a realization of the process. Then* $\{X_k\}$ *is* stochastically $(\bar{\epsilon}, \bar{\delta})$–separated at (i, j) at time n *(or, alternately,* X_n *is* $(\bar{\epsilon}, \bar{\delta})$–*separated at* (i, j)*) if* $\mathbf{P}[|X_n'(i) - X_n'(j)| > \bar{\epsilon}] > \bar{\delta}$. $\{X_k\}$ *is* stochastically $(\bar{\epsilon}, \bar{\delta})$–separated at time n *(alternately,* X_n *is* $(\bar{\epsilon}, \bar{\delta})$*–separated) if* $\mathbf{P}[min_{i \neq j}|X_n'(i) - X_n'(j)| > \bar{\epsilon}] > \bar{\delta}$.

We will also say X_n is stochastically separated or separated or $\bar{\epsilon}$-separated at (i, j) or X_n is stochastically separated or separated or $\bar{\epsilon}$-separated where the $(\bar{\epsilon}, \bar{\delta})$ or $\bar{\delta}$ is implied or is to be taken from context. With this definition we have

Lemma 7 *Given* X_n *is stochastically* $\bar{\epsilon}$*-separated at* (k, j)*, if a moving event occurs beginning at time* $n + 1$ *then* $X_{n+N(\alpha,\bar{\epsilon})}$ *is stochastically separated at* (k, j).

Proof: Let i and θ be as in Lemma 6. Let X_n be $(\bar{\epsilon}, \bar{\delta})$–separated at (k, j). There are three cases. If $k, j \in B(i)$ then by Obs. 2 $|X_{n+N(\alpha,\bar{\epsilon})}(k) - X_{n+N(\alpha,\bar{\epsilon})}(j)| > \bar{\epsilon}(1 - \alpha)^{N(\alpha,\bar{\epsilon})}$. If $k, j \in V \setminus B(i)$ then $|X_{n+N(\alpha,\bar{\epsilon})}(k) - X_{n+N(\alpha,\bar{\epsilon})}(j)| > \bar{\epsilon}$ since they do not move. If $k \in V \setminus B(i)$ and $j \in B(i)$ then $|X_{n+N(\alpha,\bar{\epsilon})}(k) - X_{n+N(\alpha,\bar{\epsilon})}(j)| > \bar{\epsilon}/8$ and similarly if the roles of k and j are reversed. Hence $X_{n+N(\alpha,\bar{\epsilon})}$ is stochastically $(\bar{\epsilon}(1 - \alpha)^{N(\alpha,\bar{\epsilon})}, \theta(\alpha, \bar{\epsilon})\bar{\delta})$–separated at (k, j). ∎

In the following, we write "HIT i" as shorthand for "a moving event occurs at particle i".

We wish to show that for certain r there exists $\kappa > 0$ and n_1 such that for all $X \in \Upsilon$ the probability X_{n_1} is monotonic exceeds κ. (Note that κ and n_1 are independent of initial conditions.) We know this cannot be true for all r. For example, if $r = 2, \ell = 4$ then an ordering 1 3 2 4 (that is, $X(1) < X(3) < X(2) < X(4)$) cannot be undone through the process. This is because particles 2 and 3 are in the neighborhoods of every particle for this case and hence every hit is effective for them. In general, if $r > \lfloor\frac{\ell-1}{2}\rfloor$ then two or more particles are in every neighborhood and since every hit is effective for them by Observation 2 their order never changes. (Note that the process causes these particles to begin collapsing on themselves and so such a chain would be asymptotically monotone.) So eventual monotonicity may be impossible in these cases. For $r \leq \lfloor\frac{\ell-1}{2}\rfloor$ at most one particle is in every neighborhood and for every particle there is at least one other particle r away and in all cases but one a particle $r+1$ away. This is important for reordering as the proofs below demonstrate.

Proposition 3 *Consider the standard Kohonen process with* $1 \leq r \leq \lfloor\frac{\ell-1}{2}\rfloor$, *then for all* $X \in \Upsilon$ *there exists* $\kappa > 0$ *and* n_1 *such that the probability* X_{n_1} *is monotonic exceeds* κ.

Proof: The proof is based on two lemmas. We will use $X_{\_}$ (or $Y_{\_}$) to refer to the chain $\{X_n\}$ at some point in time when we are not concerned with the particular time or the time is "loosely" understood from the context.

Lemma 8 *Let* $X_0 \in \Upsilon$ *and* $1 \leq r \leq \lfloor\frac{\ell-1}{2}\rfloor$. *If* X'_m *is stochastically* $(\bar{\epsilon}, \bar{\delta})$*-separated then there exists* $\kappa_1 > 0$ *and* n' *such that the probability that* $X'_{m+n'}$ *is monotone exceeds* κ_1.

Proof: The desired event is a sequence of "well-placed" moving events. This sequence can be divided into three sections: (1) Move 1 to an end. (2) Move 2 adjacent to 1. (3) Finish, using induction.

(1) Without loss of generality, assume $X'_0(\ell) > X'_0(1)$. HIT ℓ. This makes $B(\ell)$ consecutive (but perhaps unordered). Continuing with the sequence of moving events, HIT $\max(\ell - r, r+2)$. If $\max(\ell - r, r+2) \neq r+2$ then HIT $\max(\ell - 2r, r+2)$. Continue in this manner until $\max(\ell - kr, r+2) = r+2$

for some k. After a moving event at $r+2$, $X_-(j) > X_-(1)$ for all $j \in V, j \neq 1$. Thus particle 1 is on an end.

(2) HIT 1 and then HIT $r+3$. At this point, particle 1 is still on the end and 2 is adjacent to 1.

(3) Assume particles $1, \ldots, m$ are consecutive and ordered, $X_-(1) < \ldots < X_-(m)$. Also, assume 1 is still an end particle with $X_-(j) > X_-(1)$ for all $j \in V, j \neq 1$ and that there are still separations everywhere. We want to show that with positive probability, in a finite number of steps, we can get m and $m+1$ adjacent with $X_-(m) < X_-(m+1)$. If $r < m+1$ then HIT $m-r+1$ and we are done. If $r \geq m+1$ then HIT m and, subsequently, HIT $m+r+2$ which exists since $m+r+2 \leq 2r+1 \leq \ell$. This completes the induction argument.

After the first moving event, $X'_{m+N(\alpha,\bar{\epsilon})}(j) > X'_{m+N(\alpha,\bar{\epsilon})}(1)$ for all $j \in B(\ell)$. Let $\bar{\epsilon}_1 = \bar{\epsilon}$,

$$\bar{\epsilon}_{n+1} = (1-\alpha)^{N(\alpha,\bar{\epsilon}_n)}\bar{\epsilon}_n$$

and

$$\theta_n = \theta(\alpha, \bar{\epsilon}_n).$$

Then $X'_{m+N(\alpha,\bar{\epsilon})}$ is stochastically $(\bar{\epsilon}_1, \theta_1\bar{\delta})$–separated. By induction and by the Moving Lemma and Observation 2, after M moving events, $X_{m+\sum_{n=1}^{M} N(\alpha,\bar{\epsilon}_n)}$ is $(\bar{\epsilon}_M, \bar{\delta}\prod_{n=1}^{M}\theta_n)$–separated.

Then $n' = \sum_{n=1}^{M} N(\alpha, \bar{\epsilon}_n)$ and $\kappa_1 = \prod_{n=1}^{M'}\theta_n$ where $M' = \lceil \frac{\ell-2}{r} - 1 \rceil + \ell + r - 2$ is the number of moving events. ∎

The next *Separation Lemma* proves to be important in it's own right.

Lemma 9 *Let* $1 \le r \le \lfloor \frac{\ell-1}{2} \rfloor$. *For* $X_0 \in \Upsilon$ *there exists* $\kappa_2 > 0$ *and* n'' *such that the probability of achieving a separation everywhere in* n'' *steps exceeds* κ_2.

Proof: For the initial X_0 we know there is a separation of at least $s = 1/(\ell + 1)$ somewhere. Assume this separation occurs at particle i. HIT i. By the Moving Lemma, $X_{N(\alpha,s)}$ is $(\bar{\epsilon}, \theta(\alpha, s))$–separated at two or three places where $\bar{\epsilon} = s(1 - \alpha)^{N(\alpha,s)}$.

Now have a moving event at one of the "outside" separations. The hope is that with every moving event a new separation is created until all particles are separated. The proof is by contradiction. Assume we reach a point where we cannot create any new separations. To facilitate the discussion we define some new terms. Define a *patch* to be a collection of two or more consecutive particles such that there are separations immediately to the left of the leftmost particle and to the right of the rightmost particle. Call the leftmost and rightmost particles of a patch the *exterior* or *outer* particles and call the remaining particles of the patch *interior* particles. We say the interior particles are *covered* and that the exterior particles are *open.* A patch is said to be *stable* if it cannot be reduced in size due to hits at any of the separations i.e. hitting any of the separations cannot create a separation between any members of the patch. A map $X \in \Upsilon$ is said to be stable if all its patches are stable. If we have a patch G we will use $|G|$ to represent the number of elements in G.

Assume $X_{\_}$ is stable and let G be a patch of $X_{\_}$ with k elements $i_1 < \ldots < i_k$. For G to be stable then the members of $U := \{i_1 + r + 1, i_1 + r + 2, \ldots, i_k + r\} \cap V$ and $L := \{i_1 - r, i_1 - r + 1, \ldots, i_k - r - 1\} \cap V$ must be covered. Let i_{j_1} and i_{j_2}, $i_{j_1} < i_{j_2}$, be the exterior particles of G. Note that since i_{j_1} and i_{j_2} are open they can be hit and, consequently,

$$G \subseteq B(i_{j_1}) \cap B(i_{j_2}),$$

i.e. hitting the exterior elements of a stable patch must be effective for all members of the patch. In particular, we have

$$|i_{j_1} - i_1|,\ |i_{j_1} - i_k|,\ |i_{j_2} - i_1|,\ |i_{j_2} - i_k| \ \leq r$$

Without loss of generality, assume $i_k + r \in V$ and just consider U. (At least one of $i_k + r$ or $i_k - r$ exists by our requirement that $1 \leq r \leq \lfloor \frac{\ell-1}{2} \rfloor$.)

Note that $J_1 := \{i_1+r+1, \ldots, i_{j_1}+r\} = U \cap B(i_{j_1})$ will be moved by hitting at i_{j_1} and so J_1 must be contained in patches separate from $U \setminus J_1$. Similarly, if $J_2 := U \cap B(i_{j_2})$ then $J_2 \setminus J_1$ must be contained in patches separate from J_1 and $U \setminus J_2$. Now $J_2 \setminus J_1$ contains at least one member, $i_{j_2} + r$. Particles that can possibly cover elements of the latter are in $C_1 := \{i_{j_1} + 1, \ldots, i_1 + r\}$ and $C_2 := \{i_k + r + 1, \ldots, i_{j_2} + 2r\}$. However, for such a patch to be stable we cannot use elements of C_2 since hitting i_{j_2} will cause the patch to break up by moving particles of $J_2 \setminus J_1$ but not particles of C_2. Members of C_1 cannot be used since i_{j_1} will move them but not the elements of $J_2 \setminus J_1$. (Note that $J_2 \setminus J_1$ is disjoint from G.) Hence G and thus X_-, are not stable.

So, at each step we can achieve a separation at a new place with positive probability. Consequently, since there are at least two separations after the first moving event, we need at most $\ell + 1$ moving events (taking the endpoints 0 and 1 to be particles). Let $\bar{\epsilon}_1 = s$ and define $\bar{\epsilon}_n$, N_n, and θ_n as in the previous lemma. Then $n'' = \sum_{n=1}^{\ell+1} N_n$ and $\kappa_2 = \prod_{n=1}^{\ell+1} \theta_n$. ∎

(Alternatively, suppose that K particles are in patches, $\ell - K$ are not. Denote the patches $G_1, G_2, \ldots, G_n$ with $|G_j| = k_j$. Thus $\sum_{j=1}^{n} k_j = K$. Let U_j and

L_j correspond to G_j as above and let each patch have particles $i_{1_j} < \ldots < i_{k_j}$. Let $i_{k_j} - i_{1_j} = m_j$. So, $k_j \le m_j$. For a given patch G_j, if either $i_{1_j} - r \in V$ or $i_{k_j} + r \in V$, then the number of particles that must be covered relative to G_j, $|U_j \cup L_j|$, is greater than or equal to m_j. If neither of these particles are in V then $i_{1_j} < 1 + r \le \ell - r < i_{k_j}$. So $|U_j| = \ell - i_{1_j}$ and $|L_j| = i_{k_j} - r - 1$. Thus, $|U_j \cup L_j| = \ell - r - 1 + i_{k_j} - i_{1_j} = \ell - r - 1 + m_j > m_j$. To cover $U_j \cup L_j$, we need at least two additional particles. Hence, $K \ge \sum_{j=1}^{n}(m_j + 1) \ge \sum_{j=1}^{n}(k_j + 1) = K + n$. Thus, $n = 0$ and, consequently, $K = 0$.)

Proof of the Proposition: The proposition follows from the above two lemmas: $n_1 = n' + n''$ and $\kappa = \kappa_1 \kappa_2$. ∎

Remark: Note that from Proposition 3 it follows that the hitting time to Υ_{m} is finite almost surely regardless of the initial positions.

From Lemma 9 we can prove

Lemma 10 *Let X_0, $Y_0 \in \Upsilon$. There exists a $\kappa_3 > 0$ such that the probability X' and Y' are both separated in $2n''$ steps exceeds κ_3.*

Proof: Separate X as in Lemma 9. Now consider Y. There is a separation $s > 1/(\ell + 1)$ between two of the Y–particles. Without loss of generality, suppose the separation is to the right of $Y(i)$. Thus if $\omega \in (Y(i), Y(i) + s/2)$ then $Y(i)$ will be hit. Within the latter interval are at most ℓ X–particles. There is a separation $s' > s/2(\ell + 1)$ between two of these particles. Without loss of generality, suppose the separation is to the right of $X(k)$. Now start the first moving event with a hit in $(X(k) + s'/4, X(k) + s'/2)$. Now both $Y(i)$ and $X(k)$ will be hit with this moving event. Follow this procedure for every moving event in Lemma 9. ∎

The following *No-Split Lemma* shows that close particles are not likely to split.

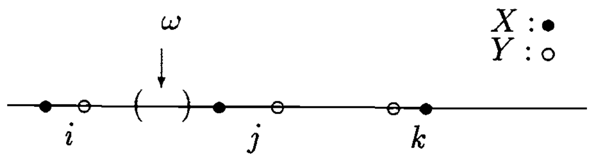


FIGURE 5.2. Splitting

Lemma 11 *For each $\epsilon > 0$, if the initial X_0 and Y_0 satisfy $D(X_0, Y_0) \leq \epsilon$ and are paired, then the probability that after one step X and Y split is bounded by $2\ell\epsilon$.*

Proof: Assume $X_0(i)$, $Y_0(i) < X_0(j))$, $Y_0(j) < X_0(k)$, $Y_0(k)$ and that the X_0 particle that is hit is the jth particle. If there is a split, then the Y_0 particle that is hit is either the jth particle or the kth particle. Suppose the latter is hit then $(Y_0(i) + Y_0(k))/2 < \omega < (X_0(i) + X_0(k))/2$. But the length of this interval is bounded by ϵ. So the probability that it is split in this way is bounded by ϵ. Similarly, the probability of a split where the Y_0 particle hit is i is bounded by ϵ. Since there are ℓ particles, the probability of a split for some j is bounded by ℓ times 2ϵ. ∎

5.3.2. *Existence of the Invariant Measure*

Recall that a stationary Markov chain, $\{X_n\} \subset \Lambda$, is *(weak) Feller* if the map $x \mapsto \boldsymbol{E}[f(x)]$ is continuous for all bounded, continuous $f : \Lambda \to \boldsymbol{R}$.

We can now prove

Proposition 4 *Let the environmental measure, μ, be as in Theorem 16, then the process starting in Υ is Feller.*

Proof: Fix $X_0 \in \Upsilon$. Let $\eta' = \min_{i\neq j} |X_0(i) - X_0(j)|$. Let $\delta' < \eta'/4$ and

$$(1-\alpha)(\frac{\eta'}{2} - 2\delta') > \eta > 0 \tag{5.20}$$

Define

$$A_\eta = \{(x_1, x_2, \ldots, x_\ell) \in [0,1]^\ell : \min_{i\neq j} |x_i - x_j| \geq \eta\} \tag{5.21}$$

Since A_η is compact f is uniformly continuous on it. Pick $\epsilon > 0$ and let δ'' be such that

$$|f(x_1, x_2, \ldots, x_\ell) - f(y_1, y_2, \ldots, y_\ell)| < \epsilon/2 \tag{5.22}$$

when

$$\|(x_1, x_2, \ldots, x_\ell), (y_1, y_2, \ldots, y_\ell)\| < \delta'' \tag{5.23}$$

and $(x_1, x_2, \ldots, x_\ell), (y_1, y_2, \ldots, y_\ell) \in A_\eta$. Now let h be the Radon-Nikodym derivative of μ with respect to Lebesgue measure λ, with $0 < a_1 \leq h \leq a_2$. Let

$$0 < \delta < \min\{\delta', \delta'', \frac{\epsilon}{4M(\ell-1)a_2}\} \tag{5.24}$$

and Y_0 be such that

$$D(X_0, Y_0) < \delta. \tag{5.25}$$

Then $Y_0 \in \Upsilon$ and X_0 and Y_0 are paired by (5.20) and (5.24). Note that

$$\min_{i\neq j} |Y_0(i) - Y_0(j)| > \eta' - 2\delta \tag{5.26}$$

Define

$$\bar{x}_0 = 0, \quad \bar{x}_k = \frac{x_k + x_{k+1}}{2} \quad 1 \le k \le \ell - 1, \quad \bar{x}_\ell = 1 \tag{5.27}$$

and similarly for $\bar{y}_k$. So,

$$|\boldsymbol{E}f(X_0) - \boldsymbol{E}f(Y_0)| =$$
$$|\sum_{k=1}^{\ell-1} \int_{x_{k-1}}^{x_k} f(F(\omega, X_0))\mu(d\omega) - \sum_{k=1}^{\ell-1} \int_{y_{k-1}}^{y_k} f(F(\omega, Y_0))\mu(d\omega)| \tag{5.28}$$

Let

$$\tilde{x}_0 = 0, \quad \tilde{x}_{2k-1} = \min(\hat{x}_k, \hat{y}_k), \quad \tilde{x}_{2k} = \max(\hat{x}_k, \hat{y}_k), \quad \tilde{x}_{2\ell-1} = 1 \tag{5.29}$$

for $1 \le k \le \ell - 1$.

From (5.28) then,

$$|\boldsymbol{E}f(X_0) - \boldsymbol{E}f(Y_0)| =$$
$$|\sum_{k=0}^{2\ell-1} \int_{\tilde{x}_k}^{\tilde{x}_{k+1}} [f(F(\omega, X_0)) - f(F(\omega, Y_0))]\mu(d\omega)| \tag{5.30}$$
$$\le \sum_{k=0}^{2\ell-1} \int_{\tilde{x}_k}^{\tilde{x}_{k+1}} |f(F(\omega, X_0)) - f(F(\omega, Y_0))|\mu(d\omega) \tag{5.31}$$

There are two cases to consider with respect to (5.30):

(1) The intervals $(\tilde{x}_{2k}, \tilde{x}_{2k+1})$: In this case, particle $k+1$ is hit for both X_0 and Y_0. Thus there is a joint hit and, consequently, $D(X_1, Y_1) < \delta$. Though, the distance between particles in the hit neighborhoods has decreased it is still bounded below by η by (5.20), (5.24), (5.26). Hence, both $F(\omega, X_0),\ F(\omega, Y_0) \in A_\eta$. So

$$\int_{\tilde{x}_{2k}}^{\tilde{x}_{2k+1}} |f(F(\omega, X_0)) - f(F(\omega, Y_0))|\mu(d\omega) \le \frac{\epsilon}{2}\mu((\tilde{x}_{2k}, \tilde{x}_{2k+1})) \tag{5.32}$$

(2) The intervals $(\tilde{x}_{2k-1}, \tilde{x}_{2k})$: On these intervals there is a split. From the proof of the No-Split Lemma,

$$\tilde{x}_{2k} - \tilde{x}_{2k-1} < \delta \tag{5.33}$$

Let h be the Radon-Nikodym derivative of μ with respect to Lebesgue measure λ, with $0 < a_1 \le h \le a_2$. Then

$$\int_{\tilde{x}_{2k-1}}^{\tilde{x}_{2k}} |f(F(\omega, X_0)) - f(F(\omega, Y_0))|\mu(d\omega) \le 2M\mu((\tilde{x}_{2k-1}, \tilde{x}_{2k})) \tag{5.34}$$

$$\le 2Ma_2\delta \tag{5.35}$$

Thus,

$$\begin{aligned}
&|\boldsymbol{E}f(X_0) - \boldsymbol{E}f(Y_0)| \le \\
&\quad \sum_{k=0}^{\ell-1} \int_{\tilde{x}_{2k}}^{\tilde{x}_{2k+1}} |f(F(\omega, X_0)) - f(F(\omega, Y_0))|\mu(d\omega) + \\
&\qquad \sum_{k=1}^{\ell-1} \int_{\tilde{x}_{2k-1}}^{\tilde{x}_{2k}} |f(F(\omega, X_0)) - f(F(\omega, Y_0))|\mu(d\omega) \\
&\le \frac{\epsilon}{2} \sum_{k=0}^{\ell-1} \mu((\tilde{x}_{2k}, \tilde{x}_{2k+1})) + 2M(\ell - 1)a_2\delta \\
&< \epsilon
\end{aligned} \tag{5.36}$$

Hence the result. ∎

Let ϵ be the final separation referred to in Lemma 9. Let $A := \{X \in \Upsilon : |X(i) - X(j)| \ge \epsilon,\ i \ne j\}$. Then by Lemma 9

$$\boldsymbol{P}(X_n \in A\ i.o.) = 1 \tag{5.37}$$

The latter along with the fact that the chain is Feller imply by Theorem 12.3.2 in [29] that $\{X_n\}$ has an invariant measure π that is finite on compact sets. Since Lemma 9 implies that the expected return time to A is finite, we may assume an invariant probability measure. To see this, define a probability measure π_A by

$$\pi_A(C) := \frac{\pi(A \cap C)}{\pi(A)}$$

for measurable C. Let

$$S_A = \inf\{n \geq 0 : X_n \in A\}$$

$$T_A = \inf\{n > S_A : X_n \in A\}$$

Define a measure,

$$\bar{\mu}(C) = \sum_{n=1}^{\infty} \boldsymbol{P}_{\pi_A}(T_A > n, X_n \in C), \qquad C \subseteq \Upsilon \setminus A.$$

Now, for measurable C, let

$$\bar{\nu}(C) = \pi_A(C) + \bar{\mu}(C \cap \Upsilon \setminus A)$$

The reader can check that $\bar{\nu}$ is invariant for $\{X_n\}$ (see section 3.4.4). So let

$$\nu = \frac{\bar{\nu}}{\boldsymbol{E}_{\pi_A} T_A}.$$

5.3.3. *Monotonic Maps*

The next few lemmas will deal with monotonic maps. We now prove the *Shrinking Lemma* which is based on the notion of a pairing.

Lemma 12 *For $r \geq 1$ there exists $\epsilon_0 > 0, \gamma > 0$ and n_0 such that for all X_0 and Y_0 in Υ_{m} with $D(X_0, Y_0) \leq \epsilon_0$ and X_0 and Y_0 paired, the probability that after n_0 steps there is a shrinking exceeds γ.*

Proof: Assume there are ℓ particles and $\epsilon_0 = \epsilon_0(\ell, \alpha) > 0$ is to be determined. Let X_0, Y_0 be the initial particle positions and, without loss of generality, assume they are monotonic increasing. First, note that if $r \geq \ell$ then we are done since every hit is a shrinking event. Second, note that it is not necessary to have a joint hit at every particle to have a shrinking event i.e. given X_0, Y_0 we could achieve a shrinking event with $\lfloor \ell/2r \rfloor + 1$ well-placed hits. Third, note that if

$X_n(i+1) - X_n(i) \geq t$ then $Y_n(i+1) - Y_n(i) \geq t - 2D(X_n, Y_n)$. Furthermore, if $D(X_n, Y_n) < \epsilon_0 < t/2$ and $0 < \tilde{t} < (t/2) - \epsilon_0$ we can achieve a joint hit at i if $\omega_{n+1} \in (X_n(i) + \tilde{t}/2, X_n(i) + \tilde{t})$. Similarly, we can achieve a joint hit at $i+1$ if $\omega_{n+1} \in (X_n(i+1) - \tilde{t}, X_n(i+1) - \tilde{t}/2)$. Finally, notice that if $i > r$ in the above, then for $\omega_{n+1} \in (X_n(i) + (\tilde{t}/2), X_n(i) + \tilde{t})$ we are guaranteed a separation at $X_{n+1}(i-r)$ and $X_{n+1}(i-r-1)$. Specifically, we have that

$$\begin{aligned} X_{n+1}(i-r) &= \alpha\omega_1 + (1-\alpha)X_n(i-r) \\ &\geq \alpha(X_n(i-r) + (\tilde{t}/2)) + (1-\alpha)X_n(i-r) \\ &= X_n(i-r) + (\alpha\tilde{t}/2). \end{aligned}$$

Since $X_{n+1}(i-r-1) = X_n(i-r-1)$, $X_{n+1}(i-r) - X_{n+1}(i-r-1) \geq \alpha\tilde{t}/2$. We know there is a separation of at least $s = 1/(\ell+1)$ between $X_0'(i)$ and $X_0'(i+1)$ for some $i = 0, 1, \ldots, \ell+1$ i.e. $X_0'(i+1) - X_0'(i) \geq s$. Then $Y_0(i+1) - Y_0(i) \geq s - 2\epsilon_0$. Let $s_1 = ((s/2) - 2\epsilon_0)/2$ and let $I_1 = (X_0(i) + s_1, X_0(i) + 2s_1)$. Then if $\omega_1 \in I_1$, by the above discussion, we have a joint hit at i. The latter argument holds for all i including the case where the separation is between $X_0(\ell)$ and 1 $(= X_0'(\ell+1))$. If the separation is between 0 $(= X_0'(0))$ and $X_0(1)$ then let $I_1 = (X_0(1) - 2s_1, X_0(1) - s_1)$. The probability of this event (a joint hit at i) is then greater than s_1.

Such an event creates separations between $i-r-1$ and $i-r$ and between $i+r$ and $i+r+1$ for $i > r, l-i > r$. For $i > r$ then $X_1(i-r) - X_1(i-r-1) \geq \alpha s_1$ by the above. Similarly, for $l - i > r$, $X_1(i+r+1) - X_1(i+r) \geq 2\alpha s_1$. With these separations we can achieve joint hits at $i-r-1$ and $i+r+1$. (If, for example, $l - i > r$, then to have a joint hit at $i+r+1$, let $s_2 = (\alpha s_1/4) - \epsilon_0$, $\omega_2 \in I_2 = (X_1(i+r+1) - 2s_2, X_1(i+r+1) - s_2)$.

An induction argument, using

$$s_k = \frac{\alpha s_{k-1}}{4} - \epsilon_0 = (\frac{\alpha}{4})^{k-1}(\frac{s}{4}) - \epsilon_0 \frac{1-(\frac{\alpha}{4})^k}{1-\frac{\alpha}{4}},$$

shows that we can achieve a shrinking event in $n_0 = \lfloor \ell/(r+1) \rfloor + 1$ steps. We must have then that

$$0 < \epsilon_0 < (\frac{1-\frac{\alpha}{4}}{1-(\frac{\alpha}{4})^{\lfloor \ell/(r+1) \rfloor + 1}})(\frac{\alpha}{4})^{\lfloor \ell/(r+1) \rfloor}(\frac{s}{4}).$$

Further,

$$\gamma \geq \prod_{k=1}^{\lfloor \ell/(r+1) \rfloor + 1} s_k. \blacksquare \tag{5.38}$$

A *gluing* is the event that two particles never split in the future. The previous two lemmas combine to give a *Gluing Lemma.*

Lemma 13 *For all $\rho < 1$ there exists $\epsilon > 0$ such that if the starting $X_0, Y_0 \in \Upsilon_{\mathbf{m}}$ are paired and satisfy $D(X_0, Y_0) \leq \epsilon$, then the probability of a gluing (never having a split in the future) exceeds ρ.*

Proof: Choose $\rho < 1$. Take $0 < \epsilon < \epsilon_0$, where ϵ_0 is as in the Shrinking Lemma. Let T_k be the time to the kth shrinking event (that all ℓ particles are hit with no split). If there is has been no split up to T_k, then the particle distance has decreased so that $D(X_{T_k}, Y_{T_k}) \leq \epsilon(1-\alpha)^k$. After T_k look at blocks of n_0 steps, with the n_0 of the Shrinking Lemma. The probability that T_{k+1} has not been reached and that two particles are split for the first time in the jth block of length n_0 after T_k is bounded by $(1-\gamma)^j 2\ell n_0 \epsilon (1-\alpha)^k$, by the Shrinking Lemma and the No-Split Lemma.

Sum over $k \geq 0$ and $j \geq 0$. This gives a bound for the probability of a split ever occurring of $2\ell n_0 \epsilon/(\gamma\alpha)$. This is less than $1-\rho$ for ϵ small enough. ■

The next step is the following *Starting Lemma.*

Lemma 14 *Let $X_0, Y_0 \in \Upsilon_{\mathrm{m}}$, then there exists n_2 and $\delta > 0$ such that the probability that $D(X_{n_2}, Y_{n_2}) < \epsilon_0$ and X_{n_2}, Y_{n_2} are paired exceeds δ.*

Proof: We break the proof up into a series of steps.

1.) *Separating X particles and Y particles.* By Lemma 10 there is a strictly positive probability that both X' and Y' become separated in $2n''$ steps.

2.) *Establishing an upper bound on the distance between same numbered particles.* Once there is a lower bound on the separations between X–particles and the separations between Y–particles, we cannot say much about the distance $|X_{-}(k)-Y_{-}(k)|$, $1 \leq k \leq \ell$. We can remedy this through a series of moving events in which all the hits are joint hits. Without loss of generality, assume that X_0 and Y_0 are both monotonic increasing. Once both X and Y particles are $\bar{\epsilon}$–separated a joint hit is guaranteed at particle 1 if $\omega \in (0, \bar{\epsilon})$ and at particle ℓ if $\omega \in (1-\bar{\epsilon}, 1)$. We will start with joint hits at particle ℓ below but first we consider the following prototypical situation: Fix a vertex k and let $m = \min(X(k), Y(k))$, $M = \max(X(k-1), Y(k-1))$. Suppose that for some $d_1, d_2 > 0$ it holds that $|X(k) - Y(k)| < d_1, m - M > d_2$. If $s > 0$ and $\omega \in (m - s/2, m - d_3/4)$, then to have a joint hit at k for all such ω , $d_1 - s/2 < d_2 - s/2$ or $s < d_2 - d_1$.

For the (n+1)st moving event let $\bar{s}_n$ correspond to s above (as in the Moving Lemma). Let $N_n = \sum_{k=0}^{n-1} N(\alpha, \bar{s}_k) + 2n''$ and

$$\begin{aligned}
x_0 &= 1, \\
x_n &= \min(X_{N_n}(\ell - nr), Y_{N_n}(\ell - nr)), \\
x_n^* &= \max(X_{N_n}(\ell - nr - 1), Y_{N_n}(\ell - nr - 1)), \\
d_{1,n} &= |X_{N_n}(\ell - nr) - Y_{N_n}(\ell - nr)|, \\
d_{2,n} &= x_n - x_n^*,
\end{aligned}$$

for $1 \leq n \leq m_1$ where $2 \leq \ell - m_1 r < r + 2$, i.e.

$$\frac{\ell - 2}{r} - 1 < m_1 \leq \frac{\ell - 2}{r}. \tag{5.39}$$

Let $I_n = (x_n - \bar{s}_n/2, x_n - \bar{s}_n/4), I'_n = (x_n - 3\bar{s}_n/4, x_n - \bar{s}_n/8), n \geq 0, n \neq m_1 + 1$, (see below) where I_n is the "hit interval" in the Moving Lemma.

If $\bar{s}_0 = \bar{\epsilon}$ then $I_0 = (1 - \bar{\epsilon}/2, 1 - \bar{\epsilon}/4)$ and ℓ is joint hit for all $\omega_j \in I_0, 2n'' < j < N_1$. After this event, $X_{N_1}(\ell - r), Y_{N_1}(\ell - r) \in I'_0$ and $x_1 > x_0 - 3\bar{\epsilon}/4$. Since $1 - X_{2n''}(\ell - p) > (p+1)\bar{\epsilon}$ and the same is true substituting Y for X then $d_{2,1} = x_1 - x^*_1 > (r + 5/4)\bar{\epsilon} > 9\bar{\epsilon}/4$. And since $d_{1,1} < 5\bar{\epsilon}/8$, then $d_{2,1} - d_{1,1} > 13\bar{\epsilon}/8$. Choosing $\bar{s}_1 = \bar{\epsilon}$, for example, the second moving event at x_1 ($\omega \in I_1$) is a series of joint hits at $\ell - r$.

Suppose $\bar{s}_{n-1} = \bar{\epsilon}$ for $1 \leq n \leq m_1$. By induction,

$$x_n > \frac{3}{4}\sum_{k=0}^{n} \bar{s}_k = 1 - \frac{3n}{4}\bar{\epsilon} \tag{5.40}$$

for $n > 1$ and

$$1 - x^*_n > (nr + 2)\bar{\epsilon} \tag{5.41}$$

so $d_{2,n} > (n(r - 3/4) + 2)\bar{\epsilon} > (n/4 + 2)\bar{\epsilon} > \bar{\epsilon}$. Thus $\bar{s}_n = \bar{\epsilon}$ is sufficient for our purposes for $n \leq \ell/r$.

Note that for $\ell - m_1 r \leq k \leq \ell$ and $m_1 > 0$ particle k has "undergone" two moving events while particles 1 through $\ell - m_1 r - 1$ have undergone one moving event. After the first moving event, if the initial separation is s (see the Moving Lemma) then the distance between particles of the hit neighborhood is bounded above by $5s/8$. For a particle k undergoing a second moving event with a possibly different second separation s',

$$|X_{\_}(k) - Y_{\_}(k)| < \frac{5s}{8}(1 - \alpha)^{N(\alpha, s')} < \frac{5s}{8}\frac{s'}{16} \tag{5.42}$$

by equation 5.17. For $\ell - m_1 r \leq k \leq \ell$ then $|X_{N_{m_1+1}}(k) - Y_{N_{m_1+1}}(k)| < 5\bar{\epsilon}^2/128$.

3.) *Pairing of particles 1 through $\ell - r - 1$.* Let $y_1 = 0$ and $I_{m_1+1} = (\bar{\epsilon}/4, \bar{\epsilon}/2)$ i.e. $\bar{s}_{m_1+1} = \bar{\epsilon}$. A moving event at 0 will be a series of joint hits of 1 since by Lemma 9, Lemma 7 and the above $X_{N_{m_1+1}}(1), Y_{N_{m_1+1}}(1) > \bar{\epsilon}$. We are now ready to begin pairing the particles. If there is a joint moving event at $r+2$ then there will be a pairing at 1. If subsequently there are joint moving events at 2 and $r+3$ then there will be a pairing at 2 and so on.

Let

$$
\begin{aligned}
y_{2k} &= \min(X_{N_{m_1+2k}}(r+k+1), Y_{N_{m_1+2k}}(r+k+1)),\\
y_{2k+1} &= \min(X_{N_{m_1+2k+1}}(k+1), Y_{N_{m_1+2k+1}}(k+1)),\\
y^*_{2k} &= \max(X_{N_{m_1+2k}}(r+k), Y_{N_{m_1+2k}}(r+k)),\\
y^*_{2k+1} &= \max(X_{N_{m_1+2k+1}}(k), Y_{N_{m_1+2k+1}}(k))\\
d_{1,m_1+2k} &= |X_{N_{m_1+2k}}(r+k+1) - Y_{N_{m_1+2k}}(r+k+1)|,\\
d_{1,m_1+2k+1} &= |X_{N_{m_1+2k+1}}(k+1) - Y_{N_{m_1+2k+1}}(k+1)|,\\
d_{2,m_1+n} &= y_n - y^*_n,
\end{aligned}
$$

for $1 \le k \le \ell - r - 1$. Then

$$
\begin{aligned}
y_2 = \min(X_{N_{m_1+2}}(r+2), Y_{N_{m_1+2}}(r+2)) &\ge x_{m_1}\\
&> 1 - \frac{3}{4} m_1 \bar{\epsilon}\\
&> 1 - \frac{3(\ell-2)}{4r}\bar{\epsilon}\\
&\ge (\ell+1)\bar{\epsilon} - \frac{3(\ell-2)}{4r}\bar{\epsilon}\\
&= \frac{\ell+10}{4}\bar{\epsilon}
\end{aligned}
$$

using (5.40) and (5.39) and the facts that $1 \ge (\ell+1)\bar{\epsilon}$ and that $r \ge 1$. Since $y^*_2 = \max(X_{N_{m_1+2}}(r+1), Y_{N_{m_1+2}}(r+1)) < 3\bar{\epsilon}/4$, $d_{2,m_1+2} = y_2 - y^*_2 \ge (\ell+7)/4 > 5\bar{\epsilon}/2$ for $m_1 > 0$. From the above, $d_{1,m_1+2} < 5\bar{\epsilon}^2/128$. So $\bar{s}_{m_1+2} = \bar{\epsilon}$ will insure joint hits at $r+2$.

After a moving event at y_2 all particles except 1 are above $y_2 - 3\bar{\epsilon}/4 > 5\bar{\epsilon}/2 - 3\bar{\epsilon}/4 = 7\bar{\epsilon}/4$. Note also that $|X_{N_{m_1+2}}(1) - Y_{N_{m_1+2}}(1)| < 5\bar{\epsilon}/8$ so there is a pairing at 1.

Now $d_{2,m_1+3} = y_3 - y_3* > 7\bar{\epsilon}/4 - 3\bar{\epsilon}/4 = \bar{\epsilon}$ and

$$d_{1,m_1+3} = |X_{N_{m_1}+2}(2) - Y_{N_{m_1}+2}(2)| < \frac{5\bar{\epsilon}}{8}\frac{\bar{\epsilon}}{16}\frac{\bar{\epsilon}}{16} \tag{5.43}$$

(since particle 2 was in hit neighborhoods of x_{m_1}, y_0, y_1). Thus $\bar{s}_{m_1+3} = \bar{\epsilon}/2$ guarantees joint hits at 2.

Recall that for a given initial separation s an moving event creates a separation greater than $s/8$ between the particles of the hit neighborhood and the other particles. A moving event at particle ℓ will be the last moving event for this stage, causing a pairing at $\ell - r - 1$. For each particle (in this stage) there are two moving events, so let $m_2 = 2(\ell - r - 1)$. Suppose

$$d_{2,m_1+n} > \frac{\bar{s}_{m_1+n-1}}{8} \tag{5.44}$$

and let

$$\bar{s}_{m_1+n} = \frac{\bar{\epsilon}}{2 \cdot 16^{n-3}} \tag{5.45}$$

for $3 \leq n \leq m_2$. Thus,

$$\bar{s}_{m_1+n} = \frac{\bar{s}_{m_1+n-1}}{16} \tag{5.46}$$

for $4 \leq n \leq m_2$. For joint hits, recall it is sufficient that $\bar{s}_{m_1+n+1} < d_{2,n_1+n+1} - d_{1,m_1+n+1}$. If $n = 2k$ then $y_n = \min(X_{N_{m_1+n}}(r+k+1), Y_{N_{m_1+n}}(r+k+1))$ and particle $r+k+2$ is in the hit neighborhood for the moving event at y_n. If $n = 2k-1$ then $y_n = \min(X_{N_{m_1+n}}(k+1), Y_{N_{m_1+n}}(k+1))$ and particles $r+k+1$ and $k+2$ are in the hit neighborhood for the moving event at y_n. In both cases,

$$d_{1,m_1+n+1} < \frac{5\bar{\epsilon}}{8} \cdot \frac{\bar{\epsilon}}{16} \cdot \frac{\bar{s}_{m_1+n-1}}{16} = \frac{5\bar{\epsilon}}{8} \cdot \frac{\bar{\epsilon}}{16} \cdot \bar{s}_{m_1+n} < \frac{\bar{s}_{m_1+n}}{16}. \tag{5.47}$$

So

$$d_{2,m_1+n+1} - d_{1,m_1+n+1} > \frac{\bar{s}_{m_1+n}}{8} - \frac{\bar{s}_{m_1+n}}{16} = \frac{\bar{s}_{m_1+n}}{16} = \bar{s}_{m_1+n+1} \tag{5.48}$$

After the moving event at m_2, particles 1 through $\ell - r - 1$ are adjacent. Consideration of $d_{2,m_1+n} - d_{1,m_1+n}$ shows they are also paired.

4.) *Pairing the remaining particles (ℓ through $\ell - r$).* Let

$$\begin{aligned}
z_1 &= 1, \\
z_{2k} &= \min(X_{N_{m_1+m_2+2k+1}}(\ell - r - k), Y_{N_{m_1+m_2+2k+1}}(\ell - r - k)), \\
z_{2k+1} &= \min(X_{N_{m_1+m_2+2k+2}}(\ell - k), Y_{N_{m_1+m_2+2k+2}}(\ell - k)), \\
z^*_{2k} &= \max(X_{N_{m_1+m_2+2k+1}}(\ell - r - k - 1), Y_{N_{m_1+m_2+2k+1}}(\ell - r - k - 1)) \\
z^*_{2k+1} &= \max(X_{N_{m_1+m_2+2k+2}}(\ell - k - 1), Y_{N_{m_1+m_2+2k+2}}(\ell - k - 1)) \\
d_{1,m_1+m_2+2k+1} &= |X_{N_{m_1+m_2+2k+1}}(\ell - r - k) - Y_{N_{m_1+m_2+2k+1}}(\ell - r - k)| \\
d_{1,m_1+m_2+2k+2} &= |X_{N_{m_1+m_2+2k+2}}(\ell - k) - Y_{N_{m_1+m_2+2k+2}}(\ell - k)| \\
d_{2,m_1+m_2+n} &= z_n - z^*_n
\end{aligned}$$

Note that from the Moving Lemma and the description of the above process that $\max(X_{N_{m_1+m_2}}(\ell), Y_{N_{m_1+m_2}}(\ell)) < 1 - \bar{\epsilon}/2$ and so choosing $s_{m_1+m_2+1} = \bar{\epsilon}/4$ insures joint hits at ℓ. Now proceed as with the y's.

5.) *Getting the maps arbitrarily close.* With the maps paired and still stochastically separated they can be made arbitrarily close by a finite number of shrinking events. The result follows. ∎

Lemma 15 *The probability of eventual glueing (finitely many splits) is one.*

Proof: The probability of becoming monotonic in a finite time is 1 so we assume the initial maps are monotonic. Let τ_1 be the time of the first split after n_2 steps, where n_2 is from the starting lemma. Let τ_k be the time of the first split after $\tau_{k-1} + n_2$ steps, given $\tau_{k-1} < \infty$. If $\tau_{k-1} = \infty$ then set $\tau_k = \infty$. Then the $\{\tau_k\}$ are stopping times. We wish to show $\boldsymbol{P}$(infinitely many splits) = 0. Note that

$$[\text{infinitely many splits}] = \cap_{k=1}^{\infty}[\tau_k < \infty].$$

By the Starting and Gluing Lemmas,

$$\boldsymbol{P}(\tau_1 < \infty) < 1 - \delta\rho.$$

Assuming

$$\boldsymbol{P}(\tau_k < \infty) < (1 - \delta\rho)^k,$$

then

$$\begin{aligned}\boldsymbol{P}(\tau_{k+1} < \infty) &= \boldsymbol{P}(\tau_{k+1} < \infty | \tau_k < \infty)\boldsymbol{P}(\tau_k < \infty)\\ &= \mathbf{P}(\text{there is a split after } \tau_k + n_2 \text{ steps})\mathbf{P}(\tau_k < \infty)\\ &< (1 - \delta\rho)^{k+1}.\end{aligned}$$

where in line 3 we use the fact that the Starting Lemma does not depend on initial conditions given the maps are monotonic. Hence $\mathbf{P}$(infinitely many splits) = 0. ∎

5.3.4. *The Waiting Time W*

We now begin to define W from the definition of super-stability. We consider a skeleton chain with a scaled, shifted 'version' of W, W'. We wish to show that the distance between maps $X, Y \in \Upsilon$ diminishes quickly after a waiting time W which is itself geometrically bounded. We do this by showing that shrinkings occur "regularly" while splits and "not-close" events, i.e. not having enough shrinkings per time, die off rapidly. Let $n^* = \max\{n_0, n_1, n_2\}$ where n_0 is from the Shrinking Lemma, n_1 is from the Monotonic Proposition and n_2 is from the Starting Lemma. Divide the time axis into n^*-blocks. Given $X_0, Y_0 \in \Upsilon$ and $\{\omega_k\}$, determine $\{X_k\}, \{Y_k\}$. Let

$$\begin{aligned}
M_{X,n} &= \text{monotonic event in the nth } n^*\text{-block for } \{X_k\} \\
M_{Y,n} &= \text{monotonic event in the nth } n^*\text{-block for } \{Y_k\} \\
E_n &= \text{at least one starting event and no splits in the nth } n^*\text{-block} \\
S_n &= \text{at least one split in the nth } n^*\text{-block} \\
H_n &= \text{at least one shrinking event and no splits in the nth } n^*\text{-block}
\end{aligned}$$

where, by a starting event in the nth n^*-block we mean the event detailed in the Starting Lemma so $D(X_{nn^*}, Y_{nn^*}) < \epsilon_0$ where ϵ_0 is from the Shrinking Lemma. Note that E_n and H_n are only defined on when the maps are monotonic so $E_n, H_n \subseteq (\cup_{k=1}^{n-1} M_{X,k}) \cap (\cup_{k=1}^{n-1} M_{Y,k})$. Denote an element of the probability space $(\Omega, \mathcal{F}, \boldsymbol{P})$ by $\bar{\omega}$. If A_n is a sequence of events with n denoting the nth n^*-block, then define $T(\{A_n\}, t)(\bar{\omega}) := T(A_n, t)(\bar{\omega}) := \inf\{n > t : \bar{\omega} \in A_n\}$.

Inductively, we can construct [9] an independent sequence of sets H'_n such that $H'_n \subseteq H_n$ and $\boldsymbol{P}[H'_n | E_k \cap (\cap_{i=k+1}^{n-1} S_i^c)] = \gamma$ for $0 < k < n$. Let $h'_n = \mathbf{1}_{H'_n}$ then h'_n are iid Bernoulli random variables with mean γ. From the theory of large deviations [32],

$$\boldsymbol{P}[\frac{h'_1 + \cdots + h'_n}{n} < \gamma - \epsilon] \leq e^{-2\epsilon^2 n}. \tag{5.49}$$

We can now define so-called "not-close" events. If $C_{n,k} = [\frac{h'_{n+1}+\cdots+h'_{n+k}}{k} < \frac{\gamma}{2}]$ then $\boldsymbol{P}(\cup_{k=1}^{\infty} C_{n,k}) \le \sum_{k=1}^{\infty} \boldsymbol{P}(C_{n,k}) \le e^{-\frac{\gamma^2}{2}k}/(1-e^{-\frac{\gamma^2}{2}})$. Let M be large enough so that if

$$C_n = [\frac{h'_{n+1} + \cdots + h'_{n+p}}{p} < \frac{\gamma}{2} \quad \text{for some} \quad p > M]$$

then

$$\boldsymbol{P}(C_n) \le \sum_{k=M+1}^{\infty} e^{-\frac{\gamma^2}{2}k} := \eta < 1$$

That is, $M > 1 - \frac{2}{\gamma^2}\log(1-e^{-\frac{\gamma^2}{2}})$. Thus C_n is our "not-close" event. Further, let

$$\tilde{T}(C_n)(\bar{\omega}) = \inf\{n + p : \frac{h'_{n+1} + \cdots + h'_{n+p}}{p} < \frac{\gamma}{2},\ p > M,\ \bar{\omega} \in C_n\}$$

Now let

$$\begin{aligned}
T_0 &= 0 \\
T_1 &= \min\{T(M_{X,n}, 0), T(M_{Y,n}, 0)\} \\
T_2 &= \max\{T(M_{X,n}, 0), T(M_{Y,n}, 0)\} \\
T_{2k+1} &= T(E_n, T_{2k}), \qquad k \ge 1 \\
T_{2k} &= \min\{T(S_n, T_{2k-1}), \tilde{T}(C_{T_{2k-1}})\}, \qquad k \ge 2 \\
L &= 1 + \sup\{2k : T_{2k} < \infty\} \\
W_k &= T_k - T_{k-1}, \qquad k \ge 1 \\
W' &= M + \textstyle\sum_{k=1}^{L} W_k
\end{aligned}$$

The waiting times W_k are geometrically bounded as is L.

Lemma 16 *There exists constants $a > 0$, $0 < b < 1$ such that $\boldsymbol{P}(W_k = n) < ab^n$ for all $k \ge 1$.*

Proof: For $k = 1, 2$, $\boldsymbol{P}(W_k = n) < (1-\kappa)^{n-1}$ by the Monotonic Proposition. By the Starting Lemma, $\boldsymbol{P}(W_{2k+1} = n) < (1-\delta)^{n-1}$, for $k \ge 1$. If $W_{2k} = \tilde{T}(C_{T_{2k-1}}) - T_{2k-1}$ then by equation 5.49 $\boldsymbol{P}(W_{2k} = n) < e^{-\frac{\gamma^2}{2}n}$. If $W_{2k} = T(S_n, T_{2k-1}) - T_{2k-1}$ then

$$\begin{aligned}
\boldsymbol{P}(W_{2k}=n) &= \sum_{m=0}^{n-1} \boldsymbol{P}(W_{2k}=n|\sum_{j=1}^{n-1} h'_{2k+j}=m)\boldsymbol{P}(\sum_{j=1}^{n-1} h'_{2k+j}=m) \\
&< \sum_{m=0}^{n-1} n^* 2\ell\epsilon_0 (1-\alpha)^m \binom{n-1}{m} \gamma^m (1-\gamma)^{n-1-m} \\
&= n^* 2\ell\epsilon_0 (1-\alpha\gamma)^{n-1}
\end{aligned}$$

So let $a = \max\{\frac{1}{1-\kappa}, \frac{1}{1-\delta}, \frac{n^* 2\ell\epsilon_0}{1-\alpha\gamma}\}$ and $b = \max\{1-\kappa, 1-\delta, 1-\alpha\gamma, e^{-\frac{\gamma^2}{2}}\}$. ∎

Lemma 17 *There exists constants $c > 0$, $0 < d < 1$ such that $\boldsymbol{P}(L = 3+2n) < cd^{2n}$.*

Proof: From the Monotonic Proposition and Starting Lemma, $\boldsymbol{P}(W_k < \infty) = 1$ for $k = 1,2,3$ (see previous lemma). Thus $\boldsymbol{P}(L=3) = 1$. Let

$$\begin{aligned}
L_1 &= \sum_{k=2}^{(L-3)/2} \mathbf{1}_{[T_{2k}=T(S_n, T_{2k-1})]} \\
L_2 &= \sum_{k=2}^{(L-3)/2} \mathbf{1}_{[T_{2k}=\widetilde{T}(C_{T_{2k-1}})]}.
\end{aligned}$$

So L_1 counts splits and L_2 counts not-close events. From the definition of W', each of the splits or not-close events counted above is followed by a starting event and this occurs with probability one. Thus $L = 3 + 2(L_1 + L_2)$. From the Monotonic Proposition and the Starting Lemma the probability that X_0 and Y_0 both become monotonic and ϵ_0-close in $3n^*$ steps is greater $\kappa^2\delta$. By the Gluing Lemma the probability of never having a split after T_3 exceeds $\kappa^2\delta\rho$ for some ρ, $0 < \rho < 1$. Similarly, the probability of never having a split and never having a not-close event after T_3 exceeds $\kappa^2\delta\rho\eta$. Let $d^4 = 1 - \kappa^2\delta\rho\eta$. Then

$$\begin{aligned}
\boldsymbol{P}(L-3=2n) &= \sum_{k=1}^{2n} \boldsymbol{P}(2L_1 = 2k|2L_2 = 2(n-k))\boldsymbol{P}(2L_2 = 2(n-k)) \\
&< \sum_{k=1}^{2n} (1-\kappa^2\delta\rho)^k (1-\kappa^2\delta\rho\eta)^{n-k} \\
&< 2n(d^4)^n \\
&< cd^{2n}
\end{aligned}$$

where $c = -(2\log d)^{-1}$. ∎

Remark: In the following we use (see section 3.6)

(i) a random variable Y is *lag geometric* if $Y = C + Z$ where C is a constant and Z is a geometric random variable.

(ii) if X is a random variable then the following are equivalent:

1.) X is stochastically dominated by Y where Y is lag geometric.

2.) $\boldsymbol{P}(X = n) \leq ke^{-n\beta}$

3.) $\boldsymbol{P}(X > n) \leq k'e^{-n\beta'}$

4.) $\boldsymbol{P}(X > nR) \leq k''e^{-n\beta''}$

for positive R and appropriate, positive k, k', k'', β, β', β''.

We can use the last two lemmas to show

Lemma 18 *W' is geometrically bounded.*

Proof: Let $\{\widetilde{W}_k\}$ be a sequence of iid lag geometric random variables with

$$\boldsymbol{P}(\widetilde{W}_k > n) = \begin{cases} 1 & , \quad n \leq \frac{\log a'}{\log \bar{b} - \log b} \\ \bar{b}^n & , \quad \text{else} \end{cases}$$

where $b < \bar{b} < 1$, $a' = ab/(1-b)$ and a, b as above. Let $\widetilde{L}$ be a lag geometric random variable independent of $\{\widetilde{W}_k\}$ with

$$\boldsymbol{P}(\widetilde{L} > n) = \begin{cases} 1 & , \quad n \leq \frac{\log c'}{\log \bar{d} - \log d} \\ \bar{d}^n & , \quad \text{else} \end{cases}$$

where $d < \bar{d} < 1$, $c' = cd/(1-d)$ and c, d as above. Let

$$W'_k = \begin{cases} W_k & , \quad k \leq L \\ \widetilde{W}_k & , \quad \text{else} \end{cases}$$

and $\boldsymbol{w} = (w_0, w_1, \ldots, w_J) \in \{\boldsymbol{N} \cup \{0\}\}^{J+1}$ for all J. Then

$$\begin{aligned}
&\boldsymbol{P}[L > w_0, W_k' > w_k, 1 \le k \le J] = \\
&\quad \boldsymbol{P}[L > w_0]\boldsymbol{P}[W_1' > w_1 | L > w_0] \cdot \\
&\quad \boldsymbol{P}[W_2' > w_2 | W_1' > w_2,\ L > w_0] \cdots \\
&\quad\quad \boldsymbol{P}[W_J' > w_J | W_k' > w_k,\ L > w_0,\ 1 \le k \le J-1] \\
&\le \bar{d}^{\hat{w}_0} \prod_{k=1}^{J} \bar{b}^{\hat{w}_k} \\
&= \boldsymbol{P}[\widetilde{L} > w_0, \widetilde{W}_k > w_k, 1 \le k \le J]
\end{aligned}$$

where

$$\hat{w}_0 = \begin{cases} w_0 & : \quad w_0 > \frac{\log c'}{\log \bar{d} - \log d} \\ 0 & : \quad \text{otherwise} \end{cases}$$

and, for $k > 0$

$$\hat{w}_k = \begin{cases} w_k & : \quad w_k > \frac{\log a'}{\log \bar{b} - \log b} \\ 0 & : \quad \text{otherwise.} \end{cases}$$

From Theorem 1 of [22] there exists random variables L', $\widetilde{L}'$, W_k'', $\widetilde{W}_k'$ such that $L' \le \widetilde{L}'$ and $W_k'' \le \widetilde{W}_k'$ almost surely for $1 \le k \le N$ and $(L', W_1'', \ldots, W_J'')$ is distributed as $(L, W_1', \ldots, W_J')$ and $(\widetilde{L}', \widetilde{W}_1', \ldots, \widetilde{W}_J')$ is distributed as $(\widetilde{L}, \widetilde{W}_1, \ldots, \widetilde{W}_J)$. Then, with probability one,

$$W_1'' + \cdots + W_{L'}'' \le \widetilde{W}_1' + \cdots + \widetilde{W}_{\widetilde{L}'}'.$$

If $\widetilde{W}' = \sum_{k=1}^{\widetilde{L}} \widetilde{W}_k'$, $W'' = \sum_{k=1}^{L'} W_k''$, $\widetilde{W} = \sum_{k=1}^{\widetilde{L}} \widetilde{W}_k$ then the above implies

$$\boldsymbol{P}(W'' > n) \le \boldsymbol{P}(\widetilde{W}' > n). \tag{5.50}$$

Since $\widetilde{W}$ is geometrically bounded and $\widetilde{W}'$ is distributed as $\widetilde{W}$, $\widetilde{W}'$ is geometrically bounded. With W' distributed as W'' the result follows from Equation 5.50. ∎

5.3.5. *Proof of Theorem 16*

Proof of Theorem 16: Let μ be the probability measure of the theorem which is absolutely continuous with respect to Lebesque measure λ. Let the Radon-Nikodym derivative of μ with respect to λ be h where $0 < a_1 \leq h \leq a_2$. Then θ in the Moving Lemma becomes $\prod_{k=0}^{N} a_1 s_k = \left(\frac{a_1 s}{4}\right)^N \prod_{k=0}^{N}(1-\alpha)^k$.

Equation 5.19 becomes

$$\theta(s,\alpha) = \left(\frac{a_1 s}{4}\right)^{N(s,\alpha)} \prod_{k=0}^{N(s,\alpha)} (1-\alpha)^k. \tag{5.51}$$

In the Shrinking Lemma (5.38) becomes

$$\gamma \geq \prod_{k=1}^{\lfloor \ell/(r+1) \rfloor + 1} a_1 s_k. \tag{5.52}$$

The probability δ in the Starting Lemma is changed in a similar fashion.

In the statement of the No-Split Lemma, $2\ell\epsilon$ is replaced by $2\ell a_2 \epsilon$

Let $j = j(n, W')$ be the number of shrinkings after $W' - M$ by the nth n^*-block, $n > W'$. Since there are no more not-close events after W' and since T_L is a starting event with probability one, then

$$\frac{j}{n - W' + M} \geq \frac{\gamma}{2}.$$

Consequently,

$$j \geq \frac{(n - W' + M)\gamma}{2}$$

Thus

$$D(X_{nn^*}, Y_{nn^*}) < (1-\alpha)^j = e^{j \log(1-\alpha)} < e^{\frac{(n-W'+M)\gamma}{2} \log(1-\alpha)} < e^{-\beta'(n-W')}$$

where $\beta' := -\frac{\gamma}{2}\log(1-\alpha)$. For $k > W'$, $D(X_k, Y_k)$ is non-increasing. For a given k then, $\lfloor \frac{k}{n^*} \rfloor \leq \frac{k}{n^*} < \lfloor \frac{k}{n^*} \rfloor + 1$ and

$$D(X_k, Y_k) \le D(X_{n^*\lfloor\frac{k}{n^*}\rfloor}, Y_{n^*\lfloor\frac{k}{n^*}\rfloor}) < e^{-\beta'(\lfloor\frac{k}{n^*}\rfloor - W')}$$
$$< e^{-\beta'(\frac{k}{n^*} - 1 - W')}$$
$$= e^{-\frac{\beta'}{n^*}(k - n^*(W'+1))}$$
$$= e^{-\beta(k-W)}$$

where $\beta := \frac{\beta'}{n^*}$ and $W := n^*(W'+1)$. (Note that W is still lag geometric.) ∎

5.4. The Linear Kohonen Process with Strictly Decreasing Kernel Function

Let K be the kernel function of Theorem 17 and let $\alpha_0 = K(0)$ and $\alpha_r = K(r)$. We will be using α_1 and α_2 below but do not intend that they necessarily refer to $K(1)$ and $K(2)$, respectively.

With the kernel of Theorem 16 the members of the hit neighborhood maintain their relative order. This is not necessarily the case in this new setting as will be seen below. We do have

Observation 7 *Let $i_1, i_2 \in B(i'_n)$, $K(|i'_n - i_1|) := \alpha_1 < K(|i'_n - i_2|) := \alpha_2$ and either $X_n(i_1) < X_n(i_2) < X_n(i'_n)$ or $X_n(i_1) > X_n(i_2) > X_n(i'_n)$. Then i_1, i_2, and i'_n do not change their relative order when i'_n is hit.*

We wish to generalize (5.6) under the conditions of the Moving Lemma as well as prove a Moving Lemma for this setting. Without loss of generality, assume there is a separation s between $X_0(i)$ and the next adjacent particle above i. Let $s_0 = s/4$, $s_n = (1-\alpha_0)^n s_0$ and $\omega_{n+1} \in (X_n(i) + s_n, X_n(i) + 2s_n)$. Further, let $i_1, i_2 \in B(i)$ and let α_1, α_2 be the respective associated shrinking parameters. The claim is that for such ω there are n consecutive hits at i. First, note that ω_n is always closer to $X_{n-1}(i)$ than to $X_{n-1}(i_1)$ if $X_0(i_1) < X_0(i)$. For $n = 1$,

$$\begin{aligned} X_1(i) - X_1(i_1) &= X_0(i) - X_0(i_1) + \alpha_0(\omega_1 - X_0(i)) - \alpha_1(\omega_1 - X_0(i_1)) \\ &= (1-\alpha_1)(X_0(i) - X_0(i_1)) + (\alpha_0 - \alpha_1)(\omega_1 - X_0(i)) \\ &> 0 \end{aligned}$$

since $\alpha_0 \geq \alpha_1$, $\omega_1 > X_0(i)$, $0 < \alpha_1 < 1$ and $X_0(i) > X_0(i_1)$. An induction argument shows that $X_n(i) > X_n(i_1)$. If $X_0(i_1) > X_0(i)$ then

$$\begin{aligned} X_1(i_1) - X_1(i) &= X_0(i_1) - X_0(i) + \alpha_1(\omega_1 - X_0(i_1)) - \alpha_0(\omega_1 - X_0(i)) \\ &= X_0(i_1) - X_0(i) + \alpha_1(\omega_1 - X_0(i) + X_0(i) - X_0(i_1)) \\ &\quad -\alpha_0(\omega_1 - X_0(i)) \\ &= (1-\alpha_1)(X_0(i_1) - X_0(i)) - (\alpha_0 - \alpha_1)(\omega_1 - X_0(i)) \\ &> (1-\alpha_1)s - (\alpha_0 - \alpha_1)2s_0 \\ &> (1-\alpha_0)s \end{aligned}$$

By induction,

$$X_n(i_1) - X_n(i) > (1-\alpha_0)^n s \tag{5.53}$$

and so the claim. With the above and the next two cases we can extend the Moving Lemma.

(1) Let $0 < d_0 < X_0(i) - X_0(i_1) < d_0'$ then

$$X_1(i) - X_1(i_1) = (1-\alpha_1)(X_0(i) - X_0(i_1)) + (\alpha_0 - \alpha_1)(\omega_1 - X_0(i)).$$

So,

$$(1-\alpha_1)d_0 + (\alpha_0 - \alpha_1)s_0 < X_1(i) - X_1(i_1) < (1-\alpha_1)d_0' + (\alpha_0 - \alpha_1)2s_0 \tag{5.54}$$

In general, let

$$d_n = (1-\alpha_1)d_{n-1} + (\alpha_0 - \alpha_1)s_{n-1} \tag{5.55}$$

and

$$d'_n = (1-\alpha_1)d'_{n-1} + (\alpha_0 - \alpha_1)2s_{n-1} \tag{5.56}$$

then by induction

$$(1-\alpha_1)d_{n-1} + (\alpha_0-\alpha_1)s_{n-1} < X_n(i) - X_n(i_1) < (1-\alpha_1)d'_{n-1} + (\alpha_0-\alpha_1)2s_{n-1}.$$

More induction arguments give

$$(1-\alpha_1)^n d_0 + (\alpha_0-\alpha_1)s_0 \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_0)^{n-1-j} < X_n(i) - X_n(i_1) \tag{5.57}$$

and

$$X_n(i) - X_n(i_1) < (1-\alpha_1)^n d'_0 + (\alpha_0-\alpha_1)2s_0 \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_0)^{n-1-j} \tag{5.58}$$

Roughly, then

$$(1-\alpha_0)^n d_0 < X_n(i) - X_n(i_1) < A_1(1-\alpha_r)^{\frac{n}{2}} \tag{5.59}$$

for some constant A_1.

(2) Suppose $0 < d_0 < X_0(i_1) - X_0(i) < d'_0$ then

$$X_1(i_1) - X_1(i) = (1-\alpha_1)(X_0(i_1) - X_0(i)) - (\alpha_0-\alpha_1)(\omega_1 - X_0(i)).$$

So,

$$(1-\alpha_1)d_0 - (\alpha_0-\alpha_1)2s_0 < X_1(i_1) - X_1(i) < (1-\alpha_1)d'_0 - (\alpha_0-\alpha_1)s_0.$$

In general, let

$$d_n = (1-\alpha_1)d_{n-1} - (\alpha_0 - \alpha_1)2s_{n-1} \tag{5.60}$$

and

$$d'_n = (1-\alpha_1)d'_{n-1} - (\alpha_0 - \alpha_1)s_{n-1} \tag{5.61}$$

then by induction

$$(1-\alpha_1)d_{n-1} - (\alpha_0-\alpha_1)2s_{n-1} < X_n(i_1) - X_n(i) < (1-\alpha_1)d'_{n-1} - (\alpha_0 - \alpha_1)s_{n-1}.$$

More induction arguments give

$$(1-\alpha_1)^n d_0 - (\alpha_0 - \alpha_1)2s_0 \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_0)^{n-1-j} < X_n(i_1) - X_n(i) \tag{5.62}$$

and

$$X_n(i_1) - X_n(i) < (1-\alpha_1)^n d'_0 - (\alpha_0-\alpha_1)s_0 \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha)^{n-1-j} \tag{5.63}$$

Roughly, then

$$(1-\alpha_0)^n d_0 < X_n(i_1) - X_n(i) < (1-\alpha_r)^n \tag{5.64}$$

using (5.53).

In general then we have,

Observation 8 *If $j \in B(i)$ and there are n consecutive hits at i*

$$(1-\alpha_0)^n d_0 < |X_n(j) - X_n(i)| < A_1(1-\alpha_r)^{\frac{n}{2}} \tag{5.65}$$

where $|X_0(j) - X_0(i)| > d_0$.

We are now in position to establish the *General Moving Lemma*

Lemma 19 *Let* $\min\{X'_m(k) - X'_m(i) | k \neq i,\ X'_m(k) > X'_m(i)\} > s > 0$ *then there exists an integer* $\bar{N}(K,s)$ *and a real number* $\bar{\theta}(K,s) > 0$ *such that*

$$\boldsymbol{P}\{X_{m+\bar{N}(K,s)}(B(i)) \subset (X_m(i) + \frac{s}{8}, X_m(i) + \frac{3s}{4})\} > \bar{\theta}(K,s) \tag{5.66}$$

If $\min\{X'_m(i) - X'_m(k) | k \neq i,\ X'_m(i) > X'_m(k)\} > s > 0$ *then for the same* $\bar{N}(K,s)$ *and* $\bar{\theta}(K,s)$ *as above*

$$\boldsymbol{P}\{X_{m+\bar{N}(K,s)}(B(i) \subset (X_m(i) - \frac{3s}{4}, X_m(i) - \frac{s}{8})\} > \bar{\theta}(K,s) \tag{5.67}$$

Proof: Let $s_0 = s/4$ and define $s_n(1-\alpha_0)^n$ as above. Let

$$\omega_{n+1} \in I_n := (X_n(i) + s_n, X_n(i) + 2s_n) \tag{5.68}$$

and $1 \leq |p| \leq r$. As before,

$$I_{n+1} \subset I_n \tag{5.69}$$

and

$$I_n \subset I_0 \quad \forall n. \tag{5.70}$$

With the above and the facts that

$$|X_n(i+p) - X_n(i)| \geq (1-\alpha_0)^n s. \tag{5.71}$$

and

$$\omega_{n+1} - X_n(i) < (1-\alpha_0)^n \frac{s}{2},$$

ω_n will always hit i while maintaining a separation greater than $s/2$ with non-neighbor particles above i. (Of course, if the hits are above i we do not have to worry about hitting particles below i since the particles maintain their relative positions.)

To get $B(i)$ adjacent and such that $X_{\bar{N}}(i+p) \in (X_0(i)+(s/8), X_0(i)+(3s/4))$. choose $\bar{N}$ such that

$$A_1(1-\alpha_r)^{\frac{\bar{N}}{2}} < s/16 \tag{5.72}$$

The proof then follows from the above discussion. So

$$\bar{N} > 2\frac{\log(s/16A_1)}{\log(1-\alpha_r)} \tag{5.73}$$

and

$$\bar{\theta} = \left(\frac{s}{4}\right)^{\bar{N}} \prod_{k=0}^{\bar{N}} (1-\alpha_0)^k \tag{5.74}$$

Define

$$\bar{N}(K,s) = \lceil 2\frac{\log(s/16A_1)}{\log(1-\alpha_r)}\rceil + 1 \tag{5.75}$$

and

$$\bar{\theta}(K,s) = \frac{s}{4}\prod_{k=0}^{\bar{N}(K,s)} (1-\alpha_0))^k \quad \blacksquare \tag{5.76}$$

In order to partially extend (5.6) a bit more, consider the following two cases:

(3) Let $0 < t_0 < X_0(i_2) - X_0(i_1) < t_0'$ with $X_0(i_2) < X_0(i)$ and $\alpha_2 > \alpha_1$ then

$$\begin{aligned} X_1(i_2) - X_1(i_1) = (1-\alpha_1)(X_0(i_2) - X_0(i_1)) + (\alpha_2-\alpha_1)(X_0(i) - X_0(i_2)) \\ + (\alpha_2-\alpha_1)(\omega_1 - X_0(i)). \end{aligned}$$

If d_n and d_n' are as in (1) above then

$$(1-\alpha_1)t_0 + (\alpha_2-\alpha_1)d_0 + (\alpha_2-\alpha_1)s_0 < X_1(i_2) - X_1(i_1)$$

and

$$X_1(i_2) - X_1(i_1) < (1-\alpha_1)d_0' + (\alpha_2-\alpha_1)d_0' + (\alpha_2-\alpha_1)2s_0.$$

Induction arguments give

$$\begin{aligned} &X_n(i_2) - X_n(i_1) > \\ &(1-\alpha_1)^n t_0 + (\alpha_2 - \alpha_1) d_0 \sum_{j=0}^{n-1} (1-\alpha_1)^j (1-\alpha_2)^{n-1-j} \\ &+ (\alpha_2 - \alpha_1) s_0 \sum_{j=0}^{n-1} (1-\alpha_1)^j (1-\alpha_0)^{n-1-j} \\ &+ (\alpha_2 - \alpha_1)(\alpha_0 - \alpha_2) s_0 \sum_{\substack{u+v+w=n-2 \\ u,v,w \geq 0}} (1-\alpha_0)^u (1-\alpha_1)^v (1-\alpha_2)^w \end{aligned} \tag{5.77}$$

and

$$\begin{aligned} &X_n(i_2) - X_n(i_1) < \\ &(1-\alpha_1)^n t_0' + (\alpha_2 - \alpha_1) d_0' \sum_{j=0}^{n-1} (1-\alpha_1)^j (1-\alpha_2)^{n-1-j} \\ &+ (\alpha_2 - \alpha_1) 2 s_0 \sum_{j=0}^{n-1} (1-\alpha_1)^j (1-\alpha_0)^{n-1-j} \\ &+ (\alpha_2 - \alpha_1)(\alpha_0 - \alpha_2) 2 s_0 \sum_{\substack{u+v+w=n-2 \\ u,v,w \geq 0}} (1-\alpha_0)^u (1-\alpha_1)^v (1-\alpha_2)^w \end{aligned} \tag{5.78}$$

for $n \geq 2$. Thus,

$$(1-\alpha_0)^n t_0 < X_n(i_2) - X_n(i_1) < A_3 (1-\alpha_r)^{\frac{n}{2}} \tag{5.79}$$

for some constant A_3.

(4) Let $0 < t_0 < X_0(i_1) - X_0(i_2) < t_0'$ with $X_0(i) < X_0(i_2)$ and $\alpha_2 > \alpha_1$ then

$$\begin{aligned} X_1(i_1) - X_1(i_2) = (1-\alpha_1)(X_0(i_1) - X_0(i_2)) + (\alpha_2 - \alpha_1)(X_0(i_2) - X_0(i)) \\ - (\alpha_2 - \alpha_1)(\omega_1 - X_0(i)). \end{aligned}$$

If d_n and d_n' are as in (2) above then

$$(1-\alpha_1) t_0 + (\alpha_2 - \alpha_1) d_0 - (\alpha_2 - \alpha_1) 2 s_0 < X_1(i_1) - X_1(i_2)$$

and

$$X_1(i_1) - X_1(i_2) < (1-\alpha_1) d_0' + (\alpha_2 - \alpha_1) d_0' + (\alpha_2 - \alpha_1) s_0.$$

Induction arguments give

$$\begin{aligned}
&X_n(i_1) - X_n(i_2) > \\
&(1-\alpha_1)^n t_0 + (\alpha_2-\alpha_1)d_0 \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_2)^{n-1-j} \\
&-(\alpha_2-\alpha_1)2s_0 \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_0)^{n-1-j} \\
&(\alpha_2-\alpha_1)(\alpha_0-\alpha_2)2s_0 \sum_{\substack{u+v+w=n-2\\u,v,w\geq 0}} (1-\alpha_0)^u(1-\alpha_1)^v(1-\alpha_2)^w
\end{aligned} \tag{5.80}$$

and

$$\begin{aligned}
&X_n(i_1) - X_n(i_2) < \\
&(1-\alpha_1)^n t_0' + (\alpha_2-\alpha_1)d_0' \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_2)^{n-1-j} \\
&-(\alpha_2-\alpha_1)s_0 \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_0)^{n-1-j} \\
&-(\alpha_2-\alpha_1)(\alpha_0-\alpha_2)s_0 \sum_{\substack{u+v+w=n-2\\u,v,w\geq 0}} (1-\alpha_0)^u(1-\alpha_1)^v(1-\alpha_2)^w
\end{aligned} \tag{5.81}$$

for $n \geq 2$.

Note that we can take $d_0 \geq s = 4s_0$. Using the latter and (5.53), we find that

$$X_n(i_1) - X_n(i_2) >$$

$$(1-\alpha_1)^n t_0 + (\alpha_2-\alpha_1)2s_0 \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_0)^{n-1-j} \tag{5.82}$$

$$> (1-\alpha_0)^n t_0 \tag{5.83}$$

Thus,

$$(1-\alpha_0)^n t_0 < X_n(i_1) - X_n(i_2) < A_4(1-\alpha_r)^{\frac{n}{2}} \tag{5.84}$$

for some constant A_4.

From cases (3) and (4) we have

Observation 9 *Let* $i_1,\ i_2 \in B(i)$, $K(|i-i_1|) := \alpha_1 < K(|i-i_2|) := \alpha_2$ *and either* $X_0(i_1) < X_0(i_2) < X_0(i)$ *or* $X_0(i_1) > X_0(i_2) > X_0(i)$, *then*

$$(1-\alpha_0)^n t_0 < |X_n(i_2) - X_n(i_1)| < A(1-\alpha_r)^{\frac{n}{2}} \tag{5.85}$$

where $A = \max\{A_3, A_4\}$ *and* $|X_0(i_2) - X_0(i_1)| > t_0$.

The latter justifies Observation 7 which implies

Observation 10 *If the kernel function is non-increasing then monotone maps remain monotone under the process.*

In the next two cases we show that the particle order can change under consecutive hits. We use these to extend the Monotonic Proposition. We just consider upper bounds and we will assume from now on that K is strictly decreasing.
(5) Let the conditions be the same as (3) above except that $\alpha_2 < \alpha_1$ then

$$X_1(i_2) - X_1(i_1) = (1-\alpha_1)(X_0(i_2) - X_0(i_1)) - (\alpha_1 - \alpha_2)(X_0(i) - X_0(i_2))$$
$$- (\alpha_1 - \alpha_2)(\omega_1 - X_0(i)).$$

If d_n is as in (1) except that α_1 is replaced by α_2 above then

$$X_1(i_2) - X_1(i_1) < (1-\alpha_1)t_0' - (\alpha_1 - \alpha_2)d_0 - (\alpha_1 - \alpha_2)s_0.$$

An induction argument gives

$$\begin{aligned} &X_n(i_2) - X_n(i_1) < \\ &(1-\alpha_1)^n t_0' - (\alpha_1-\alpha_2)d_0 \sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_2)^{n-1-j} \\ &-(\alpha_1-\alpha_2)s_0\sum_{j=0}^{n-1}(1-\alpha_1)^j(1-\alpha_0)^{n-1-j} \\ &-(\alpha_1-\alpha_2)(\alpha_0-\alpha_2)s_0 \sum_{\substack{u+v+w=n-2 \\ u,v,w\geq 0}} (1-\alpha_0)^u(1-\alpha_1)^v(1-\alpha_2)^w \end{aligned} \tag{5.86}$$

for $n \geq 2$. There is a change in ordering then when n is such that,

$$(1-\alpha_1)^n t_0' - (\alpha_1 - \alpha_2) d_0 n (1-\alpha_1)^{n-1} < 0 \tag{5.87}$$

that is, when

$$n > \frac{(1-\alpha_1)t_0'}{(\alpha_1 - \alpha_2)d_0} \tag{5.88}$$

or, in general,

$$n > \frac{(1-\alpha_r)}{\min_{j_1 \neq j_2} |\alpha_{j_1} - \alpha_{j_2}| d_0} \tag{5.89}$$

(6) Let the conditions be the same as (4) above except that $\alpha_2 < \alpha_1$ then

$$X_1(i_1) - X_1(i_2) = (1-\alpha_2)(X_0(i_1) - X_0(i_2)) - (\alpha_1 - \alpha_2)(X_0(i_2) - X_0(i)) \\ + (\alpha_1 - \alpha_2)(\omega_1 - X_0(i)).$$

If d_n is as in (2) above then

$$X_1(i_1) - X_1(i_2) < (1-\alpha_2)t_0' - (\alpha_1 - \alpha_2)d_0 + (\alpha_2 - \alpha_1)2s_0.$$

An induction argument gives

$$X_n(i_1) - X_n(i_2) < \\ (1-\alpha_1)^n t_0' - (\alpha_1 - \alpha_2) d_0 \sum_{j=0}^{n-1} (1-\alpha_1)^j (1-\alpha_2)^{n-1-j} \\ + (\alpha_1 - \alpha_2) 2s_0 \sum_{j=0}^{n-1} (1-\alpha_1)^j (1-\alpha_0)^{n-1-j} \\ + (\alpha_1 - \alpha_2)(\alpha_0 - \alpha_2) 2s_0 \sum_{\substack{u+v+w=n-2 \\ u,v,w \geq 0}} (1-\alpha_0)^u (1-\alpha_1)^v (1-\alpha_2)^w \tag{5.90}$$

for $n \geq 2$. Continuing from (5.90)

$$
\begin{aligned}
&X_n(i_1) - X_n(i_2) < \\
&\quad (1-\alpha_1)^n t_0' - (\alpha_1 - \alpha_2) d_0 n (1-\alpha_1)^{n-1} \\
&\quad + (\alpha_1 - \alpha_2) 2 s_0 n (1-\alpha_1)^{n-1} \\
&\quad + (\alpha_1 - \alpha_2)(\alpha_0 - \alpha_2) 2 s_0 (n-1)(1-\alpha_2)^{n-2} \\
&< (1-\alpha_1)^n t_0' - (\alpha_1 - \alpha_2) d_0 n (1-\alpha_1)^{n-1} + C s_0 \\
&< (1-\alpha_r)^n [1 - \min_{j_1 \neq j_2} |\alpha_{j_1} - \alpha_{j_2}| d_0 n (1-\alpha_r)^{-1}] + C s_0
\end{aligned} \tag{5.91}
$$

for some constant C. Let

$$
f(n, K, d_0) := (1-\alpha_r)^n [1 - \min_{j_1 \neq j_2} |\alpha_{j_1} - \alpha_{j_2}| d_0 n (1-\alpha_r)^{-1}]
$$

then f has a minimum at

$$
n_{\min} := \frac{1-\alpha_r}{\min_{j_1 \neq j_2} |\alpha_{j_1} - \alpha_{j_2}| d_0} - \frac{1}{\log(1-\alpha_r)} \tag{5.92}
$$

and is negative for

$$
n > \frac{1-\alpha_r}{\min_{j_1 \neq j_2} |\alpha_{j_1} - \alpha_{j_2}| d_0}. \tag{5.93}
$$

To have have i_1 and i_2 change order (i.e. $X_n(i_1) - X_n(i_2) < 0$) take

$$
s_0 = \frac{(1-\alpha_r)^{\lceil n_{\min} \rceil} [1 - \min_{j_1 \neq j_2} |\alpha_{j_1} - \alpha_{j_2}| d_0 \lceil n_{\min} \rceil (1-\alpha_r)^{-1}]}{2C} \tag{5.94}
$$

We are now in a position to redefine the moving event so that when one occurs at a particle then the neighborhood of that particle reorders given that separations exist. This will be useful for our purposes.

Recall that a particle is called open if a separation that is bounded below exists to the left of, to the right of or on both sides of the particle. If the separation exists on both sides of a particle we say it is *isolated.*

As is evidenced by case (5) it is necessary to have separations (bounded below) on both sides in order to determine the number of hits necessary to reorder a particle's neighborhood. That is, it is useful to have the particle isolated. This is not hard to do if the particle is open on one side. Suppose $X_0 \in \Upsilon$ and i is open on the right with a separation greater than s. As usual, let $\omega_1 \in (X_0+(s/4), X_0+(s/2))$ then, after the hit at i, it is separated on the right by an amount greater than $(1-\alpha_0)s$ by (5.65) and on the left by an amount greater than $(K(0)-K(1))(s/4)$ using (5.54) noting that in the latter equation $K(0)-K(1) \leq \alpha_0 - \alpha_1$. Thus we have

Observation 11 *If a particle is open then it can be isolated in one hit.*

We now define a new variety of 'moving event' that uses the latter observation and will reorder neighborhoods (if they are not monotonic). Let particle i be open and hit it as decribed above. Then i is isolated. Let $\bar{d}$ be the minimum of the two separations at i. Let $\bar{s} = \bar{s}(K, \bar{d})$ be the actual, initial 'hit-interval' length once i is isolated—this corresponds to $s/4$ in the original Moving Lemma. Define $\bar{s}$ as in (5.94) in terms of $\bar{d}$. Redefine

$$\bar{N}(K, \bar{d}) = \lceil \max\{2\frac{\bar{s}(K, \bar{d})/4A_1}{\log(1-\alpha_r)}, n_{\min}(K, \bar{d})\}\rceil + 2 \tag{5.95}$$

We will still use the phrase 'moving event' in this new context to denote the above sequence of hits with the understanding (as before) that the number of steps and the interval lengths will in general change over a series of events. As shorthand for such an event at particle i we will write "hHIT$'$ i." Though we will not keep track of the number of hits and the associated probabilities, the situation is typically as follows: There is an initial separation $\bar{d}_0$ at a particle. Denote the subsequent separations by $\bar{d}_n$. Let $\bar{s}_0^0 = \bar{d}_0/4$, $d_1 = \min\{(1-K(0))\bar{d}_0, (K(0)-K(1))\bar{d}_0/4\}$, and, for $n \geq 1$, let $\bar{s}_0^n = \bar{s}_0^n(K, \bar{d}_n)$ be the initial hit-interval length after the nth so-called moving event. Let $\bar{N}_n := \bar{N}(K, \bar{d}_{n-1})$. Then define

$$\bar{d}_n = \min\{\frac{\bar{s}_0^{n-1}}{2}, (1-\alpha)^{\bar{N}_{n-1}} 4\bar{s}_0^{n-1}\} \quad (5.96)$$

Note that when a moving event occurs at particle i, the members of $B(i)$ will order themselves according to their K-values. For example, suppose that $B(i)\setminus\{i\}$ is to the left of i and that i is isolated. A moving event at i (hHIT′ i) will cause $i+p$ and $i-p$ to pair provided the particles exist for $0 < p \leq r$. Further, $i \pm r$ will be left of $i \pm (r-1)$ which will be left of $i \pm (r-2)$ etc.. To describe this phenomenon we say that the particles *α-order* or are *α-ordered.* Also we have

Observation 12 *Let $i_1, i_2 \in B(i)$ with $K(|i-i_1|) = K(|i-i_2|) := \alpha'$ and $X_0 \in \Upsilon$. If there are n consecutive hits at particle i then*

$$|X_n(i_1) - X_n(i_2)| = (1-\alpha')^n |X_0(i_1) - X_0(i_2)|. \quad (5.97)$$

5.4.1. *Getting to Monotonicity*

We thank the referee for informing us that the self-organizing property has been established in [17]. We would also like the say something about the rate at which this occurs. Without much difficulty we can prove

Lemma 20 *Let $X_0 \in \Upsilon$, K be strictly decreasing and $r \geq \ell$ then there exists constants $\bar{\kappa}_1$ and n_4 (depending on K) such that the probability X_{n_4} is monotonic exceeds $\bar{\kappa}_1$.*

Proof: There exists a particle i which is open with a separation greater than $s = 1/(\ell+1)$. hHIT′ i. This causes $B(i)$ to α-order. Consequently, either particle 1 or particle ℓ is on an end (i.e. open). hHIT′ whichever of these is open. The map is now monotonic. ∎

The case when $r < \ell$ requires a little more work.

Lemma 21 *Let $X_0 \in \Upsilon$. There exists constants $\bar{\kappa}_2 > 0$ and n_5 such that the probability X_{n_5} is monotonic exceeds $\bar{\kappa}_2$.*

Proof: We give a recipe using a series of moving events. Given a particle j, define $j_{\min} := \min B(j) = \max\{1, j - r\}$ and $j_{\max} := \max B(j) = \min\{\ell, j + r\}$. Below is a table listing "moving events" with a column titled "Open Particles." We justify calling the particle(s) open with the following:

(i) a particle of hHIT′ particle's neighborhood becomes open due to a combination of a moving event and α-ordering

(ii) a particle not of the hHIT′ particle's neighborhood becomes open when the moving event causes surrounding particles to move away

(iii) a previously open particle remains open by Observation 9 or Observation 8

(iv) a previously open particle remains open by Observation 12

There exists a particle i which is open with a separation greater than $s = 1/(\ell + 1)$. (In the open particles column, we may not list all the open particles.)

No.	hHIT′	Open Particles	Comment
1.)	i	i (iii), $i_{\min}$ and/or $i_{\max}$ (i)	$B(i)$ is α-ordered

Without loss of generality, assume $i_{\max}$ is open above i. Now suppose $i_{\max} = \ell$ i.e.

$i + r \geq 0$.

No.	hHIT′	Open Particles	Comment
2.)	ℓ	$\ell - r$ (i), ℓ (iii)	$B(\ell)$ is α-ordered above $i_{\min} \ldots \ell - r - 1$
3.)	$\ell - r$	$\ell - r$ (iii), ℓ (i) and possibly (iv)	$B(\ell - r)$ α-ordered

Suppose $\ell - 2r < 1$ then all particles are below ℓ.

No.	hHIT′	Open Particles	Comment
4.)	ℓ	1 (ii), $\ell - r - 1$ (ii), $\ell - r$ (i), ℓ (iii)	$\ell - r, \ldots, \ell$ are monotonic increasing above $1, \ldots, \ell - r - 1$
5.)	$\ell - r$	same particles open for various reasons	$1, \ldots, \ell$ monotonic increasing

Suppose $\ell - 2r \geq 1$ and below ℓ.

No. hHIT$'$	Open Particles	Comment
$4'$.) ℓ	$\ell-2r$ (ii), $\ell-r-1$ (ii), $\ell-r$ (i), ℓ (iii)	Particles $\ell-r,\ldots,\ell$ are monotonic above $\ell-2r,\ldots,\ell-r-1$
$5'$.) $\ell-r$	$\ell-2r$ (i), $\ell-r-1$ (ii), $\ell-r$ (iii), ℓ (ii)	$\ell-2r,\ldots,\ell$ monotonic increasing
6.) $\ell-2r$	$\ell-2r$ (iii), $\ell-r+1$ (ii)	$B(\ell-2)$ α-ordered below $\ell-r+1,\ldots,\ell$
7.) $\ell-r+1$	$\ell-2r+1$ (i), $\ell-r+1$ (iii)	Particles $\ell-2r+1,\ldots,\ell$ are monotonic increasing above $(\ell-2r)_{\min},\ldots,\ell-2r$
8.) $\ell-2r+1$	$(\ell-2r)_{\min}$ (ii), $\ell-2r$ (iii), $\ell-2r+1$ (iii)	Particles $(\ell-2r)_{\min},\ldots,\ell$ are monotonic increasing.

If $(\ell-2r)_{\min} \leq 1$ then we are done. If not, an induction argument based on steps 6–8, will carry through to particle 1.

Suppose $\ell-2r \geq 1$ and above ℓ.

No. hHIT$'$	Open Particles	Comment
$4''$.) $\ell-2r$	$\ell-2r$ (iii), $\ell-r$ (i) or (iv), ℓ (ii)	$B(\ell-2r)$ is α-ordered above $\ell-r+1,\ldots,\ell$
$5''$.) $\ell-r$	$\ell-2r$ (i), $\ell-r$ (iii), ℓ (i)	Particles $\ell-2r,\ldots,\ell$ are monotonic decreasing

Now follow steps 6–8 above and induction, if necessary, to get to particle 1. Of course in these cases "monotonic increasing" is replaced by "monotonic decreasing."

Suppose $i + r < \ell$.

No. hHIT′	Open Particles	Comment
2′.) $i + r$	i (i) or (iv), $i + r$ (iii)	$B(i + r)$ is α-ordered above $i_{\min}, \ldots, i - 1$
3′.) i	$i_{\min}$ (i), i (iii), $i + r$ (i)	Particles $i_{\min}, \ldots, i + r$ are monotonic increasing.

If $i = 1$ then the argument follows similarly to the case where ℓ is open. Suppose $i > 1$.

No. hHIT′	Open Particles	Comment
4‴.) $i + r$	$i_{\min}$ (i), $i-1$ (iii), i (iii), $i+r$ (iii)	Particles $i_{\min}, \ldots, i - 1$ are monotonic increasing below $B(i + r)$ (which is α-ordered).
5‴.) $i - 1$	$i_{\min}$ (iii), $i - 1$ (iii), i (iii), $i + r - 1$ (i), $i + r$ (ii)	Particles $i_{\min}, \ldots, i + r - 1$ are monotonic increasing below $i + r, \ldots, (i + r)_{\max}$.
6‴.) $i+r-1$	$i_{\min}$, $i-1$ (i), i (iii), $i+r-1$ (iii), $i + r$ (ii), $(i+r)_{\max}$ (ii)	$i_{\min}, \ldots, (i+r)_{\max}$ are monotonic increasing.

If $i + 2r < \ell$ then use induction following steps 4‴ through 6‴. If $i + 2r > \ell$ and $i - r > 1$ then follow steps 3 through 5 or 3, 4′, 5′, 6, 7, 8 as the case may be. ∎

Once maps are monotone they stay monotone and consequently we can use moving events as defined in the General Moving Lemma. Further separations are now preserved under moving events by (5.65) and (5.85) (see Lemma 7).

5.4.2. *Existence of the Invariant Measure*

We first note that the No-Split Lemma goes through unchanged.

We prove the existence of an invariant measure as before, by showing the process is Feller and that there exists a compact set A' that the process visits infinitely often almost surely and with finite expected return time.

Proposition 5 *Let the environmental measure, μ, and the process be as in Theorem 17, then the process starting in Υ is Feller.*

Proof: The proof is the same as that of Proposition 4 if "α" in that proof is changed to "α_0." ∎

To prove the existence of the set A' we will use the following

Lemma 22 *For every $\epsilon > 0$ there exists m and $\delta' > 0$ so that for every $X_0 \in \Upsilon_m$ the probability that $X_m(i) \le \epsilon$ for all $i \in V$ exceeds δ'.*

Proof: The strategy is show that all the particles can come very close to one endpoint, say zero. Assume, without loss of generality, that X_0 is monotonic increasing.

Case (1): Assume there are no particles in $[0, \epsilon]$. Any hits in $[0, \epsilon]$ will bring $r + 1$ particles closer to the interval. Let $\{z_n, n \ge 0\}$ be defined by $z_1 = 1, z_n = (\alpha_r \epsilon/4) + (1 - \alpha_r) z_{n-1}$ for $n > 0$. Then if there are n consecutive hits in $[0, \epsilon/4]$, $X_n(1 + p) \le z_n$, where $0 \le p \le r$. Now $\lim_{n \to \infty} z_n = \epsilon/4$. Consequently, these particles will be in $[0, \epsilon]$ with a finite number of consecutive hits.

Case (2): Assume there are k particles in $[0, \epsilon]$, $k < l$. The object is to move $k + 1$ below ϵ and then use an induction argument to finish the lemma. We know there is a separation of at least $s_0 = \epsilon/(k + 1)$ between two of the particles or between an end particle and an end of the interval. If the separation is between i and $i+1$ and $k-i-1 < r$, then use the Moving Lemma for the decreasing kernel, hitting $i+1$ on the left and bring $k+1,\ldots,i+1+r$ below ϵ. If the separation is between i and $i+1$ and $k-i-1 > r$ then let $s_0 = s/4$, $\omega_1 \in (X_0(i+1)-2s_0, X_0(i+1)-s_0)$. This guarantees a hit at $i+1$ and by the calculations following Observation 4 (see equations 5.7 and 5.8) we have a separation greater than $\alpha_r s/4$ between $i+r+1$ and $i+r+2$ ($\notin B(i+1)$).

Let $s_j = \alpha_r s_{j-1}/4 = (\alpha_r/4)^j(s/4)$, $\omega_{j+1} \in (X_j(i+1+j)-2s_j, X_j((i+1+j)-s_j)$ for $0 \leq j \leq \lfloor (k-i-2)/r \rfloor$. Induction shows we can reach parti cle k in a finite number of steps and that $\epsilon - X_{\lfloor (k-i-2)/r \rfloor}(k) > (\alpha_r/4)^{[k-i-2]/r}(s/4)$. If $k+1$ is below ϵ then use induction. If not, then use the General Moving Lemma and induction. Determine δ' and m by the worst case, namely, the disallowed case where all particles are at 1 and the modified algorithm where in the case of a hit at coincident particles the lowest index particle is taken as the hit particle. ∎

Proposition 6 *There exists a compact set A' such that $\mathbf{P}(X_n \in A'$ i.o.$) = 1$.*

Sketch of proof: First we note that with a strictly decreasing kernel function, hits create separations between neighborhood members. Let $X_0 \in \Upsilon$ and, without loss of generality, let X_0 be monotonic increasing. Assume particle i is hit on the right and that $i_1, i_2 \in B(i)$ and $X_0(i_2) < X_0(i_1) < X_0(i)$. Let $\omega_1 - X_0(i) \geq s$, $K(|i-i_j|) = \alpha_j$, $j = 1,2$ and $a := \min_{j_1 \neq j_2} |K(j_1) - K(j_2)|$, then

$$
\begin{aligned}
X_1(i_1) - X_1(i_2) &= X_0(i_1) + \alpha_1(\omega_1 - X_0(i_1)) - [X_0(i_2) + \alpha_2(\omega_1 - X_0(i_2))] \\
&= X_0(i_1) - X_0(i_2) + (\alpha_1 - \alpha_2)(\omega_1 - X_0(i_1)) \\
&\quad -\alpha_2(X_0(i_1) - X_0(i_2)) \\
&\geq (\alpha_1 - \alpha_2)s \geq as
\end{aligned}
$$

Thus we can create a lower bound for our separations. Let $\ell' = \lceil \frac{\ell}{r} \rceil$. As per the above lemma, move all the particles into the interval $[0, \frac{1}{4\ell'})$. If there is a moving event in the interval $(1 - \frac{1}{8\ell'}, 1]$ (i.e. let the hit particle be $\ell + 1 = 1$ and the initial hit is in $(1 - \frac{1}{8\ell'}, 1]$), then $B(\ell) \subseteq (1 - \frac{1}{4\ell'}, 1]$. If there is then a moving event in the interval $(1 - \frac{3}{8\ell'}, 1 - \frac{1}{4\ell'})$ then v (i.e. let the hit particle be $\ell - r$ and the initial hit is in $(1 - \frac{3}{8\ell'}, 1 - \frac{1}{4\ell'})$), $B(\ell - r) \subseteq (1 - \frac{2}{4\ell'}, 1 - \frac{1}{4\ell'})$. Continue in this fashion until all particles are moved above $1 - \frac{\ell'}{4\ell'} = \frac{3}{4}$.

The initial distance s between the hit intervals and the particles below $\frac{1}{4\ell'}$ is always $\geq \frac{1}{2}$. After the first hit, then, the distances between particles of the hit neighborhood are bounded below by $\frac{a}{2}$. In that any neighborhood is hit a finite number of times and there are a finite number of neighborhoods hit, the distance between particles has a positive lower bound by equation 5.85 (Observation 9). Now we can define A' as we did A. Since the event described here happens with positive probability in a finite number of steps, the result follows. ∎

5.4.3. *Proof of Theorem 17*

Both the Shrinking Lemma and the Starting Lemma rely on joint hits to monotonic maps. The situation for the general kernel function is much the same as for the simple kernel function of Theorem 16.

Observation 13 *If there is a joint hit at $i'_n = i$ for $X_n, Y_n \in \Upsilon_m$ then*

$$|X_{n+1}(i) - Y_{n+1}(j)| = (1 - K(|i - j|)|X_n(i) - Y_n(j)|. \tag{5.98}$$

for $j \in B(i)$.

Lemma 23 (General Shrinking Lemma) *For $r \geq 1$ there exists $\bar{\epsilon}_0 > 0, \bar{\gamma} > 0$ and $\bar{n}_0$ such that for all X_0 and Y_0 in Υ_{m} with $D(X_0, Y_0) \leq \bar{\epsilon}_0$ and X_0 and Y_0 paired, the probability that after $\bar{n}_0$ steps there is a shrinking exceeds $\bar{\gamma}$.*

Proof: The proof is the same as that of the Shrinking Lemma for Theorem 1 using the above observation and changing "α" to "α_r" and "ϵ_0," "γ" and "n_0" to "$\bar{\epsilon}_0$," "$\bar{\gamma}$" and "$\bar{n}_0$," respectively. The neighborhood function need only be non-increasing. ∎

Lemma 24 (General Glueing Lemma) *For all $\rho < 1$ there exists $\epsilon > 0$ such that if the starting $X_0, Y_0 \in \Upsilon_{\mathrm{m}}$ satisfy $D(X_0, Y_0) \leq \epsilon$, then the probability of a glueing (never having a split in the future) exceeds ρ.*

Proof: The proof is the same as that of the Glueing Lemma for Theorem 16 with the following changes: "α," "γ," "ϵ_0," and "n_0" become "α_r," "$\bar{\gamma}$," "$\bar{\epsilon}_0$," and "$\bar{n}_0$," respectively. Replace "Shrinking Lemma" with "General Shrinking Lemma." ∎

Lemma 25 *Let $1 \leq r \leq \lfloor \frac{\ell-1}{2} \rfloor$. For $X_0 \in \Upsilon$ there exists $\bar{\kappa}_2 > 0$ and $\bar{n}''$ such that the probability of achieving a separation everywhere in $\bar{n}''$ steps exceeds $\bar{\kappa}_2$.*

Proof: The proof follows that of Lemma 9 with the following changes: The moving events are those of the General Moving Lemma. Change "Moving Lemma" to "General Moving Lemma." "$N(\alpha, s)$" and "$\theta(\alpha, s)$" become "$\bar{N}(K, s)$" and "$\bar{\theta}(K, s)$," respectively. "θ_n" becomes "$\bar{\theta}(K, s_{n-1})$" where $s_n = (1 - \alpha_0)^{\bar{N}(K, s_{n-1})} s_{n-1}$ and $s_0 = 1/4(\ell + 1)$. ∎

Lemma 26 (General Starting Lemma) *Let $X_0, Y_0 \in \Upsilon_{\mathrm{m}}$, then there exists $\bar{n}_2$ and $\bar{\delta} > 0$ such that the probability that $D(X_{\bar{n}_2}, Y_{\bar{n}_2}) < \epsilon_0$ and $X_{\bar{n}_2}, Y_{n_2}$ are paired exceeds $\bar{\delta}$.*

Proof: Again the proof is the same as the that of the Starting Lemma with the following changes: "Moving Lemma" becomes "General Moving Lemma" and the moving events are those of the latter. "n''" changes to "$\bar{n}''$." Replace "N_{n+1}" with "$\bar{N}_{n+1}$" and let $\bar{N}_{n+1} = \sum_{k=1}^{n} \bar{N}(K, \bar{s}_{n-1}) + 2\bar{n}''$. "$N(\alpha, s')$" becomes "$N(\bar{K}, s')$" in (5.42). ∎

In the section on the waiting time W, the definitions of $M_{X,n}, \ldots, H_n, T_0, \ldots, W'$ remain unchanged. We list the lemmas of that section. The proofs are similar to their counterparts for Theorem 16; the changes should be clear from the above adaptations. For example, "Monotonic Proposition" is replaced by "Lemma 21."

Lemma 27 *There exists constants $a > 0$, $0 < b < 1$ such that $\boldsymbol{P}(W_k = n) < ab^n$ for all $k \geq 1$.*

Lemma 28 *There exists constants $c > 0$, $0 < d < 1$ such that $\boldsymbol{P}(L = 2n) < cd^{2n}$ and L is always even.*

Lemma 29 *W' is geometrically bounded.*

Proof of Theorem 17: The proof follows that of Theorem 16 using the above changes, e.g., "θ" becomes "$\bar{\theta}$," "γ" becomes "$\bar{\gamma}$."∎

6. CONCLUSIONS AND REMARKS ON THE KOHONEN STRING IN HIGHER DIMENSIONS

6.1. Conclusions

We have shown that in one dimension, both the standard Kohonen process and the Kohonen process with a strictly decreasing neighborhood function are super-stable with exponential rate if the environmental measure μ is absolutely continuous with respect to Lebesgue measure with a density bounded above and away from zero and $1 \leq r \leq \lfloor \frac{\ell-1}{2} \rfloor$ in the case of the standard Kohonen process.

This type of stability is different than that considered in the literature (see, for example, [20]) in that the shrinking/gain parameter α remains fixed instead of going ('slowly') to zero. Our approach is simple (though, at times, tedious) and constructive, and while it parallels that of Burton and Faris in [9], it differs from the approaches of others. Finally, our bound on r in the standard Kohonen process is new, though hinted at previously (see the condition $\mathcal{S}$, equation 2.4 in section 2.2.2).

Though a number of the proofs did not rely on monotonicity and the existence of the absorbing class $\Upsilon_{\mathbf{m}}$, the latter were important in, for example, the Starting (Pairing) Lemma. When considering the Kohonen string in the d-dimensional unit cube, where $d > 1$, the situation is more complex. Although some states appear to be more stable than others, there does not appear to be any absorbing class [19].

6.2. Remarks on the Kohonen String in d Dimensions

6.2.1. *Introduction*

Some of the results for the one-dimensional case extend with little difficulty to higher dimensional case. For example, we give a proof of the Moving Lemma below. With these extensions (especially, the generalization of Lemma 9, the Separation Lemma), it appears one can show super-stability for cases where ℓ is small and $r = 1$. With increasing ℓ and r, the latter method becomes unwieldy.

6.2.2. *Observations*

In the following, let $\|\cdot\|$ be the standard Euclidean norm. We consider the standard Kohonen process—the Kohonen string—in d dimensions. So $V = \{1, 2, \ldots, \ell\}$, $X \in \Upsilon^d$, and $\{\omega_n\} \subset [0,1]^d$.

Recall the

Definition 21 *The* Voronoï tessellation $C_i(x)_{i \in V}$ *of* $x \in \Upsilon$ *is defined by*

$$C_i(x) = \{z \in [0,1]^d | \|x_i - y\| < \|x_k - y\|,\ k \neq i\}, \quad \forall i \in V$$

Thus, if $\omega \in C_i(X)$ then particle i is hit. If there were just two particles then the Voronoï tessellation would consist of two regions divided by the hyperplane orthogonal to the line segment $\overline{X(1)X(2)}$ through $(X(1) + X(2))/2$ or the set of points $x \in [0,1]^d$ satisfying

$$(X(1) - X(2)) \cdot (x - \frac{(X(1) + X(2))}{2}) = 0.$$

Let

$$H_X(i;j) := \{x \in [0,1]^d | (X(i) - X(j)) \cdot (x - \frac{(X(i) + X(j))}{2}) > 0\}.$$

Then

$$C_X(i) = \cap_{j=1}^{\ell} H_X(i;j).$$

Consequently, $C_X(i)$ is convex for all i [31].

All but one of the observations from the one dimensional case generalize to higher dimensions.

Observation 14 *Hitting at i causes a shrinking of the distance between between i and its neighbors by a factor $(1-\alpha)$.*

If $j \in B(i)$ then

$$\|X^1(j) - X^1(i)\| =$$
$$\sqrt{(X_1^1(j) - X_1^1(i))^2 + (X_2^1(j) - X_2^1(i))^2 + \ldots + (X_d^1(j) - X_d^1(i))^2}$$
$$= \sqrt{(1-\alpha)^2[(X_1^0(j) - X_1^0(i))^2 + (X_2^0(j) - X_2^0(i))^2 + \ldots + (X_d^0(j) - X_d^0(i))^2]}$$
$$= (1-\alpha)\|X^0(j) - X^0(i)\|$$

Consequently, equation 5.6 holds in the higher dimensional case, i.e.,

$$(1-\alpha)^n s < \|X^{m+n}(i+p_1) - X^{m+n}(i+p_2)\| < \sqrt{d}(1-\alpha)^n. \qquad (6.1)$$

where $|X^m(i+p_1) - X^m(i+p_2)| > s, \quad 0 \leq |p_1|, |p_2| \leq r.$

Observation 15 *Hits at i do not change the relative position of i and its neighbor(s).*

Consider the convex hull of $X(B(i))$. A hit at i causes the latter to shrink however its 'shape' remains the same. The direction of the line through $X^0(i)$ and $X^0(j)$ is given by $(X^0(j) - X^0(i))/\|X^0(j) - X^0(i)\|$. The direction of the line through $X^1(i)$ and $X^1(j)$ is

$$\frac{(X^1(j)-X^1(i))}{\|X^1(j)-X^1(i)\|} = \frac{(1-\alpha)(X^0(j)-X^0(i))}{\|(1-\alpha)(X^0(j)-X^0(i))\|} = \frac{X^0(j)-X^0(i)}{\|X^0(j)-X^0(i)\|}$$

Observation 16 *If particle i is hit and $\omega \in$ convex hull of $X(B(i))$ then there is an increase in separation between the neighbors of i and particles outside the convex hull of $X(B(i))$.*

Observation 17 *If for the maps X_n, $Y_n \in \Upsilon$ there is a joint hit at i then*

$$|X_{n+1}(i+p) - Y_{n+1}(i+p)| = (1-\alpha)|X_n(i+p) - Y_n(i+p)|$$

where $|p| \leq r$.

Observation 18 *The only way to increase the distance between maps is to have a split.*

6.2.3. *The Moving and No-Split Lemmas*

Given two sets A and B let $\|A - B\| := \inf\{\|x-y\| \ : \ x \in A,\ y \in B\}$. Let $V(A) := \int_A \mu(dx)$ where μ is d-dimensional Lesbegue measure and let S_d be the unit sphere centered at the origin in $\boldsymbol{R}^d$. We can now generalize the Moving Lemma.

Lemma 30 *Let C be a convex body, $C \subset C_{X_m}(i)$, where $\|C - \partial C_{X_0}(i)\| > s > 0$ and $V(C) > 0$. Then there exists an $\widetilde{N}(\alpha, s')$ and $\widetilde{\theta} > 0$ such that*

$$X_{m+\widetilde{N}(\alpha,s')}(B(i)) \subset C + s'S_d \tag{6.2}$$

with probability greater than $\widetilde{\theta}$, where $0 > s' < s$.

Proof: As before, let $m = 0$, $I_0 := C$ and $\omega_{n+1} \in I_n := \alpha\omega_n + (1-\alpha)I_n$. Consequently,

$$I_n \subset I_{n-1} \subset \ldots \subset I_0.$$

Note that $\{I_n\}$ behaves as $\{X_n(B(i))\}$. In particular, the trajectories of elements of I_0 and $X_0(B(i))$ maintain their relative positions. Thus

$$I_n \subset C_{X_n}(i), \quad \forall\, n.$$

Let $x_0 \in I_0$ and $x_{n+1} := \alpha\omega_{n+1} + (1-\alpha)x_n$ then

$$\|X_n(i+p) - x_n\| = (1-\alpha)^n \|X_0(i+p) - x_0\| \quad \forall\, n$$

where $|p| \leq r$ and $i + p \in V$. Since $\{x_n\} \subset C$ then

$$\|X_n(B(i)) - C\| < (1-\alpha)^n \sqrt{d}.$$

Thus, choose n so that

$$(1-\alpha)^n \sqrt{d} < s'$$

Let

$$\widetilde{N}(\alpha, s') := \lceil \frac{\log(s'/\sqrt{d})}{\log(1-\alpha)} \rceil + 1 \tag{6.3}$$

and

$$\widetilde{\theta} := \prod_{n=0}^{\widetilde{N}(\alpha,s')} V(I_n) \quad \blacksquare. \tag{6.4}$$

The No-Split Lemma generalizes as

Lemma 31 *For each $\epsilon > 0$, if the initial X_0 and Y_0 satisfy $D(X_0, Y_0) \leq \epsilon$, are paired and $\min_{i \neq j} \|X_0(i) - X_0(j)\| > h > 0$ then the probability that after one step X and Y split is bounded by $C_0 l\epsilon/h$.*

Proof: See [9].

BIBLIOGRAPHY

[1] Bhattacharya, R. N., and E. C. Waymire, 1990. *Stochastic Processes with Applications.* New York: Wiley-Interscience.

[2] Billingsley, P. 1971. *Weak Convergence of Measures: Applications in Probability.* Philadelphia: SIAM.

[3] Billingsley, P. 1979. *Probability and Measure.* New York: John Wiley & Sons. 1979.

[4] Bouton, Catherine and Pagés, Gilles. 1993. Self-organization and a.s. convergence of the one-dimensional Kohonen algorithm with non-uniformly distributed stimuli. *Stochastic Processes and their Applications* 47:249-274.

[5] Bouton, Catherine and Gilles Pagés. 1994. Convergence in distribution of the one-dimensional Kohonen algorithms when the stimuli are not uniform. *Adv. Appl. Prob.* 26:80-103.

[6] Bouton, Catherine and Gilles Pagés. Weak convergence of the Kohonen Learning Vector Quantization algorithm: the multidimensional case. preprint.

[7] Breiman, Leo. 1992. *Probability.* Philadelphia: SIAM.

[8] Burton, Robert M., and Harold G. Dehling, *Mathematical Aspects of Neural Computing.* preprint.

[9] Burton, Robert M., and William G. Faris, 1996. A Self-Organizing Cluster Process. *The Annals of Applied Probability* 6:1232–1247.

[10] Burton, Robert M.and Gerhard Keller. Stationary measures for randomly chosen maps. preprint.

[11] Chung, K. L. 1974. *A Course in Probability Theory.* Second Edition. New York: Academic Press.

[12] Chung, K. L. 1982. *Lectures from Markov Processes to Brownian Motion.* New York: Springer-Verlag.

[13] Cottrell, Marie and Jean-Claude Fort, 1987. Étude d'un processus d'auto-organization. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 23:1–20.

[14] Doob, J. L. 1953. *Stochastic Processes.* New York: Wiley.

[15] Dudley, R. M. 1989. *Real Analysis and Probability.* Belmont, CA:Wadsworth and Brooks/Cole.

[16] Durrett, R. 1991. *Probability: Theory and Examples.* Pacific Grove, CA: Wadsworth and Brooks/Cole.

[17] Flanagan, J. A. 1994. Self-organizing Neural Networks. PhD Thesis. Ecole Polytechnique Fédérale de Lausanne (Switzerland).

[18] Fort, Jean-Claude and Gilles Pagés. 1993. Sur la convergence *p.s.* de l'algorithme de Kohonen généralisé. *Comptes Rendus Académie des Sciences.* Série I. 317:389-394.

[19] Fort, Jean-Claude and Gilles Pagés. 1995. About the Kohonen Algorithm: Strong or Weak Self-Organization?. *Neural Networks* 8:1–13.

[20] Fort, Jean-Claude and Gilles Pagés. 1995. On the A.S. Convergence of the Kohonen Algorithm with a General Neighborhood Function. *Annals of Applied Probability* 5:1177–1216.

[21] Hecht-Nielsen, R. 1991. *Neurocomputing.* New York: Addison-Wesley.

[22] Kamae, T., U. Krengel and G. L. O'Brien. 1977. Stochastic inequalities on partially ordered spaces. *Annals of Probability* 5:899-912.

[23] Kifer, Yuri. 1986. *Ergodic Theory of Random Transformations.* Boston: Birkhäuser.

[24] Kohonen, Teuvo. 1981. Automatic formation of topological maps of patterns in a self-organizing system. In *Proc. 2nd Scand. Conf. on Image Analysis.* Oja, E., Simula, O. (eds.). 214–220. Espoo: Suomen Hahmontunnistustutkismuksen Seura.

[25] Kohonen, Teuvo. 1981. Hierarchical ordering of vectoral data in a self-organizing process. Report TKK-F-A461. Helsinki University of Technology.

[26] Kohonen, Teuvo. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics.* 43:59–69.

[27] Kohonen, Teuvo. 1982. Analysis of a Simple Self-Organizing Process. *Biological Cybernetics* 44:135–140.

[28] Teuvo Kohonen, 1989 *Self-Organization and Associative Memory* Third Edition. Berlin: Springer-Verlag.

[29] Meyn, S.P and R.L. Tweedie, 1993. *Markov Chains and Stochastic Stability.* London: Springer-Verlag.

[30] Nasrabadi, N.M and Y. Feng, 1988. Vector Quantization of Images Based Upon the Kohonen Self-organizing Feature Maps. *Neural Networks (Suppl. 1 "Abstracts of the 1st annual INNS Meeting")* 1:518–519.

[31] Okabe, A., B. Boots, K. Sugihara. 1992. *Spatial Tessellations: Concepts and Applications of Voronoï Diagrams.* Chichester, England. John Wiley & Sons.

[32] Pollard, David. 1984. *Convergence of Stochastic Processes.* New York: Springer.

[33] Revuz, D. 1975. *Markov Chains.* Amsterdam: North-Holland.

[34] Ritter, H., T. Martinetz, and K. Schulten. 1992. *Neural Computation and Self-Organizing Maps: An Introduction.* Reading, MA: Addison-Wesley.

[35] Sarzeaud, O., Y. Stéphan, C. Touzet. 1990. Application des cartes auto-organizatrices à la génération de maillages aux éléments finis. *Proceedings of Neuro-Nîmes* (EC2 Editions, Nantere, France). 1:81–96.

[36] Ultsch, A. 1990. Kohonen's Self-organizing Feature Maps for Exploratory Data Analysis. *International Neural Network Conference of Paris.* (Kluwer Academic Publ., Dordrecht, Nederlands) 1:305–308.