The Annals of Applied Probability
1997, Vol. 7, No. 3, 679–710

ABOUT THE MULTIDIMENSIONAL COMPETITIVE LEARNING VECTOR QUANTIZATION ALGORITHM WITH CONSTANT GAIN

BY CATHERINE BOUTON AND GILLES PAGÈS

Université Paris 1 and Université Paris 6 (URA 224) and Paris 12

The competitive learning vector quantization (CLVQ) algorithm with constant step $\varepsilon > 0$—also known as the Kohonen algorithm with 0 neighbors—is studied when the stimuli are i.i.d. vectors. Its first noticeable feature is that, unlike the one-dimensional case which has $n!$ absorbing subsets, the CLVQ algorithm is "irreducible on open sets" whenever the stimuli distribution has a path-connected support with a nonempty interior. Then the Doeblin recurrence (or uniform ergodicity) of the algorithm is established under some convexity assumption on the support. Several properties of the invariant probability measure ν^ε are studied, including support location and absolute continuity with respect to the Lebesgue measure. Finally, the weak limit set of ν^ε as $\varepsilon \to 0$ is investigated.

Introduction and main definitions. The origin of the competitive learning vector (or space) quantization essentially comes from the difficulties encountered in data analysis in compressing the information contained in huge sets of data. Two possible ways can be investigated. One is to reduce the dimension of the data state space using, for example, a PCA or some similar technique. The other way is to drastically reduce the size of the data set by building a small number of prototypes, each of them representing a specified subset of the data set. This second approach is called vector quantization.

This paper is devoted to the study of an "on line" quantization process that relies on the minimization of a quadratic criterion. Namely, if the data take their values in a compact set of $\mathbb{R}^d$, if they are i.i.d. and μ distributed and if n is the number of searched prototypes, the potential to be minimized is

$$(1)\quad \forall\, x := (x_1, \dots, x_n) \in (\mathbb{R}^d)^n, \qquad E_n^\mu(x) := \tfrac{1}{2}\int \min_{1\le k\le n} \|x_k - \omega\|^2 \mu(d\omega).$$

It is well known that, under some diffuseness assumption on μ [see (3)], E_n^μ is differentiable at every n-tuple $x \in (\mathbb{R}^d)^n$ having pairwise distinct components and that its gradient has an integral representation with respect to μ (see, e.g., [17], [21] when $d \geq 2$). The algorithm that will be considered in this paper—the competitive learning vector quantization (CLVQ)—is merely the constant step stochastic gradient related to this representation [see (2) and (5)].

In information theory, this potential is called the *distortion* of μ at x. The minimizing problem(s) related to E_n^μ (existence of $\min E_n^\mu$, structure of

Received July 1995; revised December 1996.
AMS 1991 *subject classifications*. Primary 60J20; secondary 60J10, 60F99.
Key words and phrases. Vector quantization, neural networks, Markov chain, uniform ergodicity.

argminloc E_n^μ, estimation of min E_n^μ, minimizing algorithms, etc.) is known as the *quantization* of the distribution μ. These questions have been widely investigated by various authors in the framework of information theory until the 1980's. Let us mention, for example, Conway, Lloyd, Kieffer, Trushkin and Zador (see [13] for a survey). Most results deal with the asymptotics of min E_n^μ as $n \to +\infty$ in any dimension d, the uniqueness of argmin(loc)E_n^μ in the one-dimensional setting under log-concavity assumptions on the density of μ, or the search for some efficient (deterministic) algorithms for tracking the minimum of E_n^μ, still in the one-dimensional setting.

Surprisingly, the CLVQ algorithm is not mentioned in the literature, neither in the constant nor decreasing gain setting. This may be explained by the fact that a stochastic gradient descent is usually not competitive when compared to a deterministic procedure. On the other hand, in a higher-dimensional setting, there seems to be no alternative to the CLVQ for quantization purposes.

In fact, the CLVQ algorithm came to light for the first time in the mid-1980's as the degenerate setting of a self-organizing algorithm based on some "neuro-mimetic" intuition: the Kohonen algorithm (see [15], [16]); this is the reason for its second denomination: Kohonen algorithm with 0 neighbors. Following the usual neural network terminology, one may see the data as stimuli exciting some sensitive units (or neurons). The number n of units is supposed to be small with respect to the number of stimuli. The stimuli excite the unit set one after the other at successive times $t \in \mathbb{N}^*$. At every time t the unit that gives the *best* response to the stimulus ω^t tends to *improve* its response to similar stimuli to come while the others do not modify their sensitivity. After a while, one usually observes that the system reaches some stationary distribution. Beyond its historical interest, this approach provides an interpretation of the CLVQ algorithm with constant step as a simple competition mechanism with local updating. The asymptotic stationarity requires that the updating is not time dependent, that is, is constant. Obviously, our first task will be to establish the existence of such a stationary distribution.

Looking back to the statistical applications, such a feature is useful to track some eventual slow alterations of the statistical characteristics of the stimuli. Thus, one may think of the digitization of images shot in slightly different lights. It also prevents an early "freezing" of the process at some undesired "flat" area of the potential as is observed with its decreasing step version. On the other hand, a small enough updating step parameter provides a stationary distribution generally supposed to concentrate around the local minima of the potential E_n^μ. Subsequently, it is important to study both the structure of the stationary distribution for a given updating step $\varepsilon > 0$ and its asymptotic behavior as the step $\varepsilon \to 0$.

Mathematical definition of the CLVQ algorithm (*constant gain*). Let C be a convex compact set of $\mathbb{R}^d$, $d \geq 1$, endowed with the Euclidean norm $\|\cdot\|$, let $I := \{1, \dots, n\}$ a so-called unit set, let $(\omega^t)_{t\in\mathbb{N}}$ be a sequence of C-valued independent identically distributed i.i.d. random stimuli with distribution μ and let ε be a $(0, 1)$-valued real number. The d-dimensional Kohonen CLVQ algorithm with n units (or simply *points*) is a C^n-valued process $(X^t)_{t\in\mathbb{N}}$ recursively defined by $X^0 := x \in C^n$ and

$$\text{(2)} \qquad \forall\, t \in \mathbb{N},\ \forall\, i \in I, \qquad X_i^{t+1} = \begin{cases} X_i^t - \varepsilon(X_i^t - \omega^{t+1}), & \text{if } i = i_0(\omega^{t+1}, X^t),\\ X_i^t, & \text{otherwise,}\end{cases}$$

where $i_0(\omega, x)$ is the (lowest) index s.t. $\|x_{i_0(\omega,x)} - \omega\| = \min_{1\le j\le n} \|x_j - \omega\|$.

So at time $t+1$, the closest component to the stimuli ω^{t+1} is moved by an ω^{t+1}-centered homothety with ratio $1-\varepsilon \ge 0$; it gets closer to ω^{t+1} than it was before. This finally is the mathematical translation for "the unit that has the *best* response to the stimuli *improves* it while other units stand still."

In neural network terminology, X_i^t is known as the *weight* of unit i at time t and $i_0(\omega, x)$ is called the "winning" unit related to the stimulus ω and the weight vector x. The *gain* (or *step*) parameter of the algorithm is ε_t.

On the other hand, the optimization approach requires some further definitions, namely the notion of Voronoi tessellation of an n-tuple $x \in C^n$.

DEFINITION 1. (a) Let $D := \{x \in C^n \mid x_i \ne x_j,\ i \ne j\}$ be the set of weight vectors with pairwise distinct components. The Voronoi tessellation of C induced by $x := (x_1, \dots, x_n) \in D$ is defined as the family $(C_i(x))_{1\le i\le n}$ of open sets of C defined by

$$\forall\, i \in \{1, \dots, n\}, \qquad C_i(x) := \Big\{\omega \in C \mid \|x_i - \omega\| < \min_{k\ne i} \|x_k - \omega\|\Big\}.$$

(b) When $x \in C^n \setminus D$, the Voronoi tessellation $(C_i(x))_{1\le i\le n}$ may be defined, following the algorithm convention, by

$$C_i(x) := \begin{cases} \Big\{\omega \in C \mid \|x_i - \omega\| < \min_{x_k \ne x_i} \|x_k - \omega\|\Big\}, & \text{if } i := \min\{k \mid x_k = x_i\},\\ \varnothing, \quad \text{otherwise.} \end{cases}$$

(c) For all $x \in C^n$, if $\{k \in I \mid x_k = x_i\} = \{i\}$, i (or x_i) is said to be single. Otherwise, $J := \{k \in I \mid x_k = x_i\}$ is called a "cluster" and one sets $C_J(x) := C_{\min J}(x)$. The obvious notation x_J will be used, too.

The complement of the Voronoi tessellation of an n-tuple is obviously contained in the union of a finite number of hyperplanes. This leads to the following definition on the stimulus distribution μ:

$$\text{(3)} \qquad \mu \text{ is strongly diffuse if for every hyperplane } H,\ \mu(H) = 0.$$

So, if μ is strongly diffuse, for every $x \in C^n$, $\mu(\bigcup_{1\le i\le n} \partial C_i(x)) = 0$. The first obvious consequence is that if μ is strongly diffuse and $x \in C^n$, the winning unit $i_0(\omega, x)$ satisfies

$$\text{(4)} \qquad i_0(\omega, x) := \sum_{i=1}^{n} i\mathbf{1}_{C_i(x)}(\omega), \qquad \mu(d\omega)\text{-a.s.}$$

The second consequence is that (see [17] or [21]) the potential defined by (1) is then differentiable at every $x \in D$ with a gradient given by

$$\nabla E_n^\mu(x) = \left(\int_{C_i(x)} (x_i - \omega)\mu(d\omega)\right)_{1 \le i \le n}. \tag{5}$$

Combining (4) and (5) implies that the CLVQ algorithm (1) definitely is the constant step stochastic gradient descent related to the potential E_n^μ, completed on D by some ad hoc convention.

Equation (2) defines a C^n-valued homogeneous Markov chain since the ω^t's are i.i.d. Its transition probability $(P^\varepsilon(x, dy))_{x \in C^n}$ is defined on bounded Borel functions f by

$$\forall\, x \in C^n, \qquad P^\varepsilon(f)(x) := \int_C f(x - \varepsilon H(\omega, x))\mu(d\omega), \tag{6}$$

where $H_i(\omega, x) = (x_i - \omega)1_{\{i = i_0(\omega, x)\}}$, $1 \le i \le n$. Let $\mathbb{P}_x$ denote the distribution of the whole Markov chain $(X^t)_{t \ge 0}$, starting at $X^0 := x \in C^n$.

For the same reasons that make E_n^μ not differentiable on ${}^cD$, the transition P^ε is never Feller: $P^\varepsilon(f)$ is not continuous at clustered points $x \in {}^cD$. Actually, if μ is strongly diffuse, $P^\varepsilon(f)$ is continuous on D whenever f is. So, although the algorithm is compact-valued, the usual stability methods for nonlinear recursive models (see [9], [20]) cannot work here. Thus, the very existence of a stationary distribution ν^ε is far from being straightforward. Finally the much more technical—but often more powerful—recurrence methods finally work under some suitable assumptions on μ. The same technical difficulties are encountered when dealing with the asymptotic behavior of ν^ε as $\varepsilon \to 0$: the standard methods fail (see, e.g., [3] when $d = 1$) because ∇E_n^μ is not everywhere continuous on C^n. Some smooth enough approximations of E_n^μ are necessary.

Some two-dimensional examples of quantization obtained with the CLVQ. Nowadays, the CLVQ procedure, along with many of its variants, is available in several "neural network method" computer libraries. More recently, multidimensional quantization methods were applied to high-dimensional numerical integration problems (see [21]) and automatic meshing (see [23]) with promising results. Figure 1 displays some quantizations of several stimuli distributions μ on the unit square $[0, 1]^2$, namely

$$\mu = U([0,1]^2), \qquad \mu := \mathcal{N}\left(0; \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}\right) \text{ (suitably truncated)},$$

$$\mu := \mu_1^{\otimes 2}, \qquad \mu_1 := \exp(2) \quad \text{(suitably truncated)},$$

$$\mu := \tfrac{1}{2} U(\Delta_1 \cup \Delta_2) + \tfrac{1}{2} U\big(\mathrm{Disk}\big((\tfrac{1}{2}, \tfrac{1}{2}), \tfrac{1}{4}\big)\big),$$

where Δ_i, $i = 1, 2$ denote the diagonal lines of the unit square.

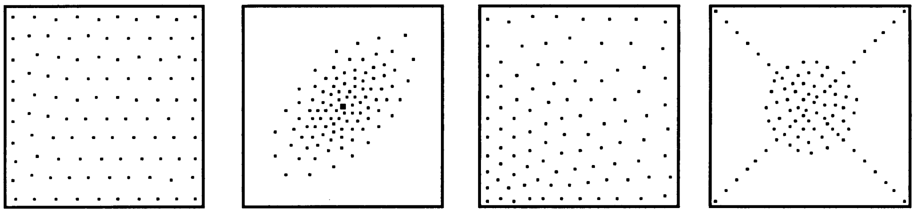

FIG. 1. *μ-quantization of* $[0,1]^2$ *by a* 100*-tuple.*

Existing results on the CLVQ algorithm. Many mathematical results on the Kohonen CLVQ algorithm have been obtained as by-products of general investigations carried out on the "regular" Kohonen algorithm (2 neighbors). They are subsequently related to one-dimensional stimuli: after the seminal paper by Cottrell and Fort [7] exclusively devoted to the Kohonen algorithm with two neighbors and uniformly distributed stimuli, the Kohonen algorithm with 0 neighbors was mentioned with a general distribution μ in [1] and was investigated in [3] along with the 2-neighbor algorithm. The uniqueness of the zero(s) of ∇E_n^μ when the stimuli distribution has a log-concave density has been established independently several times by various methods (see papers by Kieffer [14], Trushkin [24] or more recently by Lamberton and Pagès [19], Cohort [6]).

As mentioned above, the natural field of application of the CLVQ algorithm is the multidimensional setting. Anyway, most one-dimensional results cannot be satisfactorily transposed to vector-valued stimuli: as soon as $d \geq 2$, the CLVQ algorithm behaves radically differently from the one-dimensional case. Thus, as shown in the first part of the paper, the one-dimensional algorithm with n points has $n!$ absorbing sets while, under mild assumptions, such absorbing classes no longer exist whenever $d \geq 2$. Incidentally, we guess that this phenomenon is closely related to the difficulties encountered in defining a satisfactory notion of (self-)organization for the standard multidimensional Kohonen algorithm.

For some partial multidimensional results in the decreasing step case, one may cite Fort and Pagès [11], where stability of grid equilibrium points is investigated. Some conditional and regular a.s. convergence results for an equilibrium are established in [21].

In the constant step case, a very recent work by Burton and Faris [5] investigates the CLVQ algorithm with uniformly distributed stimuli on a hypercube, viewed as a random dynamical system. It is shown that the process has a unique invariant distribution and is superstable. This mainly implies that the distribution of X^t weakly converges with an exponential rate toward the invariant distribution and that two trajectories of the chain get a.s. close, still at an exponential rate.

In this paper, which is an extended version of [4], our approach is different: we consider the case of a rather general—but still compactly supported—stimulus distribution μ. Our aim is to study the property of the algorithm as a homogeneous Markov chain in terms of irreducibility, existence, uniqueness and properties of an invariant distribution ν^ε, stability, uniform ergodicity of the chain, w.r.t. to the structure of the distribution μ and its support. Finally, the asymptotic behavior of the invariant distribution ν^ε as ε approaches 0 is elucidated.

The first section is dedicated to several preliminary results along with the first theorem: under some connectivity assumption, the chain is "irreducible on open sets," that is, visits any open set from any starting value x with positive $\mathbb{P}_x$-probability. Section 2 is devoted to the uniform ergodicity (or Doeblin recurrence) result when the stimulus distribution is locally greater than the Lebesgue measure and the support of μ is a nontrivial convex set. The following sections are dedicated to the properties of the invariant probability measure ν^ε: in Section 3.1 we compare $\operatorname{supp}(\nu)$ and $\operatorname{supp}(\mu)^n$. We stress that, if $\operatorname{supp}(\mu)$ is not a convex set and an invariant probability measure ν^ε does exist, its support is wider than $\operatorname{supp}(\mu)^n$. Section 3.2 is devoted to the absolute continuity properties of ν compared to those of μ. Section 4 deals with the asymptotic behavior of ν^ε as $\varepsilon \downarrow 0$. We show that ν^ε tends to concentrate around the equilibrium set of the corresponding decreasing step algorithm.

We will assume throughout this work that C is the convex hull of $\operatorname{supp}(\mu)$; that is,

$$C = \operatorname{Conv}(\operatorname{supp}(\mu)).$$

We will denote by $\delta_C := \sup_{u,\, v \in C} \|u - v\|$ the diameter of C. Note that $\delta_C = \delta_{\operatorname{supp}(\mu)} := \sup_{u,\, v \in \operatorname{supp}(\mu)} \|u - v\|$. For every $u \in \mathbb{R}^d$, $B(u, r)$ will denote the Euclidean open ball with center u and radius $r > 0$. For every $x := (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$, we set $B_n(x, r) := \prod_{k=1}^n B(x_k, r)$. $\mathring{A}$ will denote the interior of A. Finally, $[v]$ will denote the integral part of $v \in \mathbb{R}_+$.

1. Irreducibility of the chain on open sets.

1.1. *The main result.* The aim of this section is to establish the irreducibility on open sets of the chain. This property is rigorously defined by (8). Roughly speaking, it means that, for every $x \in C^n$, X^t visits with positive $\mathbb{P}_x$-probability any open set that meets $\operatorname{supp}(\mu)^n$.

Let $\tau_A := \min\{s \in \mathbb{N} \mid X^s \in A\}$ denote the hitting time of a set A by $(X^t)_{t \geq 0}$.

THEOREM 1 (Irreducibility on open sets). *Let $d \geq 2$. If the following assumption on* $\operatorname{supp}(\mu)$ *holds*:

$$\operatorname{supp}(\mu) \text{ is path-connected and } \mathring{\operatorname{supp}}(\mu) \neq \varnothing, \tag{7}$$

then $(X^t)_{t \in \mathbb{N}}$ satisfies, for every starting value $x \in C^n$ and any open set $O \subset C^n$,

$$O \cap \operatorname{supp}(\mu)^n \neq \varnothing \quad \Rightarrow \quad \mathbb{P}_x(\tau_O < +\infty) > 0. \tag{8}$$

This result points out the basic difference between the one-dimensional and the multidimensional settings. In dimension 1 (see [3]), the CLVQ algorithm

(0-neighbor Kohonen algorithm) leaves the initial order of the starting value x unchanged: the $n!$ sets $F_\sigma := \{x \in C^n \mid x_{\sigma(1)} < \cdots < x_{\sigma(n)}\}$, σ a permutation of $\{1, \dots, n\}$, are all absorbing sets and make up a partition of D. [A set F is absorbing for $(X^t)_{t\in\mathbb{N}}$ if, for every $x \in F$, $\mathbb{P}_x(\{X^1 \in F\}) = \mathbb{P}_x(\{\forall\, t \in \mathbb{N},\ X^t \in F\}) = 1$.] On the other hand, when $d \geq 2$, the irreducibility on open sets implies the following corollary.

COROLLARY 1. *If $d \geq 2$ and assumption* (7) *hold*, *then*, *for every subset F of C^n*,

$$F \text{ is an absorbing set } \Rightarrow \operatorname{supp}(\mu)^n \subset \overline{F}.$$

Consequently, *if* $\operatorname{supp}(\mu)$ *is a convex set* (*i.e.*, $\operatorname{supp}(\mu) = C$), *$C^n$ is the smallest closed stable set*.

This diverging behavior is essentially due to the obvious topological property of $\mathbb{R}^d$, $d \geq 2$: $\forall\, x \in \mathbb{R}^d$, $\mathbb{R}^d \setminus \{x\}$ is (still) path-connected.

REMARKS. (a) It is obvious from the definition of the algorithm that $\mathring{C}^n$ is an absorbing set (this result fails if $\varepsilon = 1$).

(b) Figure 2 points out that when $\operatorname{supp}(\mu)$ is not a convex set, there are some areas in C^n that cannot be reached from any starting value x belonging to $\operatorname{supp}(\mu)^n$. So, the assumption $O \cap \operatorname{supp}(\mu)^n \neq \varnothing$ cannot be relaxed.

COROLLARY 2. *If assumption* (7) *holds*, *then any n-tuple $x \in \mathring{\operatorname{supp}}(\mu)^n$ is topologically recurrent* (*in the sense of* [20]).

The proof is obvious using the strong Markov property.

Some extensions of Theorem 1. (a) When $d \geq 3$, assumption (7) can be relaxed to "$\operatorname{supp}(\mu)$ is path-connected and contains a small two-dimensional disk."

(b) The path-connectivity assumption of $\operatorname{supp}(\mu)$ can be partly relaxed, too. Suppose, for example, that $d \geq 2$ and $\operatorname{supp}(\mu)$ splits into two path-connected

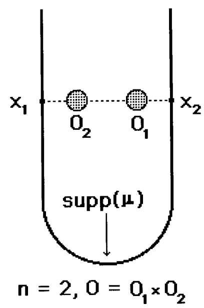

FIG. 2.

components C_1 and C_2. Then the chain is irreducible on open sets iff

$$\exists\, x_1 \in C_1,\ \exists\, x_2 \in C_2 \text{ s.t. } \|x_1 - x_2\| < \max\Bigl(\max_{u\in C_1}\|x_1-u\|, \max_{u\in C_2}\|x_2-u\|\Bigr).$$

1.2. *Proof of Theorem* 1 *when* $x \in \operatorname{supp}(\mu)^n$. The main step of the proof is the following proposition which says that a given component i of X^t can be dragged from $x_i \in C$ into the vicinity of $y_i \in C$ provided that there is some $\operatorname{supp}(\mu)$-valued path connecting these points.

PROPOSITION 1. *Let* $x \in C^n$, $y \in C^n$ *and* $i \in I$ *such that*

$$i = \min\{k \mid x_k = x_i\} \quad \textit{and} \quad y_i \neq x_j \textit{ for every } 1 \le j \le n.$$

Assume that there exists a continuous path $\gamma_i\colon [0,1] \to \operatorname{supp}(\mu)$ *such that* $x_i = \gamma_i(0)$, $y_i = \gamma_i(1)$ *and* $\min_{x_k\neq x_i} \operatorname{dist}(x_k, \gamma_i) > 0$. *Then we have the following.*

(a) *For all* $r > 0$, *there exists* $T \in \mathbb{N}^*$ *such that*

$$\mathbb{P}_x(X_i^T \in B(y_i, r);\ X_k^T = x_k \textit{ for } k \neq i) > 0. \tag{9}$$

(b) *Statement* (*a*) *holds uniformly*. *In fact*, *the time* T *can be chosen locally uniformly with respect to all the components* $x_k \neq x_i$ *of* x (*resp*., *w.r.t.* x *if* x_i *is single*); *that is*, $\forall\, r > 0$, $\exists\, T \in \mathbb{N}$, *s.t. for every* $z \in C^n$ *satisfying*

$$\begin{gathered} z_k \in B\left(x_k, \frac{\rho_i}{16}\right) \ \textit{if } x_k \neq x_i,\ z_k = x_k \textit{ if } x_k = x_i, \\ \left[\textit{resp.}, z_k \in B\left(x_k, \frac{\rho_i}{16}\right),\ 1 \le k \le n,\ \textit{if } x_i \textit{ is single}\right], \end{gathered} \tag{10}$$

$$\mathbb{P}_z(X_i^T \in B(y_i, r);\ X_k^T = z_k \textit{ for } k \neq i) > 0,$$

where $\rho_i := \min_{x_k\neq x_i} \operatorname{dist}(x_k, \gamma_i) > 0$. *Moreover*, $T := T(r, \rho_i, \gamma_i)$ *only depends on* r, ρ_i *and on the uniform continuity modulus of the path* γ_i.

In turn, Proposition 1 relies on the following "safety" lemma.

LEMMA 1. *Let* $x \in C^n$, $i \in I$ *s.t.* $i := \min\{k \mid x_k = x_i\}$ *and* $r_i \in (0, \min_{x_k\neq x_i}\|x_k - x_i\|]$. *For every* $g \in \operatorname{supp}(\mu)$ *s.t.* $(r_i/8) \le \|g - x_i\| \le (r_i/4)$, *for every* $r > 0$, *there exists some* $T \in \mathbb{N}$ *s.t.*, *for every* $z \in C^n$ *whose components satisfy the following*:

$$\begin{cases} \|z_i - x_i\| < \dfrac{r_i}{16}, & \textit{if } x_i \textit{ is single}, \\ z_i = x_i, & \textit{otherwise}, \end{cases} \tag{11(i)}$$

(ii) $z_k = x_i$ *for every* $k \neq i$ *s.t.* $x_k = x_i$,
(iii) $\|z_k - x_k\| < r_i/16$ *if* $x_k \neq x_i$,

one has

$$\mathbb{P}_z(X_i^T \in B(g,r),\ X_k^T = z_k,\ k \neq i) > 0.$$

Furthermore, T *only depends upon* r_i, r (*and* ε).

PROOF. Without loss of generality we may assume that $r < r_i/8$. We will prove by induction on t that the winning indicies at times $s = 1, \dots, t$ are all i on the event $A_t := \{\omega^s \in B(g, r/2),\ 1 \le s \le t\}$ so that

$$X_i^t = (1-\varepsilon)^t z_i + \varepsilon \sum_{s=1}^{t} (1-\varepsilon)^{t-s} \omega^s. \tag{12}$$

($t = 1$). z_i *is single*. Then

$$\|\omega^1 - z_k\| \ge \|x_i - x_k\| - (\|\omega^1 - g\| + \|g - x_i\| + \|z_k - x_k\|) > \frac{5r_i}{8} \quad \text{if } k \neq i,$$

$$\|\omega^1 - z_i\| \le \|\omega^1 - g\| + \|g - x_i\| + \|x_i - z_i\|) < \frac{3r_i}{8},$$

so $\omega^1 \in C_i(z)$.

z_i *is not single*. The above inequalities still hold, respectively, for $z_k \neq z_i$ and $z_k = z_i$. Furthermore, $i = \min\{k \mid z_k = z_i\}$.

($t \Rightarrow t+1$). Let $\omega^{t+1} \in B(g, r/2)$.

z_i *is single*. It follows from (12) that, on A_t,

$$\|X_i^t - g\| \le (1-\varepsilon)^t \left(\frac{r_i}{16} + \frac{r_i}{4}\right) + (1 - (1-\varepsilon)^t)\frac{r}{2} < \frac{5r_i}{16}$$

so $\|X_i^t - \omega^{t+1}\| < 3r_i/8$. On the other hand, if $z_k \neq x_i$,

$$\|X_k^t - \omega^{t+1}\| = \|z_k - \omega^{t+1}\| > \frac{5r_i}{8}$$

as above.

z_i *is not single*. For every k s.t. $z_k \neq x_i$, one has as above

$$\|X_i^t - \omega^{t+1}\| < \frac{3r_i}{8} < \frac{5r_i}{8} < \|z_k - \omega^{t+1}\| = \|X_k^t - \omega^{t+1}\|.$$

If $z_k = z_i$, then $\|X_i^t - \omega^{t+1}\| \le (1-\varepsilon)^t \|x_i - \omega^{t+1}\| + (1 - (1-\varepsilon)^t)(r_i/16) < \|X_k^t - \omega^{t+1}\|$ since $\|X_k^t - \omega^{t+1}\| = \|x_i - \omega^{t+1}\| \ge \|x_i - g\| - \|g - \omega^{t+1}\| > (r_i/8) - (r_i/16) = (r_i/16)$. Finally one derives from (12) that $X_i^t \in B(g, r)$ on A_t provided that $t \ge T := [\ln(1-\varepsilon)/\ln(8r/(5r_i))] + 1$. $\square$

PROOF OF PROPOSITION 1. The continuity of the path γ_i on the unit interval yields a finite γ_i-valued subdivision $(g'_0, g'_1, \dots, g'_{N+2})$ such that $g'_0 = x_i$,

$$t_k := \sup\left\{s \in [t_{k-1}, 1] \,/\, \|\gamma_i(s) - g'_{k-1}\| = \frac{\rho_i}{8}\right\} \wedge 1,$$

$$g'_k := \gamma_i(t_k).$$

Then the slightly modified subdivision $g_k := g'_k$, $0 \le k \le N$, $g_{N+1} := y_i$, satisfy $\rho_i/8 \le \|g_k - g_{k-1}\| \le \rho_i/4$, $1 \le k \le N+1$.

The proposition relies upon $N+1$ applications of the above lemma at the successive points $(x_1, \dots, x_{i-1}, g_k, x_{i+1}, \dots, x_n)$ with $r_i^{(k)} = \rho_i$, $r^{(k)} := \rho_i/16$, $1 \le k \le N$, $r_i^{(N+1)} := \rho_i$, $r^{(N+1)} := r$. It yields $N+1$ times $T_k \in \mathbb{N}^*$, and the announced result holds with $T := T_1 + \cdots + T_{N+1}$ thanks to the Markov property. □

In Proposition 2, we prove that a whole n-tuple can be moved into the neighborhood of a n-tuple y whose components lie in the same connected component of the interior of $\mathrm{supp}(\mu)$.

PROPOSITION 2. *Assume that $d \ge 2$ and assumption* (7) *holds. Let $x, y \in \mathrm{supp}(\mu)^n$ satisfying*:

$$(13)\quad \begin{array}{ll} \text{(i)} & y \in \Gamma^n \textit{ for some connected component } \Gamma \textit{ of } \mathrm{s\mathring{u}pp}(\mu), \\ \text{(ii)} & \forall\, i, j \in I, x_i \ne y_j. \end{array}$$

Then we have the following:

(a) $\forall\, r > 0$, $\exists\, T \in \mathbb{N}^*$ *s.t.* $\mathbb{P}_x(X^T \in B_n(y, r)) > 0$.
(b) *Uniformity. If $x \in D$, $\exists\, \rho > 0$, $\forall\, r > 0$, $\forall\, z \in B_n(x, \rho)$,*

$$\mathbb{P}_z(X^T \in B_n(y, r)) > 0$$

for the same T as in (a).

PROOF. Let x, y be as in assumptions (13). The main problem is to choose a suitable order to move one-by-one each component of X^0 with a positive probability using Proposition 1 at each step. First let $(\gamma_i)_{1 \le i \le n}$ be a family of $\mathrm{supp}(\mu)$-valued continuous paths s.t. $\gamma_i(0) = x_i$ and $\gamma_i(1) = y_i$, $\gamma_i \cap \{y_k \mid k \ne i\} = \varnothing$. This is possible as $\mathrm{supp}(\mu) \setminus \{y_k \mid 1 \le k \le n\}$ is still path-connected: $d \ge 2$ and all the components of y lie in the same connected component Γ of the interior of $\mathrm{supp}(\mu)$ (Γ is locally convex as an open set of $\mathbb{R}^d$).

Consider now γ_1. If, for every $x_k \ne x_1$, $x_k \notin \gamma_1$, then assumption (7) of Proposition 1 is fulfilled. Consequently, there exists some event with positive probability on which X_1^t is moved into a neighborhood of y_1, in T_1 steps. If some component(s) $x_k \ne x_1$ belong to γ_1, we proceed as follows. We define a permutation σ_1 on the indices k of the γ_1-valued components x_k's located between y_1 and x_1 according to their natural encountering order. Thus, we set $\sigma_1(1) := k_1$ if x_{k_1} is the closest component to y_1 on the way to x_1. Finally, σ_1 is conventionally defined on the packed components so as to be increasing on the corresponding set of indices. Reproducing the same process with each path γ_i, we finally define a global priority moving order σ on all the components of x.

Now we may build a new path $\gamma^*_{\sigma(1)}$, joining $x_{\sigma(1)}$ to $y_{\sigma(1)}$ so $\gamma^*_{\sigma(1)} \cap \{x_k \,/\, x_k \ne x_{\sigma(1)}\} = \varnothing$ and $\gamma^*_{\sigma(1)} \cap \{y_k \,/\, y_k \ne y_{\sigma(1)}\} = \varnothing$. We proceed as follows (see Figure 3): since $\mathrm{supp}(\mu) \setminus \{y_k,\ 1 \le k \le n\}$ is path connected, there exists a Γ-valued

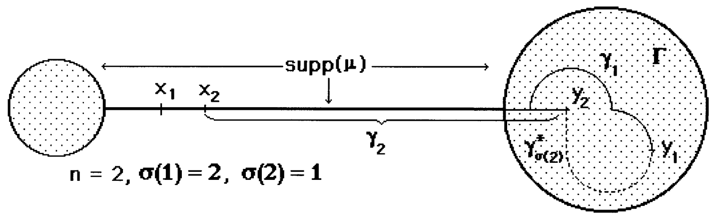

FIG. 3.

path $\tilde{\gamma}_{\sigma(1)}$ connecting y_1 and $y_{\sigma(1)}$. Then $\gamma^*_{\sigma(1)}$ is made up by sticking together the part of γ_1 that connects $x_{\sigma(1)}$ and y_1 with $\tilde{\gamma}_{\sigma(1)}$.

We construct the same way round, a path $\gamma^*_{\sigma(i)}$ for every $\sigma(i)$. Then, we are able to make up an event with positive probability on which the components are moved one by one, with respect to the priority order σ, applying at the ith step Proposition 1 to $(x_{\sigma(i)}, y_{\sigma(i)}, \gamma^*_{\sigma(i)})$. The result finally follows from the Markov property. □

Proposition 3 provides a kind of converse to Proposition 2: it is possible to redispatch the components into given small areas provided that the starting value belongs to the same connected component of $\mathring{\mathrm{supp}}(\mu)$. The proof still relies on Proposition 1 and some priority moving order. The details are left to the reader.

PROPOSITION 3. *Assume that* $d \geq 2$ *and assumption* (7) *hold. Let* $x, y \in \mathrm{supp}(\mu)^n$ *satisfying*

$$\text{(14)(i)} \qquad x \in \Gamma^n \text{ for some connected component } \Gamma \text{ of } \mathring{\mathrm{supp}}(\mu),$$

$$\text{(ii)} \qquad \forall\, i, j \in I, \;\; x_i \neq y_j.$$

Then both statements (a) *and* (b) *of Proposition* 2 *are still valid.*

PROPOSITION 4. *Theorem* 1 *holds when the starting value* x *belongs to* $\mathrm{supp}\,(\mu)^n$.

PROOF. Let $x \in \mathrm{supp}(\mu)^n$. Without loss of generality we may assume that $O = B_n(y, r)$, $y = (y_1, \dots, y_n) \in \mathrm{supp}(\mu)^n$, $r > 0$. Let $\tilde{y} \in D$ s.t. $(x, \tilde{y})$ satisfies assumption (13) of Proposition 2 and $(\tilde{y}, y)$ satisfies (14). The n-tuple $\tilde{y}$ is used as a "transit" point. Proposition 3 yields a time S and $\tilde{\rho} > 0$ s.t.

$$\forall\, z \in B_n(\tilde{y}, \tilde{\rho}), \qquad \mathbb{P}_z(X^S \in B_n(y, r)) > 0.$$

Then Proposition 2 yields a time $T \in \mathbb{N}$ s.t.

$$\mathbb{P}_x(X^T \in B_n(\tilde{y}, \tilde{\rho})) > 0.$$

Finally, the Markov property applied at time T yields

$$\begin{aligned}\mathbb{P}_x(X^{T+S} \in B_n(y,r)) &\geq \mathbb{P}_x(\{X^T \in B_n(\tilde{y},\tilde{\rho})\} \cap \{X^{T+S} \in B_n(y,r)\}),\\ &\geq \mathbb{E}_x\big(\mathbf{1}_{\{X^T \in B_n(\tilde{y},\tilde{\rho})\}} \mathbb{P}_{X^T}(X^S \in B_n(y,r))\big),\\ &> 0,\end{aligned}$$

which completes the proof. □

1.3. *Attractivity of* $\operatorname{supp}(\mu)^n$. At this stage, it remains to prove that X^t can be dragged with positive $\mathbb{P}_x$-probability into $\operatorname{supp}(\mu)^n$ from any starting value $x \in C^n$.

The main ingredient is the following geometrical lemma about the convex hull of $\operatorname{supp}(\mu)$.

LEMMA 2. *Suppose that μ is not a Dirac mass. Every $u \in C$ has at least a projection $v \in \operatorname{supp}(\mu)$, that is, s.t. $\|u - v\| = \min_{w \in \operatorname{supp}(\mu)} \|u - w\|$. Any such projection satisfies $\|u - v\| < \sup_{w \in \operatorname{supp}(\mu)} \|w - v\|$.*

PROOF. One may assume that $u \in C \setminus \operatorname{supp}(\mu)$. The point u is an extremal point of the convex closed ball $\overline{B}(v, \|u - v\|)$ and so $\overline{B}(v, \|u - v\|) \setminus \{u\}$ is still convex. If $\sup_{w \in \operatorname{supp}(\mu)} \|w - v\| \leq \|u - v\|$, then $\operatorname{supp}(\mu) \subset \overline{B}(v, \|u - v\|) \setminus \{u\}$. Hence $C \subset \overline{B}(v, \|u - v\|) \setminus \{u\}$ which is contradictory. □

The idea is now to drag one-by-one the $C \setminus \operatorname{supp}(\mu)$-valued components of a starting value $x \in C^n$, starting by the closest to $\operatorname{supp}(\mu)$. To this end, the stimuli will be picked up near the projection v of this component on $\operatorname{supp}(\mu)$. But, to make sure that an outside component will actually be attracted, all the inside components must have been previously sent to the neighborhood of a far enough point $w \in \operatorname{supp}(\mu)$. The inside components are then driven into $\mathring{\operatorname{supp}}(\mu)$ to increase the number of components that are truly inside $\operatorname{supp}(\mu)$.

First a few technical notations. For every $x \in C^n$, set $j(x) := |\{i \in I \mid x_i \in \operatorname{supp}(\mu)\}|$, $\sigma_1 := \min\{t \mid j(X^t) \geq (j(X^0) + 1) \wedge n\}$ and $\tau_k := \min\{t \mid j(X^t) \geq k\}$, $1 \leq k \leq n\}$ ($\tau_0 := 0$). If θ_t denotes the canonical shift on the X^t's, one has

$$\tau_1 \leq \sigma_1, \qquad \tau_{k+1} = \tau_k + \sigma_1 \circ \theta_{\tau_k}, \qquad 1 \leq k \leq n-1. \tag{15}$$

PROPOSITION 5. *Suppose that assumption* (7) *holds.*

(a) *For every $x \in C^n$*, $\mathbb{P}_x(\sigma_1 < +\infty) > 0$.
(b) *For every $x \in C^n$ and for every $k \in I$*, $\mathbb{P}_x(\tau_k < +\infty) > 0$. *Consequently*

$$\forall\, x \in C^n, \qquad \mathbb{P}_x(\tau_{\operatorname{supp}(\mu)^n} < +\infty) = \mathbb{P}_x(\tau_n < +\infty) > 0.$$

REMARK. In fact, we show that $\mathbb{P}_x(\tau_{\mathring{\operatorname{supp}}(\mu)^n} < +\infty) > 0$.

PROOF. (b) The proof is straightforward by induction, using the strong Markov property, assertion (a) and inequalities (15). Assume that $\mathbb{P}_x(\tau_k <$

$+\infty) > 0$ for some $k \in \{0, \dots, n-1\}$. Then

$$\begin{aligned}\mathbb{P}_x(\tau_{k+1} < +\infty) &= \mathbb{P}_x(\tau_k < +\infty \text{ and } \sigma_1 \circ \theta_{\tau_k} < +\infty)\\ &= \mathbb{E}_x(\mathbf{1}_{\{\tau_k<+\infty\}}\mathbb{P}_{X^{\tau_k}}(\sigma_1 < +\infty))\\ &> 0 \quad \text{using the induction assumption.}\end{aligned}$$

(a) If $x \in \operatorname{supp}(\mu)^n$, $\sigma_1 = 0$; otherwise we set

$$i_0 := \min\Big\{i \,/\, \operatorname{dist}(x_i, \operatorname{supp}(\mu)) = \min_{x_k \notin \operatorname{supp}(\mu)} \operatorname{dist}(x_k, \operatorname{supp}(\mu))\Big\}.$$

Lemma 2 yields $v_{i_0} \in \operatorname{supp}(\mu)$ s.t. $\|x_{i_0} - v_{i_0}\| < \sup_{w\in\operatorname{supp}(\mu)} \|w - v_{i_0}\|$.

Let $J(x) := \{k \in I \mid x_k \in \operatorname{supp}(\mu)\}$ and let $w_k \in \operatorname{supp}(\mu)$, $k \in J(x)$, $r > 0$ s.t. $\bigcup_{k\in J(x)} B(w_k, r) \subset \{w \mid \|w - v_{i_0}\| > \|x_{i_0} - v_{i_0}\|\}$ and the $B(w_j, r)$ have a pairwise empty intersection. Further restrictions on r will be added later on. Proposition 2(a) admits a straightforward extension in which some components of the starting value w may lie outside $\operatorname{supp}(\mu)$ provided that the inside ones lie inside the same connected component of $\operatorname{supp}(\mu)$. It provides a time $T_1 \in \mathbb{N}^*$ s.t. $\mathbb{P}_x(X_j^{T_1} \in B(w_j, r),\ j \in J(x)) > 0$. Roughly speaking, that means that we “hide” the $\operatorname{supp}(\mu)$-valued components far enough from v_{i_0}. So, we may now attract an outside component.

Let $z \in C^n$ s.t. $z_j \in B(w_j, r),\ j \in J(x)$, $z_i = x_i,\ i \notin J(x)$. Let $\rho_0 > 0$ s.t.

$$B(v_{i_0}, \rho_0) \cap C \subset \Bigg\{\omega \in C \mid \operatorname{dist}(\omega, \{x_k \,/\, \|x_k - v_{i_0}\| = \|x_{i_0} - v_{i_0}\|\}) < \operatorname{dist}\Bigg(\omega, \bigcup_{k\in J(x)} B(w_k, r)\Bigg)\Bigg\}$$

and $\rho_0 < \varepsilon/(2+\varepsilon)\|x_{i_0} - v_{i_0}\|$. An easy computation shows that, on the event $\{\omega^1 \in B(v_{i_0}, \rho_0)\}$, the winning index $i^1 \in \{k \,/\, \|x_k - v_{i_0}\| = \|x_{i_0} - v_{i_0}\|\}$ $\mathbb{P}_z$-a.s. and that $B(v_{i_0}, \rho_0) \subset C_{i^1}(X^1)$, $\mathbb{P}_z$-a.s. For notational convenience, this first winning index will still be denoted i_0. So continuing to pick up the ω^t's in $B(v_{i_0}, \rho_0)$ yields a time T_2 s.t. $\mathbb{P}_z(X_{i_0}^{T_2} \in B(v_{i_0}, 2\rho_0),\ X_j^{T_2} = z_j,\ j \neq i_0) > 0$ for every z defined as above. Let γ_0 be a $\operatorname{supp}(\mu)$-valued path satisfying $\gamma_0(0) := v_{i_0}$ and $v'_{i_0} := \gamma_0(1) \in \mathring{\operatorname{supp}}(\mu)$. At the very worst, all the w_j, $j \in J(x)$, may lie on γ_0. Now let $\tilde{v}_j,\ j \in J(x) \cup \{i_0\}$, be some elements satisfying $\tilde{v}_{i_0} = v'_{i_0}$ such that $\tilde{v}_j$ lies in the same connected component of $\mathring{\operatorname{supp}}(\mu)$ as v'_{i_0}. Provided that $2\rho_0$ and r are small enough, a straightforward adaptation of Proposition 2(b) yields a time T_3 s.t., for every $y \in C^n$ satisfying $y_j \in B(w_j, r),\ j \in J(x),\ y_{i_0} \in B(v_{i_0}, 2\rho_0),\ y_i = x_i,\ i \notin J(x) \cup \{i_0\}$,

$$\mathbb{P}_y\Big(X_k^{T_3} \in \underbrace{B(\tilde{v}_k, r)}_{\subset \mathring{\operatorname{supp}}(\mu)},\ k \in J(x) \cup \{i_0\},\ X_k^{T_3} = y_k,\ k \notin J(x) \cup \{i_0\}\Big) > 0.$$

Finally, the Markov property successively applied at times T_ℓ, $1 \le \ell \le 3$, leads to

$$\forall\ x \in C^n \setminus \operatorname{supp}(\mu)^n,\ \exists\ T\,(= T_1 + T_2 + T_3) \in \mathbb{N}^* \text{ s.t. } \mathbb{P}_x(j(X^t) \ge j(x) + 1) > 0.$$
□

END OF THE PROOF OF THEOREM 1. The result follows from Propositions 4 and 5, using some standard Markov arguments and

$$\begin{aligned}\mathbb{P}_x(\tau_O < +\infty) &\ge \mathbb{P}_x(\tau_n < +\infty \text{ and } \tau_n + \tau_O \circ \theta_{\tau_n} < +\infty)\\ &\ge \mathbb{E}_x(\mathbf{1}_{\{\tau_n < +\infty\}} \mathbb{P}_{X^{\tau_n}}(\tau_O < +\infty)) > 0.\end{aligned}$$
□

PROOF OF COROLLARY 1. Assume that F is stable and does not meet $\operatorname{supp}(\mu)^n$. Then $\max_{y\in F} j(y) \le n-1$; let $x_0 \in \operatorname{argmin}_F j$. Proposition 5(b) implies that there is some T s.t. $\mathbb{P}_{x_0}(X^T \in \operatorname{supp}(\mu)^n) > 0$. As $X^T \in F$ $\mathbb{P}_{x_0}$-a.s., $F \cap \operatorname{supp}(\mu)^n \neq \varnothing$.

Now, let $\tilde{x} \in F \cap \operatorname{supp}(\mu)^n$ and $y \in \operatorname{supp}(\mu)^n$. For every open neighborhood O of y, $\mathbb{P}_{\tilde{x}}(\tau_O < +\infty) > 0$, so $F \cap O \neq \varnothing$ hence $y \in \overline{F}$. Finally $\operatorname{supp}(\mu)^n \subset \overline{F}$. □

REMARK. The lack of uniformity with respect to the starting value x in Theorem 1 is the main gap between irreducibility on open sets and classical recurrence properties (see Theorem 3 below).

2. Doeblin recurrence of the Kohonen CLVQ algorithm. The Doeblin recurrence property of a Markov chain (X^t) consists of the convergence with a geometric rate of the $\mathbb{P}_x$-distribution of X^t to a unique invariant probability measure ν, uniformly in x. It is also known as the *uniform ergodicity* property (see [20]). The theorem that will be precisely called upon is recalled below for the reader's convenience. For some background see [8], [22] or, more recently, [20].

THEOREM 2. *Let $(X^t)_{t\in\mathbb{N}}$ be an $(E,\mathcal{E})$-valued homogeneous Markov chain with transition $P(x, dy)$. If $(X^t)_{t\ge 0}$ satisfies*

(16) $(\mathcal{G}) \equiv \exists\,\chi$, *nonnegative measure on* $(E,\mathcal{E})$, $\exists\ t_0 \ge 1$, $\exists\, c > 0$, $\exists\, G \in \mathcal{E}$

s.t. (i) $\chi(G) > 0$, *and* (ii) $\forall\ y \in E,\ \forall\ B \in \mathcal{E},\ B \subset G \Rightarrow \mathbb{P}_y(X^{t_0} \in B) \ge c\chi(B)$, *then $(X^t)_{t\ge 0}$ admits a unique invariant probability measure ν satisfying the following*:

(i) $\nu(B) \ge c\chi(B)$ *for every* $B \in \mathcal{E} \cap G$,
(ii) $\forall\ t \ge 1,\ \forall\ y \in E,\ \forall\ A \in \mathcal{E},\ |\mathbb{P}_y(X^t \in A) - \nu(A)| \le (1 - c\chi(G))^{(t/t_0)-1}$.

The chain is then said to be Doeblin recurrent (*or uniformly ergodic*).

2.1. *The main Doeblin recurrence result.* The Doeblin recurrence of the d-dimensional Kohonen CLVQ algorithm, $d \ge 2$, finally holds under the same type of assumption as in the one-dimensional setting; that is, if μ is locally minorized by the Lebesgue measure λ_{C^n} on C^n. However, one must keep in mind that, as soon as $d \geq 2$, the chain has no strict absorbing set while $n!$ absorbing sets exist when $d = 1$. Namely, we have the following theorem.

THEOREM 3. *Assume that* $\text{supp}(\mu) = C$ *and* $\mathring{C} \neq \varnothing$. *If*

$$\exists \text{ an open set } O \subset C,\ \exists\, \alpha > 0 \quad s.t.\ \mu_{|O} \geq \alpha \lambda_{|O}, \tag{17}$$

then the d*-dimensional Kohonen CLVQ algorithm is Doeblin recurrent.*

REMARKS. (a) Note that under the above assumptions the semigroup P^{ε} of the chain is not Feller (even on D). So the elementary fixed point theorems cannot provide the existence of an invariant probability measure.

(b) Assumption (17) is rather classical for nonlinear regression models like $X^t = f(X^{t-1}) + \omega^t$. It still works in this more nonlinear model $X^{t+1} = F(X^t, \omega^{t+1})$.

Application. The chain being Doeblin recurrent, it is straightforward that, for every bounded Borel function $f \colon C^n \to \mathbb{R}$, the Poisson equation $F - P(F) = f - \nu(f)$ has a bounded solution $F(x) := \sum_k P^k(f - \nu(f))(x)$. Hence the chain is positively recurrent and satisfies a central limit theorem; that is, for every $x \in C^n$,

$$\frac{1}{T} \sum_{t=0}^{T-1} f(X^t) \to_{\mathbb{P}_x\text{-a.s.}} \int_{C^n} f(u)\nu(du)$$

and

$$\sqrt{T}\left(\frac{1}{T} \sum_{t=0}^{T-1} f(X^t) - \int_{C^n} f(u)\nu(du)\right) \to_{\mathscr{D}_{\mathbb{P}_x}} \mathscr{N}(0; \sigma^2(f)),$$

where $\sigma^2(f) := \nu(F^2) - \nu(P(F)^2)$. These results are some well-known Markov background (see [9], page 302, or [20]). We recall them here as some convincing arguments in favor of the implementation of the Kohonen CLVQ algorithm with constant step: they allow some on-line numerical computation of integrals and provide some confidence intervals for the estimates.

2.2. *Proof of Theorem* 3. We will show that the condition ($\mathscr{G}$) of Theorem 2 is fulfilled with $\chi = (\lambda_{|O})^{\otimes n}$. This condition is similar to the one-dimensional one. The proof is divided into several lemmas. The first one is devoted to the existence, for any starting point y, of a "petite recurrent" set G_y around y. As a second step, we show that X^t can be dragged into G_y before a given time T with a positive $\mathbb{P}_x$-probability, uniformly with respect to the starting value $x \in C^n$ (see Proposition 6). In turn, this result relies on a "δ-parting" property (see Lemma 4). Proposition 1 stressed how the hitting time of a n-ball depends on $\min_{i \neq j} \|x_i - x_j\|$ where x denotes the starting point of the chain. So, in Lemma 4, we state that the chain $(X^t)_{t \geq 0}$ reaches $K_\delta := \{z \in C^n / \min_{i \neq j} \|z_i - z_j\| > \delta\}$, $\delta > 0$, with a positive $\mathbb{P}_x$-probability, uniformly with respect to x.

LEMMA 3 (Existence of a "petite recurrent" set). *Let $x \in D$ and $0 < \rho < \frac{1}{4}\min_{i \neq j} \|x_i - x_j\|$. Then we have the following*:

(i) *Borel set* $\forall\, B \subset B_n(x, \varepsilon\rho)$, $\mathbb{P}_x(X^n \in B) \geq \mu^{\otimes n}((B - (1-\varepsilon)x)/\varepsilon)$.

(ii) *The statement* (i) *holds locally uniformly w.r.t. x; namely, if $\eta \in]0, \varepsilon\rho[$ and $r = \varepsilon\rho - \eta$*,

$$\forall\, z \in B_n(x, \eta),\ \forall \textit{ Borel set } B \subset B_n(x, r),\ \mathbb{P}_z(X^n \in B) \geq \mu^{\otimes n}\left(\frac{B - (1-\varepsilon)z}{\varepsilon}\right).$$

PROOF. Let $A := \{\omega^1 \in B(x_1, \rho), \dots, \omega^n \in B(x_n, \rho)\}$. We will show that on A, $\mathbb{P}_z$-a.s., the winning indices are successively $i^1 = 1, \dots, i^k = k, \dots, i^n = n$.
Actually, one has on A:

$$\|z_1 - \omega^1\| \leq \|z_1 - x_1\| + \|x_1 - \omega^1\| \leq (1+\varepsilon)\rho \quad \text{on one hand},$$
$$\|z_k - \omega^1\| \geq \|x_k - x_1\| - \|z_k - x_k\| - \|x_1 - \omega^1\| > 2\rho,$$
$$k \neq 1 \text{ on the other hand.}$$

So, as $(1+\varepsilon)\rho < 2\rho$, the winning index at time 1 on A is $i^1 = 1$ $\mathbb{P}_z$-a.s. Suppose now that we have proved that, on A, $\mathbb{P}_z$-a.s., $i^1 = 1, \dots, i^k = k$, for $k < n$. Hence $X_j^k = (1-\varepsilon)z_j + \varepsilon\omega^j$, $j \leq k$, and $X_j^k = z_j$, $j > k$ $\mathbb{P}_z$-a.s. So, $B(x_j, \rho)$ being a convex set, we have $X_j^k \in B(x_j, \rho)$, $1 \leq j \leq n$. Then

$$\|X_j^k - \omega^{k+1}\| \geq \|x_j - x_{k+1}\| - \|X_j^k - x_j\| - \|x_{k+1} - \omega^{k+1}\| > 2\rho, \quad j \neq k+1,$$
$$\|X_{k+1}^k - \omega^{k+1}\| = \|z_{k+1} - \omega^{k+1}\| < 2\rho, \qquad j = k+1.$$

So, at time n, we have $X^n = (1-\varepsilon)z + \varepsilon W_n$, where $W_n :=^t (\omega^1, \dots, \omega^n)$. Note that $A = \{W_n \in B_n(x, \rho)\}$.

Now, let $z \in B_n(x, \eta)$ and $B \subset B_n(x, r)$, Borel set. We have

$$\{X^n \in B\} \supset \left\{W_n \in \frac{B - (1-\varepsilon)z}{\varepsilon}\right\} \cap A = \left\{W_n \in \frac{B - (1-\varepsilon)z}{\varepsilon}\right\}.$$

Hence $\mathbb{P}_z(X^n \in B) \geq \mu^{\otimes n}((B - (1-\varepsilon)z)/\varepsilon)$. $\square$

LEMMA 4 (Uniform δ-parting of the components). *Suppose that* $\operatorname{supp}(\mu) = C$. *Then*

$$\exists\, c > 0,\ \exists\, \delta > 0,\ \exists\, T \in \mathbb{N}^* \quad s.t.\ \forall\, x \in C^n,\ \mathbb{P}_x(X^T \in K_\delta) > c,$$

where $K_\delta := \{x \in C^n \mid \forall\, i \neq j,\ \|x_i - x_j\| \geq \delta\}$.

PROOF. Since C is a compact set, there exists some $a, b \in C$ such that $\|b - a\| = \delta_C$ (diameter of C), C being a closed convex set $[a, b] \subset C$. For every $u \in C$, let $p(u) \in [0, \delta_C]$ be the coordinate of the orthogonal projection of u on the straight line (a, b) with origin a. Notice that $p(a) = 0$, $p(b) = \delta_C$.

Let $x = (x_1, \dots, x_n) \in C^n$ and let σ be the permutation of $1, \dots, n$ such that $p(x_{\sigma(1)}) \leq \cdots \leq p(x_{\sigma(n)})$ and σ is increasing on the sets $I_i := \{1 \leq k \leq n \,/\, p(x_k) = p(x_i)\}$, $i \in I$. For notational convenience, we will denote $p_0 = 0 \leq$

$\cdots \le p_i(x) := p(x_{\sigma(i)}) \le \cdots \le p_{n+1} = \delta_C$. Notice that two components may have the same projection so that the p_i's may not be distinct.

For every $\delta > 0$ and $0 \le j \le n+1$, let

$$K_\delta^j := \{z \in C^n \mid p_{i+1}(z) - p_i(z) \ge \delta \text{ for at least } j \text{ different indices } i\}.$$

So we have $K_\delta^{n+1} \subset K_\delta$ and $K_\delta^1 = C^n$ as soon as $\delta < \delta_C/(n+1)$, since $\sum_{i=0}^n p_{i+1} - p_i = \delta_C$.

Now, assume we have proved that

$$\forall\, \delta \in (0, \delta_C/(n+1)),\ \forall\, j \in \{0, \dots, n-1\},\ \exists\, \alpha > 0,\ \exists\, t \in \mathbb{N}^* \\ \text{such that } \forall\, x \in K_\delta^j,\ \mathbb{P}_x(X^t \in K_{\delta/4}^{j+1}) > \alpha. \tag{18}$$

Then the lemma follows from a straightforward induction based on the Markov property. Now, to prove assertion (18) we take $x \in K_\delta^j$, and consider the index $i := \min\{k \mid p_{k+1} - p_k > \delta\}$. Without loss of generality we may assume that $i \ge 1$ (when $i = 0$, the proof below works with $l := \max\{k \,/\, p_{k+1} - p_k \le \delta\} \le n$).

Let us consider now the point $\xi \in C$ such that $[x_{\sigma(i)}, \xi]$ is parallel to the diameter $[a, b]$ of C, and $p(\xi) - p_i(x) = \frac{3}{8}\delta$ (see Figure 4).

Then, on the event $A_1 := \{\omega^1 \in B(\xi, r)\}$, with $r < \delta\varepsilon/8$, $\mathbb{P}_x$-a.s., the named index i_1 at time 1 [which may not be equal to $\sigma(i)$ because of the convention] is such that $0 \le p(x_{i_1}) \le p_i$. Thus $X_{i_1}^1 = (1-\varepsilon)x_{i_1} + \varepsilon\omega^1$ and $\mathbb{P}_x$-a.s. on the event $A_2 := \{\omega^1, \omega^2 \in B(\xi, r)\}$; the winning index at time 2 is still i_1. As a matter of fact,

$$\|X_{i_1}^1 - \omega^2\| < (1-\varepsilon)\|x_{i_1} - \omega^2\| + \varepsilon\|\omega^1 - \omega^2\| \le (1-\varepsilon)(3\delta/8 + r) + 2\varepsilon r,$$
$$\|X_k^1 - \omega^2\| > 3\delta/8 - r \quad \text{for every } k \ne i_1,$$

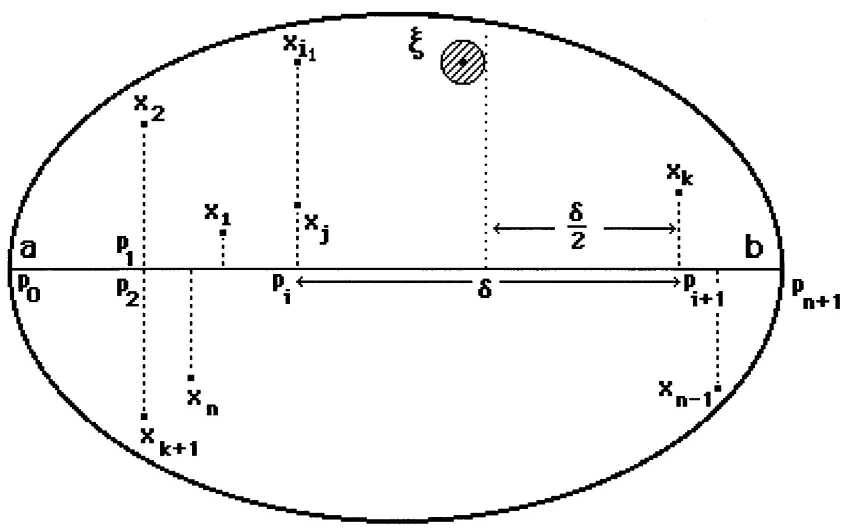

FIG. 4.

and $(1-\varepsilon)(\frac{3}{8}\delta+r)+2\varepsilon r < \frac{3}{8}\delta - r$ since $r < \delta\varepsilon/8$ ($r < 3\delta\varepsilon/8$ would be suitable at this stage of the proof).

Carrying on the process yields that $\mathbb{P}_x$-a.s., on the event $A_s := \{\omega^1, \dots, \omega^s \in B(\xi, r)\}$, the winning indices at times $1, \dots, s$ are all equal to i_1 so that $X^s_{i_1} = (1-\varepsilon)^s x_{i_1} + \varepsilon \sum_{k=1}^{s} (1-\varepsilon)^{s-k} \omega^k$.

Let $t_\delta := \min\{s \,/\, (1-\varepsilon)^s < 8\varepsilon r/3\delta\}$. On the event A_{t_δ}, $\mathbb{P}_x$-a.s., $\|X^{t_\delta}_{i_1} - \xi\| \le \delta/4$, while no other component has been moved. So $p_i + (\delta/4) \le p(X^{t_\delta}_{i_1}) \le p_{i+1} - \frac{1}{2}\delta$, and thus $X^{t_\delta} \in K^{j+1}_{\delta/4}$. Then we have

$$\mathbb{P}_x(X^{t_\delta} \in K_{\delta/4}) \ge \mu(B(\xi, r))^{t_\delta}.$$

Assertion (18) finally derives from the lower semicontinuity of the function $y \mapsto \mu(B(y, r))$, which has a positive lower bound then on the compact set $\mathrm{supp}(\mu) = C$. $\square$

We will prove that the convexity of $\mathrm{supp}(\mu)$ implies that the result in Proposition 2 actually holds uniformly with respect to the starting point $x \in C^n$.

PROPOSITION 6 (Uniform dragging). *Suppose* $\mathrm{supp}(\mu) = C$ *and* $\mathring{C} \neq \varnothing$. *Let* $y \in C^n$, $r > 0$; *then*

$$\exists\, \beta > 0,\ \exists\, T \in \mathbb{N}^* \quad \textit{such that } \forall\, x \in C^n,\ \mathbb{P}_x(X^T \in B_n(y, r)) > \beta. \tag{19}$$

PROOF. Since $D \cap B_n(y, r)$ is an open set and $\overline{D} = C^n$, there is some $y' \in D$ and $r' > 0$ such that $B_n(y', r') \subset D \cap B_n(y, r)$. So, in order to establish (19), we may assume w.l.o.g. that $y \in D$. Let $\delta := \frac{1}{2} \min_{i \neq j} \|y_i - y_j\|$ (so $y \in K_\delta$).

Now, using the δ-parting lemma (Lemma 4) and the Markov property, we may assume that $x \in K_\delta$.

In order to apply Proposition 1, we will build now a family of C-valued paths $(\gamma_i)_{i \in I}$ that connect x_i and y_i, have a uniformly bounded length L with respect to $x \in K_\delta$ and satisfy $\rho_i := \min_{x_k \neq x_i} d(x_k, \gamma_i) > 0$ (see Proposition 1). Let $\delta_1 \in]0, \delta/2[$. Assume for a while that

$$\forall\, k \in \{1, \dots, n\}, \qquad y_k \notin \bigcup_{k=1}^{n} B(x_k, \delta). \tag{20}$$

For every $i \in \{1, \dots, n\}$, we build γ_i by modifying the straight line $[x_i, y_i]$ as follows: we cancel the possible intersections of $[x_i, y_i]$ with the balls $B(x_j, \delta_1), B(y_k, \delta_1),\ 1 \le j,\ k \le n,\ j,\ k \neq i$ and replace them by an arc of the circle residing in the frontier of the intersecting ball. There are at most $2n$ intersections. Hence the length of γ_i is bounded by $L := \delta_C + 2(n-1)(\pi - 2)\delta_1$.

Without loss of generality we may assume that $0 < r < \delta_1/8$. Then one checks that $r < \rho_1$, so Proposition 1(b) applied to x_1 and y_1 yields a time $T \in \mathbb{N}^*$ and a constant $c > 0$, depending only upon L, δ and r, such that

$$\mathbb{P}_x(X^T_1 \in B(y_1, r);\ X^T_j = x_j,\ j \neq 1) > c.$$

As $r < \rho_2$, we may apply again Proposition 1(b), this time to x_2, y_2. So we get, for some time T, $\forall\, z \in C^n$ with $z_1 \in B(y_1, r)$ and $z_k = x_k,\ k \neq 1$,

$$\mathbb{P}_z(X_2^T \in B(y_2, r);\ X_k^T = z_k,\ k \neq 1) > c.$$

Applying Proposition 1(b) to the x_j, y_j's, $1 \le j \le k$ yields, for all $z \in C^n$, s.t. $z_1 \in B(y_1, r), \dots, z_k \in B(y_k, r),\ z_j = x_j,\ j > k$,

$$\mathbb{P}_z(X_{k+1}^T \in B(y_{k+1}, r);\ X_j^T = z_j,\ j \neq k+1) > c.$$

Finally, the Markov property at times $T, 2T, \dots, nT$ yields that for every x, y satisfying (20),

$$\mathbb{P}_x(X^{nT} \in B_n(y, r)) \ge c^n.$$

Let us turn now to the case where $x, y \in K_\delta$ but do not satisfy assumption (20). We will show that, provided δ is small enough, there is some "transit" n-tuple $z \in K_\delta$ such that both x, y and y, z satisfy (20). So, we may assume w.l.o.g. that $\delta < \delta_C/3n$. The diameter of $B^{x,y} := \bigcup_{j=1}^n (B(x_j, \delta) \cup B(y_j, \delta))$ is not greater than $2n\delta$. As $2n\delta < \frac{2}{3}\delta_C < \delta_C - \delta$, there exists some $z_1 \in C$ s.t. $d(z_1, B^{x,y}) > \delta$.

Carrying on the process in exactly the same way yields by induction $z_2, \dots, z_n$, s.t.

$$\forall\, k \in \{0, \dots, n-1\}, \qquad d(z_{k+1}, B(z_1, \delta) \cup \cdots \cup B(z_k, \delta) \cup B^{x,y}) > \delta.$$

One easily checks that the n-tuple $z = (z_1, \dots, z_n) \in K_\delta$ satisfies the above requirement for a transit point. □

End of the Proof of Theorem 3. Let $x_0 \in D \cap O^n$ and $\rho_0 < \frac{1}{4}\min_{i\neq j} \|x_i - x_j\|$ s.t. $B_n(x_0, \rho_0) \subset O^n$. Let η_0, r_0 be as defined in Lemma 3(ii):

$$\forall\, z \in B_n(x_0, \eta_0),\ \forall\, B \subset B_n(x_0, r_0), \qquad \mathbb{P}_z(X^n \in B) \ge \mu^{\otimes n}\left(\frac{B - (1-\varepsilon)z}{\varepsilon}\right).$$

As $(B - (1-\varepsilon)z)/\varepsilon \subset (B_n(x_0, r_0) - (1-\varepsilon)z)/\varepsilon \subset B_n(x_0, \rho_0)$, one has

$$\mathbb{P}_z(X^n \in B) \ge \alpha^n \lambda^{\otimes n}\left(\frac{B - (1-\varepsilon)z}{\varepsilon}\right) \ge \left(\frac{\alpha}{\varepsilon}\right)^n \lambda^{\otimes n}(B).$$

Now, Proposition 6 yields a time T_0 and $\beta_0 > 0$ s.t.

$$\forall\, x \in C^n, \qquad \mathbb{P}_x(X^{T_0} \in B_n(x_0, \eta_0)) \ge \beta_0.$$

Then the Markov property at time T_0 leads to

$$\begin{aligned}\forall\, x \in C^n, \qquad \mathbb{P}_x(X^{T_0+n} \in B) &\ge \mathbb{E}_x\big(\mathbf{1}_{\{X^{T_0} \in B_n(x_0,\eta_0)\}} \mathbb{P}_{X^{T_0}}(X^n \in B)\big),\\ &\ge \beta_0\left(\frac{\alpha}{\varepsilon}\right)^n \lambda^{\otimes n}(B).\end{aligned}$$

Condition $(\mathscr{G})$ of Theorem 2 is thus satisfied with $\chi = \lambda^{\otimes n}$, $t_0 = T_0 + n$ and $G = B_n(x_0, r_0)$. This completes the proof. □

REMARK. Whenever both existence and uniqueness of ν hold, then ν is symmetric; that is, for any permutation $\sigma \in \{1, \dots, n\}$, $\nu = \nu_\sigma$, where ν_σ is the image of ν by the application φ_σ from D onto D defined by $x \mapsto (x_{\sigma(1)}, \dots, x_{\sigma(n)})$. The transition P^ε obviously satisfies $P^\varepsilon(f) = P^\varepsilon(f \circ \varphi_\sigma)$ for every Borel bounded function f whose support is contained in D. So ν_r is also an invariant probability measure, uniqueness providing the expected equality.

If uniqueness assumption fails, there exists at least one symmetrical invariant probability measure $\nu^{\mathrm{sym}} := (1/n!) \sum_\sigma \nu_\sigma$.

3. About the invariant probability measure. This section is devoted to some investigations about the structure of an invariant probability measure ν for the chain. Following Theorem 3, the existence and uniqueness of ν is granted, provided that $\mathrm{supp}(\mu)$ is a convex set with nonempty interior and μ is locally minorized by the Lebesgue measure. However, our conjecture is that ν does exist at least whenever assumption (17) holds and $\overline{\mathring{\mathrm{supp}}(\mu)} = \mathrm{supp}(\mu)$ (i.e., $\mathrm{supp}(\mu)$ has no tentacle).

3.1. *Location of the support of ν.*

THEOREM 4. *Assume that assumption* (7) *holds* [$\mathrm{supp}(\mu)$ *is path-connected and* $\mathring{\mathrm{supp}}(\mu) \neq \varnothing$] *and ν is an invariant probability measure for* $(X^t)_{t\geq 0}$. *Then ν satisfies the following*:

(a) $\nu(D) = 1$ *and* $(\mathrm{supp}(\mu))^n \subset \mathrm{supp}(\nu) \subset C^n$;
(b) *if* $\mathrm{supp}(\mu)$ *is nonconvex*, *then* $(\mathrm{supp}(\mu))^n \subsetneq \mathrm{supp}(\nu)$.

REMARKS. In terms of applications, Theorem 4 shows that the algorithm with constant step necessarily explores the vicinity of the tracked (local) minima. No early freezing can occur as with the usual decreasing step algorithm.

PROOF OF THEOREM 4(a). We first notice that D is a stable set for the Markov chain (X^t); that is,

$$\forall\, x \in D, \qquad \mathbb{P}_x(X^1 \in D) = 1. \tag{21}$$

In fact, if the winning index at time 1 is i, then $X_i^1 = (1-\varepsilon)x_i + \varepsilon\omega^1$ and $X_k^1 = x_k$ for $k \neq i$. If $X_i^1 = x_j$ for some $j \neq i$, then $(1-\varepsilon)\|x_i - \omega^1\| = \|x_j - \omega^1\|$. Such an equality cannot hold, since i is the winning index.

Moreover, (21) and the Markov property straightforwardly imply that

$$\forall\, x \in D, \qquad \mathbb{P}_x(\forall\, t \geq 1,\ X^t \in D) = 1.$$

Now, let $O \subset D \cap (\mathrm{supp}(\mu))^n$, where O is a nonempty open set. Theorem 1 yields

$$\forall\, z \in C^n,\ \exists\, T_z \in \mathbb{N} \quad \text{s.t. } \mathbb{P}_z(X^{T_z} \in O) > 0.$$

The Markov property then implies, for every $z \in C^n$ and $t \in \mathbb{N}$,

$$\mathbb{P}_z(X^{T_z+t} \in D) \geq \mathbb{E}_z\big(\mathbf{1}_{\{X^{T_z} \in O\}} \mathbb{P}_{X^{T_z}}(X^t \in D)\big) = \mathbb{P}_z(X^{T_z} \in O) > 0,$$

and so

$$\forall\ z \in C^n, \qquad \liminf_t \mathbb{P}_z(X^t \in D) > 0. \tag{22}$$

Finally, the P^ε-invariance of the measure ν and Fatou's lemma imply

$$\nu(D) = \int_{C^n} \mathbb{P}_x(X^t \in D)\nu(dx) \geq \nu(D) + \int_{{}^cD} \underbrace{\liminf_t \mathbb{P}_x(X^t \in D)}_{>0} \nu(dx). \tag{23}$$

Hence $\nu({}^cD) = 0$. So $\nu(D) = 1$ and $\operatorname{supp}(\nu) \subset \overline{D} \subset C^n$. Now, let $y \in (\operatorname{supp}(\mu))^n$ and $r > 0$. According to Theorem 1,

$$\forall\ x \in \operatorname{supp}(\nu),\ \exists\ T_x \in \mathbb{N} \quad \text{s.t. } \mathbb{P}_x(X^{T_x} \in B_n(y, r)) > 0. \tag{24}$$

So there is some $\tilde{T} \in \mathbb{N}$ s.t. $\nu(\{x \,/\, \mathbb{P}_x(X^{\tilde{T}} \in B_n(y, r)) > 0\}) > 0$. Now, $(P^\varepsilon)^{\tilde{T}}$ being invariant by ν,

$$\begin{aligned} \nu(B_n(y, r)) &= \int \mathbb{P}_z(X^{\tilde{T}} \in B_n(y, r))\nu(dz) \\ &\geq \int_{\{z/\mathbb{P}_z(X^{\tilde{T}} \in B_n(y,r))>0\}} \underbrace{\mathbb{P}_z(X^{\tilde{T}} \in B_n(y, r))}_{>0} \nu(dz) > 0. \end{aligned} \tag{25}$$

Hence $\nu(B_n(y, r)) > 0$ for any $r > 0$; that is, $y \in \operatorname{supp}(\nu)$.

The proof of item (b) will be derived as a consequence of the main result of the next subsection. □

3.2. *About* supp(ν) *when* supp(μ) *is not a convex set.* For every $\varepsilon \in (0, 1)$, let $D_\varepsilon := \{\varepsilon \sum_{s=0}^t \delta_s(1 - \varepsilon)^s,\ t \in \mathbb{N},\ \delta_s \in \{0, 1\},\ 0 \leq s \leq t\}$ and let $\overline{D}_\varepsilon$ be its closure. The following lemma provides a description of the set D_ε.

LEMMA 5. (i) *If* $\varepsilon \leq \frac{1}{2}$, D_ε *is everywhere dense in* $[0, 1]$; *that is*, $\overline{D}_\varepsilon = [0, 1]$.
(ii) *If* $\varepsilon > \frac{1}{2}$, D_ε *is nowhere dense in* $(0, 1)$; *that is*, $\lambda(\overline{D}_\varepsilon) = 0$.

REMARK. The set $D_{2/3}$ is in fact the Cantor set.

PROOF. (i) $\varepsilon \leq \frac{1}{2}$. Let $z \in [0, 1]$. One sets $\delta_0 = 0$ if $z < \varepsilon$, $\delta_0 = 1$ if $z \geq \varepsilon$. Then $0 \leq z - \varepsilon\delta_0 \leq \varepsilon \mathbf{1}_{\{\delta_0=0\}} + (1-\varepsilon)\mathbf{1}_{\{\delta_0=1\}} \leq 1 - \varepsilon$ since $\varepsilon \leq 1/2$. Assume now that there exists some $\delta_0, \dots, \delta_t \in \{0, 1\}$ such that $0 \leq z - \varepsilon(\delta_0 + \cdots + \delta_t(1 - \varepsilon)^t) \leq (1 - \varepsilon)^t$. Setting $\delta_{t+1} = 0$ if $((z - \varepsilon(\delta_0 + \cdots + \delta_t(1 - \varepsilon)^t))/(1 - \varepsilon)^t) < \varepsilon$, $\delta_{t+1} = 1$ otherwise implies that

$$\begin{aligned} 0 &\leq z - \varepsilon(\delta_0 + \cdots + \delta_{t+1}(1 - \varepsilon)^{t+1}) \\ &\leq \begin{cases} \varepsilon(1 - \varepsilon)^{t+1} \leq (1 - \varepsilon)^{t+2}, & \text{if } \delta_{t+1} = 0, \\ (1 - \varepsilon)^{t+1}(1 - \varepsilon\delta_{t+1}) = (1 - \varepsilon)^{t+2}, & \text{if } \delta_{t+1} = 1. \end{cases} \end{aligned}$$

Finally,

$$\forall\, t \in \mathbb{N}, \qquad 0 \le z - \varepsilon(\delta_0 + \cdots + \delta_t(1-\varepsilon)^t) \le (1-\varepsilon)^{t+1} \longrightarrow 0 \text{ as } t \to +\infty,$$

which completes the proof of (i).

(ii) If $\varepsilon > \frac{1}{2}$,

$$\begin{cases} \varepsilon \sum_{s=0}^{t} \delta_s (1-\varepsilon)^s \le 1-\varepsilon, & \text{if } \delta_0 = 0, \\ \varepsilon \sum_{s=0}^{t} \delta_s (1-\varepsilon)^s \ge \varepsilon, & \text{if } \delta_0 = 1. \end{cases}$$

Hence $\overline{D}_\varepsilon \cap]1-\varepsilon, \varepsilon[= \varnothing$. Carrying on the process on both subintervals finally yields that $\lambda({}^c\overline{D}_\varepsilon) = \sum_t (2(1-\varepsilon))^t (2\varepsilon - 1) = 1$. $\square$

PROPOSITION 7. *Suppose that* $\text{supp}(\mu)$ *satisfies assumption* (3). *Let* a, b, $c \in \text{supp}(\mu)$ *such that* $\|a-b\| < \min(\|a-c\|, \|b-c\|)$.

(i) *If* $\varepsilon \le \frac{1}{2}$, *then* $[a,b] \times \{c\}^{n-1} \subset \text{supp}(\nu)$.

(ii) *If* $\varepsilon > \frac{1}{2}$, *then* $(b + (a-b)\overline{D}_\varepsilon) \times \{c\}^{n-1} \subset \text{supp}(\nu)$.

PROOF. Taking Lemma 5 into account, the proposition amounts to proving that $b + (a-b)D_\varepsilon \times \{c\}^{n-1} \subset \text{supp}(\nu)$.

Let $m := \lambda a + (1-\lambda)b$ with $\lambda \in D_\varepsilon$ and let $\eta > 0$ such that $\|a-b\| + 2\eta < \min(\|a-c\|, \|b-c\|)$. There exists some large enough $T \in \mathbb{N}$ such that $(1-\varepsilon)^T < \eta/\|a\|$ and $\lambda = \varepsilon \sum_{s=0}^{T} \delta_s (1-\varepsilon)^s$.

Now, let $O := B(a,\eta) \times B(c,\eta) \times \cdots \times B(c,\eta)$ be an open neighborhood of $(a, c, \dots, c)$ and let $y \in O$ be a starting value for the algorithm.

On the event $A := \bigcap_{1 \le s \le T} \{\omega^s \in B(a,\eta) \text{ if } \delta_{T-s} = 0,\ \omega^s \in B(b,\eta) \text{ if } \delta_{T-s} = 1\}$, the winning index at times $t \in \{1, \dots, T\}$ is $\mathbb{P}_y$-a.s. always 1. So

$$X_1^t = (1-\varepsilon)^t y_1 + \varepsilon \sum_{s=1}^{t} (1-\varepsilon)^{t-s} \omega^s, \qquad X_i^t = y_i,\ 2 \le i \le n,\ 1 \le t \le T.$$

This follows from an induction relying on the inequalities

$$\|X_1^t - a\| \le \eta + \|a-b\| \text{ and } \|X_1^t - b\| \le \eta + \|a-b\|,$$
$$\|X_i^t - a\| > \|a-c\| - \|y_i - c\| > \|a-b\| + \eta$$

and

$$\|X_i^t - b\| > \|a-b\| + \eta \quad \text{if } i \ge 2.$$

Now, still using the above inequalities, one has $\mathbb{P}_y$-a.s. on the event A,

$$\begin{aligned} \|X_1^T - m\| &\le (1-\varepsilon)^T \|y_1\| + \varepsilon \sum_{s/\delta_{T-s}=1} (1-\varepsilon)^{T-s} \|\omega^s - a\| \\ &\quad + \varepsilon \sum_{s/\delta_{T-s}=0} (1-\varepsilon)^{T-s} \|\omega^s - b\| \\ &\le (1-\varepsilon)^T (\|a\| + \eta) + (1 - (1-\varepsilon)^T)\eta \le 2\eta. \end{aligned}$$

Hence, writing $\theta := \sum_{s=0}^{T} \delta_s$,

$$\begin{aligned}\mathbb{P}_y(X^T \in B(m, 2\eta) \times B(c, \eta) \times \cdots \times B(c, \eta))\\ \geq \mathbb{P}_y(A)\\ \geq \mu(B(a, \eta))^{\theta} \times \mu(B(b, 2\eta))^{T-\theta} > 0.\end{aligned}$$

Now, let $x \in \operatorname{supp}(\mu)^n$ and $\alpha > 0$. Following Theorem 1, there exists some integer S such that

$$\forall\, z \in B_n(x, \alpha), \qquad \mathbb{P}_z(X^S \in O) > 0.$$

Therefore it follows, using the Markov property, that for every $z \in B_n(x, \alpha)$,

$$\begin{aligned}\mathbb{P}_z(X^{S+T} \in B(m, 2\eta) \times B(c, \eta) \times \cdots \times B(c, \eta))\\ \geq \mathbb{P}_z(X^S \in O)\mu(B(a, 2\eta))^{\theta} \times \mu(B(b, \eta))^{T-\theta} > 0.\end{aligned}$$

Now $\nu(B_n(x, \alpha)) > 0$ since $\operatorname{supp}(\mu)^n \subset \operatorname{supp}(\nu)$ and integrating w.r.t. the invariant probability measure ν finally yields

$$\begin{aligned}&\nu(B(m, 2\eta) \times B(c, \eta) \times \cdots \times B(c, \eta))\\ &\quad = \int \mathbb{P}_z\big(X^{S+T} \in B(m, 2\eta) \times B(c, \eta) \times \cdots \times B(c, \eta)\big)\nu(dz)\\ &\quad \geq \int_{B_n(x, \alpha)} \mathbb{P}_z\big(X^{S+T} \in B(m, 2\eta) \times B(c, \eta)^{n-1}\big)\nu(dz) > 0.\end{aligned}$$

This completes the proof. $\square$

PROOF OF THEOREM 4(b). All the probabilistic ingredients being included in Proposition 7, the proof now amounts to discovering three points a, b, c satisfying

$$a,\ b,\ c \in \operatorname{supp}(\mu),\ (a, b) \subset {}^c\operatorname{supp}(\mu) \quad \text{and} \\ \|a - b\| < \min(\|a - c\|, \|b - c\|) \tag{26}$$

in order to apply the proposition.

As a first step we will establish that, if $\operatorname{supp}(\mu)$ is a nonconvex compact set, then it satisfies the following (intuitive) property:

$$\forall\, \eta > 0,\ \exists\, x \neq y \in \operatorname{supp}(\mu) \\ \text{such that } (x, y) \subset {}^c\operatorname{supp}(\mu) \text{ and } \|x - y\| \leq \eta. \tag{27}$$

Then we will derive assumption (26) from (27).

Step 1. Let $a, b \in \operatorname{supp}(\mu)$ such that $(a, b) \not\subset \operatorname{supp}(\mu)$ and γ be a continuous $\operatorname{supp}(\mu)$-valued path joining a and b. Let $\eta > 0$. Now, consider a partition $g_0, g_1, \ldots, g_k, \ldots, g_{m+1}$ of γ such that $g_0 = a$, $g_{m+1} = b$, $g_i \in \operatorname{supp}(\mu)$ and $\|g_i - g_{i+1}\| \leq \eta$, $0 \leq i \leq m$. Two subcases arise.

Case (a). For some $i_0 \in \{0, \ldots, m\}$, $(g_{i_0}, g_{i_0+1}) \not\subset \operatorname{supp}(\mu)$. Let $e \in (g_{i_1}, g_{i_1+1}) \setminus \operatorname{supp}(\mu)$. Then the widest interval included in the open set $(g_{i_1}, g_{i_1+1}) \setminus \operatorname{supp}(\mu)$ and containing e, say $I := (x, y)$, is not empty; that is,

$x \neq y$ and, $\operatorname{supp}(\mu)$ being closed, x, y lie in $\operatorname{supp}(\mu)$. So assumption (27) is clearly satisfied.

Case (b). For all $i \in \{0, \dots, m\}$, $[g_i, g_{i+1}] \subset \operatorname{supp}(\mu)$. Then the piecewise affine path γ_1 joining the g_i's is $\operatorname{supp}(\mu)$-valued and has finite length. So, $\operatorname{supp}(\mu)$ being a compact set, there exists a $\operatorname{supp}(\mu)$-valued geodesic line $\tilde{\gamma}$ between a and b. The line $\tilde{\gamma}$ cannot be a straight line since $(a, b) \not\subset \operatorname{supp}(\mu)$. So, there exists a partition $\tilde{g}_0, \tilde{g}_1, \dots, \tilde{g}_m, \tilde{g}_{m+1}$ of $\tilde{\gamma}$, such that $\tilde{g}_0 = a$, $\tilde{g}_{m+1} = b$, $\|\tilde{g}_i - \tilde{g}_{i+1}\| \le \eta$, $0 \le i \le m$ and $(\tilde{g}_{i_1}, \tilde{g}_{i_1+1}) \not\subset \tilde{\gamma}$ for at least one index $i_1 \in \{0, \dots, m\}$. Now, $\tilde{\gamma}$ being a geodesic line in $\operatorname{supp}(\mu)$, $(\tilde{g}_{i_1}, \tilde{g}_{i_1+1})$ cannot be contained in $\operatorname{supp}(\mu)$.

Step 2. Assume now that no triplet $(a, b, c) \in \operatorname{supp}(\mu)^3$ satisfies assumption (26). Then,

$$\forall\, a,\ b \in \operatorname{supp}(\mu),\ (a, b) \subset {}^c\operatorname{supp}(\mu) \Rightarrow \forall\, c \in \operatorname{supp}(\mu),$$
$$\|a - b\| \ge \min(\|a - c\|, \|b - c\|).$$

Using Step 1, one may choose $a,\ b \in \operatorname{supp}(\mu)$ so as $\|a-b\| < \delta_C/8$ and $(a, b) \subset {}^c\operatorname{supp}(\mu)$ [where δ_C still denotes the diameter of $C = \operatorname{Conv}(\operatorname{supp}(\mu))$].

Now, the triangular inequality straightforwardly yields

$$\forall\, c \in \operatorname{supp}(\mu), \qquad \|a - c\| \le \|a - b\| + \min(\|a - c\|, \|b - c\|) \le 2\|a - b\|.$$

On the other hand, $\delta_C = \delta_{\operatorname{supp}(\mu)} := \max_{u,\, v \in \operatorname{supp}(\mu)} \|u - v\|$. Hence

$$\delta_C := \max_{c,\, c' \in \operatorname{supp}(\mu)} \|c - c'\| \le \max_{c \in \operatorname{supp}(\mu)} \|a - c\| + \max_{c' \in \operatorname{supp}(\mu)} \|a - c'\| \le 4\|a - b\| \le \frac{\delta_{\operatorname{supp}(\mu)}}{2}.$$

The obvious contradiction shows that (26) is fulfilled for some $a,\ b,\ c \in \operatorname{supp}(\mu)$, which completes the proof. □

3.3. *Absolute continuity properties* ($d \ge 1$). Let λ_C denote the restriction of the Lebesgue measure to the convex hull C of $\operatorname{supp}(\mu)$. Theorem 5 below shows how the absolute continuity properties can be transferred from the stimulus distribution μ to the invariant probability measure ν. It extends a first (one-dimensional) result established in [3].

THEOREM 5. (a) *Assume that* $\operatorname{supp}(\mu)$ *is path-connected and* $\overset{\circ}{\operatorname{supp}}(\mu) \neq \varnothing$ (*i.e.*, *assumption* (7) *which implies irreducibility on open sets*). *Then*

$$\mu \ll \lambda_C \quad \Rightarrow \quad \mathbf{1}_{\operatorname{supp}(\mu)^n} \cdot \nu \ll \lambda_{C^n}. \tag{28}$$

(b) *If* $\operatorname{supp}(\mu)$ *is a convex set with a nonempty interior* (*i.e.*, $\operatorname{supp}(\mu) = C$ *and* $\overset{\circ}{C} \neq \varnothing$), *then*

$$\mu \sim \lambda_C \quad \Rightarrow \quad \nu \sim \lambda_{C^n}.$$

PROOF OF THEOREM 5. Throughout the proof we will denote by φ the density of the probability measure μ.

(a) For every $J \subset \{1, \dots, n\}$, the projection from $(\mathbb{R}^d)^n$ onto $(\mathbb{R}^d)^J$ will be denoted π_J. Then we set $x_J := \pi_J(x)$, $\nu^J := \nu \circ \pi_J$ and $\lambda^J := \lambda_{C^n} \circ \pi_J = \lambda_{C^J}$. We show by induction on $k := |J|$ the following property:

$$\mathscr{P}_k \equiv \forall\ J \subset \{1, \dots, n\} \text{ with } |J| = k,\ \mathbf{1}_{\operatorname{supp}(\mu)^J} \cdot \nu^J \ll \lambda^J. \tag{29}$$

($\mathscr{P}_1$). Let $J = \{i\}$ and g: $\mathbb{R}^d \to \mathbb{R}_+$ be a nonnegative Borel function such that $g = 0$ λ-a.s. The p^ε-invariance of ν implies that

$$\begin{aligned} \int g d\nu^{\{i\}} = \int_{C^n} \nu(dx)\bigg(\int_{{}^cC_i(x)} g(x_{\{i\}})\mu(d\omega) \\ + \int_{C_i(x)} g((1-\varepsilon)x_{\{i\}} + \varepsilon\omega)\mu(d\omega)\bigg). \end{aligned} \tag{30}$$

Now, for every $x \in C^n$,

$$\int_{C_i(x)} g((1-\varepsilon)x_{\{i\}} + \varepsilon\omega)\mu(d\omega) \le \frac{1}{\varepsilon^d}\int_C g(u)\varphi\bigg(\frac{u-(1-\varepsilon)x_{\{i\}}}{\varepsilon}\bigg)\,du = 0.$$

So (30) now reads

$$\begin{aligned} \int g d\nu^{\{i\}} = \int_{C^n} \nu(dx) g(x_{\{i\}})(1-\mu(C_i(x))); \\ \text{that is, } \int_{C^n} \nu(dx) g(x_{\{i\}})\mu(C_i(x)) = 0. \end{aligned} \tag{31}$$

As $\mu(C_i(x)) > 0$ on $\{x_{\{i\}} \in \operatorname{supp}(\mu)\}$, (31) implies $\int_{\operatorname{supp}(\mu)} g(x_{\{i\}})\nu^{\{i\}}(dx_{\{i\}}) = 0$. Finally $\mathbf{1}_{\operatorname{supp}(\mu)} \cdot \nu^{\{i\}} \ll \lambda_C$.

($\mathscr{P}_k \Rightarrow \mathscr{P}_{k+1}$). Assume that $|J| = k+1$ and let g: $(\mathbb{R}^d)^J \to \mathbb{R}_+$ be a nonnegative Borel bounded function satisfying $g = 0$ λ^J-a.s. and g is everywhere null outside $\operatorname{supp}(\mu)^{J\setminus\{i\}}$.

Let $g_{i,\,x_{J\setminus\{i\}}}$ denote the partial function of g only depending on the ith component x_J, the other x_j's, $j \in J \setminus \{i\}$ being fixed. Now, the p^ε-invariance of ν yields

$$\begin{aligned} \int g\, d\nu^J = \int \nu(dx)\bigg(g(x_J)\sum_{\ell\notin J}\mu(C_\ell(x)) \\ + \sum_{i\in J}\int_{C_i(x)} g_{i,\,x_{J\setminus\{i\}}}((1-\varepsilon)x_J + \varepsilon\omega)\mu(d\omega)\bigg). \end{aligned} \tag{32}$$

Let $p \in \mathbb{N}^*$. For every $i \in J$ and $x \in C^n$, one has

$$\begin{aligned} &\int_{C_i(x)} g_{i,\,x_{J\setminus\{i\}}}((1-\varepsilon)x_J + \varepsilon\omega)\mu(d\omega) \\ &\quad \le p\int_C g_{i,\,x_{J\setminus\{i\}}}((1-\varepsilon)x_J + \varepsilon\omega)\,d\omega + \|g\|_\infty \lambda_C(\{\varphi \ge p\}) \\ &\quad \le \frac{p}{\varepsilon^d}\int_C g_{i,x_{J\setminus\{i\}}}(u)\,du + \|g\|_\infty \lambda_C(\{\varphi \ge p\}). \end{aligned}$$

Integrating w.r.t. ν leads to

$$\begin{aligned}(33)\quad &\int \nu(dx)\left(\int_{C_i(x)} g_{i,\,x_{J\setminus\{i\}}}((1-\varepsilon)x_J+\varepsilon\omega)\mu(d\omega)\right)\\ &\qquad\le \frac{p}{\varepsilon^d}\int \nu^{J\setminus\{i\}}(dx_{J\setminus\{i\}})\left(\int du_i\, g_{i,x_{J\setminus\{i\}}}(u_i)\right)+\|g\|_\infty \lambda_C(\{\varphi\ge p\}).\end{aligned}$$

Now, $g=0$ λ^J-a.s.; thus $g_{i,x_{J\setminus\{i\}}}(u_i)=0$ $\lambda(du_i)$-a.s. for $\lambda^{J\setminus\{i\}}$-almost every $x_{J\setminus\{i\}}$. Using now the induction assumption $\mathscr{P}_k$, we know that

$$\mathbf{1}_{\{\mathrm{supp}(\mu)^{J\setminus\{i\}}\}}\cdot \nu^{J\setminus\{i\}} \ll \lambda^{J\setminus\{i\}}.$$

Hence, for $\nu^{J\setminus\{i\}}$-almost every $x_{J\setminus\{i\}}\in \mathrm{supp}(\mu)^{J\setminus\{i\}}$, the first term on the right-hand side of (33) is 0. Then letting $p\to+\infty$ yields in turn that

$$\nu(dx)\text{-a.s.}\quad \int_{C_i(x)} g_{i,\,x_{J\setminus\{i\}}}((1-\varepsilon)x_J+\varepsilon\omega)\mu(d\omega)=0,$$

since g is 0 everywhere outside $\mathrm{supp}(\mu)^{J\setminus\{i\}}$. So (32) now reads

$$\begin{aligned}&\int g\,d\nu^J=\int \nu(dx)g(x_J)\Big(1-\sum_{i\in J}\mu(C_i(x))\Big);\\ (34)\quad &\text{that is,}\ \int \nu(dx)g(x_J)\underbrace{\sum_{i\in J}\mu(C_i(x))}_{>0\ \text{if}\ x_J\in\mathrm{supp}(\mu)^J}=0.\end{aligned}$$

So, $\int \nu(dx)\mathbf{1}_{\{\mathrm{supp}(\mu)^J\}}(x_J)g(x_J)=0$; that is, $\mathbf{1}_{\mathrm{supp}(\mu)^J}\cdot\nu^J\ll\lambda^J$, which completes the proof of (a).

(b) This part of the proof is based on the "improved" version of Lemma 3 stated in Section 2.2. Let us go back to the notations of Theorem 3 (Doeblin recurrence) and consider $x_0\in D$ and its related "petite recurrent" (open) set $B_n(x_0,\,r_0)$ given by Lemma 3(ii). A careful reading of the proof of Theorem 3 shows that the following inequality holds for any Borel set $B\subset B_n(x_0,\,r_0)$ and any $x\in D$ [assumption (17) is not required here]:

$$\mathbb{P}_x(X^{T_0+n}\in B)\ge \mathbb{E}_x\left(\mathbf{1}_{\{X^{T_0}\in B_n(x_0,\,\eta_0)\}}\mu^{\otimes n}\left(\frac{B-(1-\varepsilon)X^{T_0}}{\varepsilon}\right)\right).$$

Hence, integrating w.r.t. the invariant probability measure ν yields

$$(35)\qquad \nu(B)\ge\int_D \nu(dx)\mathbb{E}_x\left(\mathbf{1}_{\{X^{T_0}\in B_n(x_0,\,\eta_0)\}}\mu^{\otimes n}\left(\frac{B-(1-\varepsilon)X^{T_0}}{\varepsilon}\right)\right).$$

Assume now that $\mu^{\otimes n}(B)\neq 0$. Then $\mu\sim\lambda_C$ implies $(\lambda_C)^{\otimes n}(B)\neq 0$ which in turn implies that $(\lambda_C)^{\otimes n}((B-(1-\varepsilon)X^{T_0})/\varepsilon)\neq 0$ $\mathbb{P}_x$-a.s. since $(B-(1-\varepsilon)X^{T_0})/\varepsilon\subset D$. Finally $\mu^{\otimes n}((B-(1-\varepsilon)X^{T_0})/\varepsilon)>0$ $\mathbb{P}_x$-a.s.

A straightforward substitution in (35) yields that $\nu(B)\neq 0$; that is, $(\mu^{\otimes n})_{|B_n(x_0,\,r_0)}\ll \nu_{|B_n(x_0,r_0)}$. Now D is a countable union of open n-balls $B_n(x_0,r_0)$ since one may set $r_0:=(\varepsilon/8)\min_{i\neq j}\|(x_0)_i-(x_0)_j\|$. Finally, $\nu({}^cD)=0$ implies that $\mu^{\otimes n}\ll\nu$. Part (a) completes the proof. □

REMARK. If $\mu \sim \lambda_C$, there exists at most one invariant probability measure (see [9] or [20]). Unlike the one-dimensional setting (see [3]), this does not provide any stability property since the chain $(X^t)_{t\geq 0}$ is never Feller in C^n (because of ${}^cD$) whenever $d \geq 2$.

4. Asymptotic behavior of ν^ε when $\varepsilon \to 0$ ($d \geq 1$). The aim of this section is to prove that in some way the weight of ν^ε concentrates at the equilibrium point(s) of the algorithm when $\varepsilon \downarrow 0$. By equilibrium, we mean the zeros of the average function of the algorithm $h(x) := \int_C H(x, \omega)\mu(d\omega)$ defined on C^n. This definition fits with the convention made in Definition 1 for the Voronoi tessellation of points $x \in \partial D := C^n \setminus D$; that is,

$$\forall\ x \in C^n, \qquad h_i(x) := \int_{C_i(x)} (x_i - \omega)\mu(d\omega),\ 1 \leq i \leq n.$$

For notational convenience, if $I := I_1 \cup \cdots \cup I_p$ denotes the partition of the unit set I into the p clusters of $x \in^c D$, we set for every $\ell \in \{1, \dots, p\}$

$$C_{I_\ell}(x) := C_{\min I_\ell}(x) \quad \text{and} \quad h_{I_\ell}(x) := h_{\min I_\ell}(x).$$

Note that if $j \in I_\ell$ and $j \neq \min I_\ell$ then $h_j(x) = 0$. Hence, in some sense, the equilibrium set $\{h = 0\}$ for n points contains all the equilibrium points of the algorithm with p points, $p \leq n$.

The main ingredient of this section is an important property of the Kohonen CLVQ algorithm already mentioned in the introduction but not used since: if μ is *strongly diffuse* then h derives—on D—from a potential; that is,

$$\forall\ x \in D, \qquad h(x) := \nabla E_n^\mu(x) \quad \text{where } E_n^\mu(x) := \int \min_{1\leq i\leq n} \|x_i - \omega\|^2 \mu(d\omega).$$

In fact, the potential E_n^μ is continuous (even locally Lipschitz) on the whole space $(\mathbb{R}^d)^n$ and differentiable on $\bigcup_{i\neq j}\{x_i \neq x_j\}$. It was studied in full detail in [21] (E_n^μ is continuous, the sequence $n \mapsto \min_{C^n} E_n^\mu$ is decreasing to 0, etc.), together with some applications to high-dimensional numerical integration.

THEOREM 6. *Assume that* (i) *μ is strongly diffuse and* (ii) *μ satisfies assumption* (7).

(a) *If $d \geq 2$, any limiting point ν^0 for the weak convergence of the family $(\nu^\varepsilon)_{\varepsilon>0}$ as $\varepsilon \downarrow 0$ satisfies*

$$\operatorname{supp}(\nu^0) \subset \{h = 0\}$$

(*including the zeros of h in $\partial D := C^n \setminus D$*).

(b) *If $d = 1$, one may assume that* $\operatorname{supp}(\mu) = [0, 1]$ *and* (*see* [3]) *that X^t belongs to $F_n^+ := \{x \in [0, 1]^n / 0 < x_1 < \cdots < x_n < 1\}$ since all the* (*open*) *simplexes of $[0, 1]^n$ are absorbing sets. Then, still denoting by ν^ε the restriction to F_n^+ of the invariant probability measure, one has for any weak limiting point ν^0 of $\{\nu^\varepsilon,\ \varepsilon > 0\}$ as $\varepsilon \downarrow 0$,*

$$\operatorname{supp}(\nu^0) \subset \{h = 0\} \subset F_n^+.$$

REMARKS. Statement (b) slightly improves the original one-dimensional result in [3] where absolute continuity of μ was required.

If h were continuous on the whole state set C^n instead of D, the proof would reduce to a straightforward extension of the one-dimensional case in [3]: when $d = 1$, the key lemma is that the restriction of h to any absorbing set, say F_n^+, can be continuously extended on its closure. Then Lemma 8 completes the proof. This is no longer true when $d \geq 2$. In the multidimensional setting, the main ingredient is to approximate properly E_n^μ and its derivative by some continuously differentiable functions Φ_p satisfying:

(i) $\Phi_p \to E_n^\mu$ as $p \to +\infty$;

(ii) $\nabla\Phi_p \to \nabla\Phi_\infty$ as $p \to +\infty$ and $\nabla\Phi_\infty = \nabla E_n^\mu$ on D;

(iii) The inner product $\nabla\Phi_p \cdot h$ is continuous on C^n.

Note that the set $\{h = 0\}$ in item (a) includes the zeros of the h lying in ${}^cD$ provided by the above extension of h. Namely, it contains the points $x \in {}^cD$, whose clusters $(x_{I_1}, \dots, x_{I_p})$ are the components of an equilibrium point for the algorithm with p units, $p < n$. So this result is less satisfactory than the one-dimensional one.

4.1. *Approximation of the average function h.*

DEFINITION 2. Let $p \in \mathbb{N}^*$. For every $x \in (\mathbb{R}^d)^n$, one sets

$$\forall\, \omega \in \mathbb{R}^d,\ \varphi_p(\omega, x) := \begin{cases} \left(\sum_{k=1}^{n} \|x_k - \omega\|^{-2p} \right)^{-1/p}, & \text{if } \omega \notin \{x_i,\ 1 \leq i \leq n\}, \\ 0, & \text{otherwise,} \end{cases}$$

$$\Phi_p(x) := \int \varphi_p(\omega, x)\mu(d\omega) \in \mathbb{R}_+.$$

PROPOSITION 8. *For every $p \in \mathbb{N}^*$, we have the following*:

(i) *Φ_p is a continuously differentiable function whose gradient at $x \in (\mathbb{R}^d)^n$ is*

$$\nabla\Phi_p(x) = 2\left(\int \left(\sum_{k=1}^{n} \|x_k - \omega\|^{-2p} \right)^{-(p+1)/p} (x_i - \omega)\mu(d\omega) \right)_{1 \leq i \leq n}.$$

(ii) *For every $x \in D$, $\nabla\Phi_p(x) \longrightarrow_{p \to +\infty} h(x)$.*
If $x \notin D$ and $I = I_1 \cup \cdots \cup I_\ell \cup \cdots \cup I_p$ denotes the partition of I into the clusters of x, then

$$\forall\, i \in I_\ell, \qquad \nabla\Phi_{p,i}(x) \longrightarrow_{p \to +\infty} \frac{1}{|I_\ell|} h_{I_\ell}(x) = \frac{1}{|I_\ell|} \int_{C_{I_\ell}(x)} (x_{I_\ell} - \omega)\mu(d\omega).$$

(iii) *Let $\nabla\Phi_\infty := \lim_p \nabla\Phi_p$ on C^n. Then $\|h\|^2/n \leq h \cdot \nabla\Phi_\infty$.*

REMARKS. Furthermore, $\Phi_p \to_{p\to+\infty} E_n^{\mu}$ uniformly on compact sets of $(\mathbb{R}^d)^n$.

In spite of our notation, $\nabla\Phi_\infty$ is not a gradient on ${}^cD$.

The above results on Φ_p follow from the properties of the functions $\varphi_p(\omega, x)$ gathered in the straightforward lemma below whose proof is left to the reader.

LEMMA 6. *For every $p \in \mathbb{N}^*$, $x \in (\mathbb{R}^d)^n$, $\omega \in (\mathbb{R}^d)^n$, we have the following*:

(i) $n^{-1/p}\min_{1\le k\le n}\|x_k-\omega\|^2 \le \varphi_p(\omega, x) \le \min_{1\le k\le n}\|x_k-\omega\|^2$;

(ii) $\varphi_p(\omega,\cdot)$ *is continuously differentiable on* $(\mathbb{R}^d)^n$ *and, for every* $i \in \{1,\dots,n\}$,

$$\frac{\partial\varphi_p}{\partial x_i}(\omega, x) := \begin{cases} 2\left(\sum_{k=1}^{n}\left(\dfrac{\|x_i-\omega\|}{\|x_k-\omega\|}\right)^{2p}\right)^{-(p+1)/p}(x_i-\omega), & \textit{if } \omega \notin \{x_j,\ 1\le j\le n\},\\ 0, & \textit{otherwise}.\end{cases}$$

Hence

$$\left\|\frac{\partial\varphi_p}{\partial x_i}(\omega, x)\right\| \le 2\min_{1\le k\le n}\left(\frac{\|x_k-\omega\|}{\|x_i-\omega\|}\right)^{2(p+1)}\|x_i-\omega\| \le 2\|x_i-\omega\|.$$

PROOF OF PROPOSITION 8. (i) Using Lemma 6, for every $\omega \in C$, $x \mapsto \varphi_p(\omega, x)$ is continuously differentiable on $(\mathbb{R}^d)^n$ and $\varphi_p(\omega,\cdot)$ and $(\partial\varphi_p/\partial x_i)(\omega, x)$ are, respectively, bounded on C by δ_C^2 and $2\delta_C$. So Φ_p is well defined and continuously differentiable thanks to the suitable dominated convergence theorems.

(ii) Let $i \in I_\ell$; if $\omega \notin C_{I_\ell}(x)$,

$$\left\|\frac{\partial\varphi_p}{\partial x_i}(\omega, x)\right\| \le 2\underbrace{\min_{1\le k\le n}\left(\frac{\|x_k-\omega\|}{\|x_i-\omega\|}\right)^{2(p+1)}}_{<1}\|x_i-\omega\|.$$

Hence $(\partial\varphi_p/\partial x_i)(\omega, x) \longrightarrow_{p\to+\infty} 0$.

If $\omega \in C_{I_\ell}(x)$,

$$\sum_{j=1}^{n}\left(\frac{\|x_i-\omega\|}{\|x_j-\omega\|}\right)^{2p} = |I_\ell| + \sum_{j\notin I_\ell}^{n}\left(\frac{\|x_i-\omega\|}{\|x_j-\omega\|}\right)^{2p} \longrightarrow_{p\to+\infty} |I_\ell|.$$

Hence $(\partial\varphi_p/\partial x_i)(\omega, x) \longrightarrow_{p\to+\infty} (1/|I_\ell|)(x_i-\omega)$ and $\|\partial\varphi_p/\partial x_i\| \le 2\|x_i-\omega\|$. The dominated convergence theorem completes the proof (if $x \in D$, set $I_i := \{i\}$).

(iii) This follows from the obvious inequalities $|I_\ell| \le n$. $\square$

4.2. *Proof of Theorem* 6. Recall that, in any case, $\operatorname{supp}(\nu^\varepsilon) \subset C^n$ [cf. Theorem 4(a)].

LEMMA 7. (a) *Let I be a cluster of $x \in (\mathbb{R}^d)^n$. Then*

$$\lim_{y\to x} \sum_{i\in I} h_i(y) = \int_{C_I(x)} (x_I - \omega)\mu(d\omega) = h_I(x).$$

(b) *For every $p \in \mathbb{N}^*$, $\nabla\Phi_p \cdot h$ is continuous on the whole set C^n.*

PROOF. (a) Let $y^{(p)} \to x$. Obviously $C_I(x) \subset \liminf_p \bigcup_{i\in I} C_i(y^{(p)}) \subset \overline{\liminf_p \bigcup_{i\in I} C_i(y^{(p)})} \subset \overline{C_I(x)}$. Since μ is strongly diffuse, it follows that $\mathbf{1}_{\{\cup_{i\in I} C_i(y^{(p)})\}} \longrightarrow \mathbf{1}_{C_I(x)}$ μ-a.s. The dominated convergence completes the proof.

(b) The result follows from item (a), once it is observed that all the components of $\nabla\Phi_{p,i}$ are identical for any index i belonging to the same cluster I_ℓ. $\square$

LEMMA 8. *If $f \in \mathscr{C}^1((\mathbb{R}^d)^n, \mathbb{R})$ (continuously differentiable), then*

$$\int_{C^n} \nabla f(x)\cdot h(x)\nu^\varepsilon(dx) \longrightarrow 0 \quad \textit{as } \varepsilon \to 0.$$

PROOF. Let $x \in C^n$ and $\omega \in C$. The definition of the semigroup P^ε in (6) and the Taylor formula yield

$$|P^\varepsilon(f)(x) - f(x) + \varepsilon\nabla f(x)\cdot H(\omega, x)| \le w(\nabla f, \varepsilon\delta_C)\varepsilon\delta_C, \tag{36}$$

where $w(\nabla f, \cdot)$ denotes the continuity modulus of ∇f. Integrating (36) with respect to the invariant probability measure ν^ε then yields

$$\left|\int_{C^n} \nabla f(x)\cdot h(x)\nu^\varepsilon(dx)\right| \le \delta_C w(\nabla f, \varepsilon\delta_C) \longrightarrow_{\varepsilon\to 0} 0. \qquad \square$$

PROOF OF THEOREM 6. (a) Let $\nu^0 := \lim_k \nu^{\varepsilon_k}$ be a limiting point of the family ν^ε, $\varepsilon > 0$, for the weak topology on the probability measures on C^n. Applying Lemma 8 to the Φ_p functions (see Section 4.1) yields

$$\int_{C^n} \Phi_p \cdot h(x)\nu^{\varepsilon_k}(dx) \longrightarrow_{k\to+\infty} 0.$$

On the other hand, Lemma 7 implies that $\Phi_p \cdot h$ is continuous on C^n, so

$$\forall\ p \in \mathbb{N}^*, \qquad \int_{C^n} \Phi_p \cdot h(x)\nu^0(dx) = \lim_k \int_{C^n} \Phi_p \cdot h(x)\nu^{\varepsilon_k}(dx) = 0.$$

Then, letting $p \uparrow +\infty$, one derives from Proposition 8(ii) and 8(iii) that

$$0 \le \frac{1}{n}\int_{C^n} \|h(x)\|^2\nu^0(dx) \le \int_{C^n} \nabla\Phi_\infty \cdot h(x)\nu^0(dx) = 0.$$

Hence $\operatorname{supp}(\nu^0) \subset \{h = 0\}$.

(b) If $d = 1$, assumption (7) reads: $\mathrm{supp}(\mu)$ is a nondegenerate interval. So, w.l.o.g., one may set $\mathrm{supp}(\mu) := [0, 1]$. Furthermore, F_n^+ being an absorbing set, one may only consider F_n^+ instead of $[0, 1]^n$. The related invariance probability measure(s) on F_n^+ will still be denoted ν^ε. Lemma 8 remains obviously true in that setting. Following [2], one checks that the restriction of the average function h to F_n^+ admits a continuous extension $\tilde{h}$ to $\overline{F_n^+} \cap C^n$. Using $\tilde{h}$ instead of Φ_p in the above proof straightforwardly implies that $\mathrm{supp}(\nu^0) \subset \{\tilde{h} = 0\}$. But, still following [2], $\{\tilde{h} = 0\} = \{h = 0\} \subset \overline{F_n}^+ \cap D = F_n^+$. $\square$

PROVISIONAL REMARKS. Actually, we guess that ν^0 never weights the edge of D. Hence only the "true" n equilibria are weighted. Then, using, for example, recent results in Fort and Pagès [12], it is most likely that only D-valued attracting equilibria of the ODE $\equiv \dot{x} = -h(x)$ are possibly weighted.

Conclusion. This work is a first step toward deeper investigations of quantization algorithms. The existence of an invariant probability measure ν^ε for every $\varepsilon \in (0, 1]$ under natural assumptions on the stimuli distribution, the geometrical characteristics of its support and its absolute continuity were studied. Its asymptotic behavior as $\varepsilon \to 0$ has been elucidated. The comparison between these constant step results and those obtained with a decreasing step assumption ($\sum_t \varepsilon_t = +\infty$ and $\sum_t \varepsilon_t^2 < +\infty$; see [21] or [10]) suggests implementing some intermediary algorithms inspired by simulated annealing to reach an optimal (quadratic) quantization in the mutlidimensional case. In the one-dimensional case, some simpler and faster deterministic algorithms provide the optimal quantization (see papers by Lloyd, Kieffer and Trushkin in [13]).

REFERENCES

[1] BENVENISTE, A., MÉTIVIER, M. and PRIOURET, P. (1987). *Algorithmes Adaptatifs et Approximations Stochastiques*. Masson, Paris.

[2] BOUTON, C. and PAGÈS, G. (1993). Self-organization and convergence of the one-dimensional Kohonen algorithm with nonuniformly distributed stimuli. *Stochastic Process. Appl.* **47** 249–274.

[3] BOUTON, C. and PAGÈS, G. (1994). Convergence in distribution of the 1-dimensional Kohonen algorithms when the stimuli are not uniform. *Adv. in Appl. Probab.* **26** 80–103.

[4] BOUTON, C. and PAGÈS, G. (1994). Convergence in distribution of the multi-dimensional Kohonen algorithm with 0 neighbor. Technical Report 259, Univ. Paris 6.

[5] BURTON, R. M. and FARIS, W. G. (1996). A self-organizing cluster process. *Ann. Appl. Probab.* **6** 1232–1247.

[6] COHORT, P. (1997). *Unicité d'un quantifieur optimal par le théorème du col.* Technical report, Univ. Paris 6.

[7] COTTRELL, M. and FORT, J. C. (1987). Etude d'un algorithme d'auto-organisation. *Ann. Inst. H. Poincaré Sér. B* **23** 1–20.

[8] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[9] DUFLO, M. (1990). *Méthodes Récursives Aléatoires*. Masson, Paris.

[10] FLANAGAN, J. A. (1994). Self-organizing neural networks. Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne.

[11] FORT, J. C. and PAGÈS, G. (1995). On the a.s. convergence of the Kohonen algorithm with a general neighborhood function. *Ann. Appl. Probab.* **5** 1177–1216.
[12] FORT, J. C. and PAGÈS, G. (1996). Asymptotics of the invariant distribution of a stochastic algorithm with constant step. Technical Report 324, Univ. Paris 6.
[13] GRAY, R. M. ed. (1982). Special issue on quantization. *IEEE Trans. Inform. Theory* **28** 127–394.
[14] KIEFFER, J. C. (1982). Exponential rate of convergence for Lloyd's method I. *IEEE Trans. Inform. Theory* **28** 205–210.
[15] KOHONEN, T. (1982). Analysis of a simple self-organizing process. *Biol. Cybernet.* **44** 135–140.
[16] KOHONEN, T. (1984). *Self-Organization and Associative Memory*. Springer, New York.
[17] KOHONEN, T. (1991). Self-organizing maps: optimization approaches. In *Artificial Neural Networks* (T. Kohonen, et al., eds.) 981–990. North-Holland, Amsterdam.
[18] KUSHNER, H. J. and CLARK, D. S. (1978). *Stochastic Approximation for Constrained and Unconstrained Systems*. Springer, New York.
[19] LAMBERTON, D. and PAGÈS, G. (1996). On the critical points of the one-dimensional competitive learning vector quantization algorithm. *Proceedings ESANN 1996* 97–102.
[20] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, New York.
[21] PAGÈS, G. (1993). Voronoi tessellation, space quantization algorithms and numerical integration. *Proceedings ESANN 1993* 221–228.
[22] REVUZ, D. (1975). *Markov Chains*. North-Holland, Amsterdam.
[23] SARZEAUD, O., STÉPHAN, Y. and TOUZET, C. (1990). Application des cartes auto-organisatrices à la génération de maillages aux éléments finis. *Proceedings of Neuro-Nîmes* **1** 81–96.
[24] TRUSHKIN, A. V. (1982). Sufficient conditions for uniqueness of a locally optimal quatizer for a class of convex error weighting functions. *IEEE Trans. Inform. Theory* **28** 187–198.

SAMOS
UNIVERSITÉ PARIS 1
U.F.R. 27, 90, RUE DE TOLBIAC
F-75634 PARIS CEDEX 13
FRANCE

LABORATOIRE DE PROBABILITÉS
UNIVERSITÉ PARIS 6
URA 224
TOUR 56, F-75252 PARIS CEDEX 05
AND
UNIVERSITÉ PARIS 12
FACULTÉ DE SCIENCES
61 AVE. GÉNÉRAL DE GAULLE
F-94010 CRÉTEIL CEDEX
FRANCE
E-MAIL: gpa@ccr.jussieu.fr