Minimax bounds in nonparametric estimation of multidimensional deterministic dynamical systems

Alain BERLINET and Gérard BIAU*

Department of Probability and Statistics
University Montpellier II
Place Eugène Bataillon
34095 Montpellier Cedex 5
France

Abstract

We consider the problem of estimating a multidimensional discrete deterministic dynamical system from the first $n+1$ observations. We exhibit the optimal rate function r_n and show that the nearest neighbor estimator achieves this optimal rate, extending a recent study carried out by Guerre and Maës [18] in the univariate case. Former results strongly rely on the natural order structure of the set of real numbers. Here the optimal rate function is defined from multidimensional spacings which are edge lengths of simplices associated with a triangulation of the Voronoi cells built from the observations. A simulated example illustrates the theory.

Index Terms **— Deterministic dynamical system, nonparametric estimation, nearest neighbor estimator, minimax bound, Voronoi tessellation.**

AMS 2000 Classification**: 62G07, 37B20.**

1 Introduction

A lot of work has been devoted in recent years to understand and characterize properties of deterministic (*i.e.*, non-random) dynamical systems. The approximation of the behavior of a dynamical system is usually done by direct simulations of the individual trajectories. This method is particularly useful in the situation where a specific trajectory has to be approximated for a finite period of time. However, if one is interested in the long term behavior and if the underlying system exhibits complicated dynamics, then the information derived from single trajectories may prove inadequate. Rather in this case it seems more appropriate to determine probabilistic and statistical description of the dynamical behavior (for example in terms of invariant measure, Frobenius-Perron operator, empirical measure or observations correlations), see Boyarsky and Góra [8], Collet [13], Hall [20] and

*All correspondence to be addressed to this author. Email: biau@ensam.inra.fr.

Lasota and Mackey [23] for general results. In this way, deterministic dynamical systems have recently drawn the attention of a large part of the statistical community, see Guégan [19] as well as the survey papers by Berliner [4], Chatterjee and Yilmaz [12], Tong [26] and the discussions therein.

In the present paper, we consider the following discrete deterministic model

$$\begin{cases} X_{i+1} = \phi(X_i), & i \geq 1, \\ X_1 = x_1 \in K^d, \end{cases} \tag{1.1}$$

and we address the issue of nonparametric estimation of ϕ, a mapping from the unit hypercube $K^d = [0,1]^d$ of $\mathbb{R}^d$ into itself (for related problems, see Bosq and Guégan [6] and Maës [24]). In the case $d = 1$, Guerre and Maës [18] proposed a specific framework in order to take into account the deterministic nature of model (1.1). In the absence of noise, the nonparametric estimation of ϕ reduces to an approximation problem. This problem can be solved on the basis of the 1-nearest neighbor estimator $\tilde{\phi}_n$ defined from the first $n+1$ observations $X_1, \ldots, X_{n+1}$ drawn from ϕ as

$$\tilde{\phi}_n(x) = X_{i_n(x)+1}, \tag{1.2}$$

where $i_n(x)$ is such that $||x - X_{i_n(x)}|| \leq ||x - X_j||$, $j = 1, \ldots, n$. Thus, in the 1-dimensional context, the approximation of $\phi(x)$ is determined by the place of x among the points

$$0 = X_{(0)}, X_{(1)}, \frac{X_{(1)} + X_{(2)}}{2}, X_{(2)}, \ldots, X_{(i)}, \frac{X_{(i)} + X_{(i+1)}}{2}, X_{(i+1)}, \ldots, X_{(n)}, X_{(n+1)} = 1,$$

where $X_{(i)}$ is the i^{th} order statistics. Now, the rate of convergence in uniform norm of the estimator $\tilde{\phi}_n$ will depend on how the observations will become close to x (for any x), that is on the *spacings* $X_{(i+1)} - X_{(i)}$ where the interval $[X_{(i)}, X_{(i+1)}[$ contains x. This methodology, introduced by Guerre and Maës [18], strongly relies on the natural order structure of the real line and therefore cannot be directly extended to $\mathbb{R}^d$, with $d > 1$. Actually, most of theoretical results on deterministic dynamical systems (existence of an invariant measure, statistical stability, etc.) pertain to systems defined on real intervals (see Block and Coppel [5], Boyarsky and Góra [8] and Lasota and Mackey [23]) and few generalizations to higher dimension spaces are available. Our approach is of geometrical nature and does not rely on any order notion. It covers former univariate results.

In higher dimensions, the set of elements $y \in K^d$ such that $X_{i_n(y)} = X_i$ for some fixed index i in $\{1, \ldots, n\}$ is called the *Voronoi cell* with *nucleus* X_i. The set of all Voronoi cells obtained by varying the index i is the *Voronoi tessellation* generated by $X_1, \ldots, X_n$. In the present paper, the multidimensional rate of uniform convergence will be determined by means of a triangulation of the Voronoi cells into simplices. Thus, we need geometrical concepts in higher dimension. In a sake of completeness, they will be reviewed in Section 2.

Our final goal is to compare estimators via a *local risk minimax approach* (Devroye [16] and Ibragimov and Has'minskii [21]) using the quantity

$$\inf_{\hat{\psi}_n} \sup_{\psi \in \mathcal{V}_\epsilon(\phi)} ||r_n(.\,;\psi)(\hat{\psi}_n - \psi)||_\infty, \tag{1.3}$$

where $\mathcal{V}_\epsilon(\phi)$ is the following neighborhood of ϕ:

$$\mathcal{V}_\epsilon(\phi) = \Big\{\psi \in C^1(K^d) \ : \ ||\psi - \phi||_\infty \leq \epsilon,\, ||\mathrm{d}\psi||_\infty \leq 2\,||\mathrm{d}\phi||_\infty\Big\},$$

and where $\hat{\psi}_n$ is any estimator of ψ based on observations $X_1, \ldots, X_{n+1}$ drawn from ψ. In (1.3), $r_n(.\,;\psi)$ is the *rate* (or *weight*) *function* associated with ψ in $\mathcal{V}_\epsilon(\phi)$. In Section 3, we define the multidimensional rate function r_n and we show, under some assumptions (Theorem 3.1), that the minimax risk (1.3) remains asymptotically bounded away from 0 and infinity. The obtained positive lower bound means that r_n is an optimal rate for estimation. We also show (Proposition 3.2) that the simple 1-nearest neighbor estimator (Bosq and Lecoutre [7]) achieves the optimal rate r_n. For clarity sake, proofs are gathered in Section 5.

2 Voronoi tessellations and triangulations

2.1 Notations

Throughout, we assume the reader familiar with the standard linear and affine theory on $\mathbb{R}^d$ and with elementary topological notions. For a complete presentation, see Brøndsted [9], Møller [25] and Ziegler [27].

For d in $\mathbb{N}^*$, we equip $\mathbb{R}^d$ with the usual scalar product and we denote by $||.||$ the induced Euclidean norm. λ is the Lebesgue measure on $\mathbb{R}^d$. By the *dimension* of a subset M of $\mathbb{R}^d$, we mean the dimension of the affine hull aff M of M, and we denote it by dim M. The *relative interior* rint M of M is the interior of M in its affine hull. By a *convex polytope*, or simply a *polytope*, it is meant a set which is the convex hull conv $(x_1, \ldots, x_p)$ of a non-empty finite set $\{x_1, \ldots, x_p\}$ of points in $\mathbb{R}^d$. A polytope P is called a *k-polytope* if dim $P = k$. A convex subset F of P is called a *face* of P if for any two distinct points $y,\ z \in P$ such that $]y, z[\cap F$ is non-empty, we actually have $[y, z] \subset F$. The subsets $\emptyset$ and P of P are both faces, called the *improper* faces; all other faces are called *proper*. The faces of dimension 0, 1 and dim $P-1$, respectively, are called *vertices*, *edges* and *facets*.

A convex polytope S in $\mathbb{R}^d$ with the property that there exists an *affinely independent* family $\{x_1, \ldots, x_p\}$ such that $S = $ conv $(x_1, \ldots, x_p)$ is called a *simplex*. The set $\{x_1, \ldots, x_p\}$ is precisely the set of vertices of S. By a *k-simplex* we mean a k-polytope which is a simplex. A simplex is a k-simplex if, and only if, it has $k+1$ vertices. A 1-simplex is thus a closed segment. A 2-simplex is a *triangle* and a 3-simplex is a *tetrahedron*. Given two

d-simplices S_1 and S_2 in $\mathbb{R}^d$, there exists an affine isomorphism u_{12} from $\mathbb{R}^d$ onto $\mathbb{R}^d$ such that $u_{12}(S_1) = S_2$. In particular, any d-simplex in $\mathbb{R}^d$ may be transformed by an affine isomorphism of $\mathbb{R}^d$ onto the *standard d-simplex* Δ_d with $d+1$ vertices in $\mathbb{R}^d$:

$$\Delta_d = \text{conv } (0, e_1, \dots, e_d),$$

where e_i is the i-th unit vector in $\mathbb{R}^d$. A *pyramid* in $\mathbb{R}^d$ is a convex polytope of the form

$$P = \text{conv } \Big(\{x_0\} \cup B\Big),$$

where B is a polytope in $\mathbb{R}^d$ called the *basis* of P, and x_0 is a point of $\mathbb{R}^d -$ aff B called the *apex* of P. Note that the basis and apex need not be unique: a simplex is a pyramid where any facet may be taken as the basis, or, equivalently, any vertex may be taken as the apex. A pyramid P is called a *k-pyramid* if dim $P = k$. Clearly, a pyramid is a k-pyramid if, and only if, its basis is a $(k-1)$-polytope.

2.2 Nearest neighbor and Voronoi tessellation

Definition (1.2) of the index $i_n(x)$ leads to introduce the *Voronoi tessellation* associated with $X_1, \dots, X_n$. Basically, a *tessellation* (or, sometimes, a *mosaic*) of a k-dimensional subset M of $\mathbb{R}^d$ is a subdivision $M = \cup_i M_i$ into k-dimensional non-overlapping (*i.e.*, with disjoint relative interiors) sets M_i. Such arrangements occur in many natural situations and, depending on the context, the sets M_i might be called cells, crystals, regions, tiles, etc. Consider now a finite set $\mathcal{S} = \{x_1, \dots, x_n\}$ of points in the unit hypercube $K^d = [0,1]^d$. Each x_i is called a *nucleus* and it generates a cell $V_n(x_i)$ defined by

$$V_n(x_i) = \big\{y \in K^d \ : \ ||y - x_i|| \leq ||y - x_j|| \text{ for all } x_j \in \mathcal{S}\big\}.$$

In other words, the cell $V_n(x_i)$ consists of all points of $\mathcal{S}$ which have x_i as nearest nucleus, see Figure 1. It is called the *Voronoi cell* generated by the *nucleus* x_i. The $V_n(x_i)$'s are seen to be d-dimensional closed convex sets. These sets are either superimposed (in case of ties in the x_i's) or with disjoint interiors. Consequently, the distinct $V_n(x_i)$'s constitute a tessellation of K^d usually called *Voronoi tessellation* associated with $\mathcal{S}$. Voronoi tessellations can in fact be defined in a much more general context, where the set $\mathcal{S}$ is not necessarily finite and not necessarily embedded in the hypercube K^d (see Møller [25] for the required hypotheses and details). As far as we are concerned, one of the main results on (finite) Voronoi tessellations of K^d is as follows:

Theorem 2.1 *Let $\mathcal{S} = \{x_1, \dots, x_n\}$ be a finite set of points in the hypercube K^d. Then the Voronoi cells $V_n(x_i)$, $i = 1, \dots, n$, are d-*polytopes.

Since the Voronoi cells are polytopes, they can be triangulated into simplices. This argument will be crucial in the definition of the multidimensional rate function in Section 3. The key-theorem is as follows:

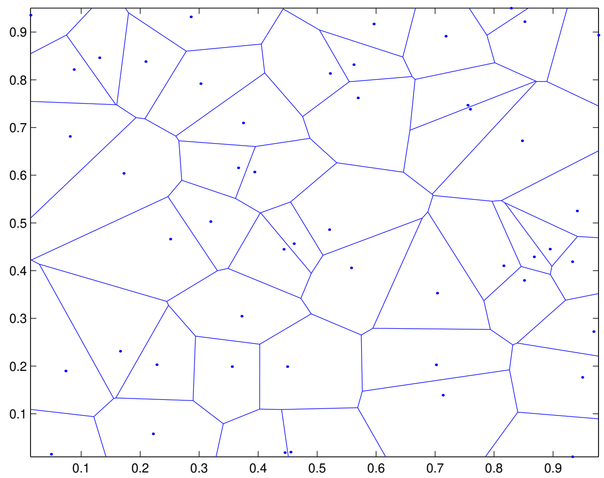

Figure 1: **2**-dimensional Voronoi tessellation generated by **50** independent and uniformly distributed data on the unit square $\mathbf{K}^2$.

Theorem 2.2 *Let P be a k-polytope in $\mathbb{R}^d$, let $x_0 \in P$ and let $F_1, \dots, F_p$ be the facets of P that do not contain x_0. Then*

(a)

$$P = \bigcup_{j=1}^{p} \operatorname{conv}\left(\{x_0\} \cup F_j\right).$$

The sets $P_j = \operatorname{conv}\left(\{x_0\} \cup F_j\right)$ are k-pyramids and the intersection of P_j and $P_{j'}$ $(j \neq j')$ is a common proper face of both P_j and $P_{j'}$.

(b) *Further, $\forall\, j = 1, \dots, p$,*

$$P_j = \bigcup_{l=1}^{p_j} T_l.$$

The T_l's are k-simplices with common apex x_0 and the intersection of T_l and $T_{l'}$ $(l \neq l')$ is a common proper face of both T_l and $T_{l'}$.

(c) *In particular,*

$$P = \bigcup_{j=1}^{p} \bigcup_{l=1}^{p_j} T_l.$$

Point (c) of the theorem means that a k-polytope P can be subdivided into non-overlapping k-simplices with common apex x_0. It is often said that the T_l's form a *triangulation* of P. As an illustration, Figure 2 depicts the triangulation of a 2-dimensional polytope (*i.e.*, a polygon). Clearly, for $k \geq 3$, such a triangulation need not be unique.

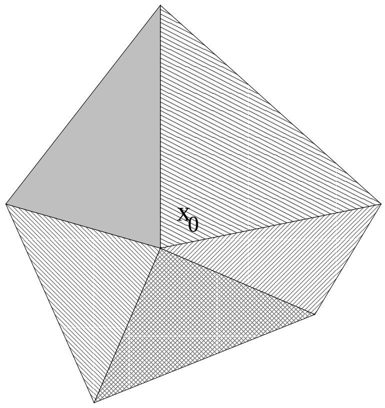

Figure 2: Triangulation of a $\mathbf{2}$-dimensional polytope as described in Theorem 2.2.

Now, let $X_1, \ldots, X_{n+1}$ be the first $n+1$ observations drawn from the dynamical system ϕ under study and let $V_n(X_1), \ldots, V_n(X_n)$ be the Voronoi cells associated with $X_1, \ldots, X_n$. Let x be a point of K^d. As the distinct Voronoi cells constitute a tessellation of K^d, we can find a cell $V_n(x)$ that contains x. In case of ties in the X_i's, the cell with the smallest index is affected to x. Since the distinct $V_n(X_i)$'s, $i = 1, \ldots, n$, are non-overlapping, $V_n(x)$ is λ-*a.e. uniquely* defined. In case of a distance tie, the candidate with the smallest index is affected to x. In this geometrical framework, the 1-nearest neighbor estimator defined in (1.2) takes the following form:

$$\tilde{\phi}_n(x) = X_{i+1} \text{ whenever } V_n(x) = V_n(X_i), \quad i = 1, \ldots, n.$$

From Theorem 2.1, it is deduced that each Voronoi polytope $V_n(X_i)$ can be triangulated as described in Theorem 2.2 into non-overlapping d-simplices, say T_{ij}^n, $j = 1, \ldots, m_i$, with common apex X_i. Consequently, for x in $V_n(X_i)$, there exists a d-simplex that contains x. We denote this d-simplex by $T_n(x)$. Here again, $T_n(x)$ is λ-*a.e. uniquely* determined. In case of tie, the simplex with the smallest index j is affected to x. As suggested by Figure 2 (case $d = 2$), the edge lengths drawn from X_i in T_{ij}^n correspond to extremal values of $||x - X_i||$ on the boundary of the Voronoi cell $V_n(X_i)$. As a matter of fact, these edge lengths will turn out to be closely related to our definition of multidimensional spacings (see Section 3 below).

Finally, for each d-simplex T_{ij}^n of the triangulation, we denote by u_{ij}^n an affine isomorphism of $\mathbb{R}^d$ such that $u_{ij}^n(T_{ij}^n) = \Delta_d$ and, for sake of convenience, such that $u_{ij}^n(X_i) = 0$.

3 Nonparametric estimation of a dynamical system

3.1 Multidimensional spacings and rate function

Let $\psi : M \subset \mathbb{R}^d \to \mathbb{R}^d$ be a bounded function defined on a subset M of $\mathbb{R}^d$. Throughout, we denote by $||\psi||_\infty$ the uniform norm of such a function ψ, *i.e.*,

$$||\psi||_\infty = \sup_{x \in M} ||\psi(x)||.$$

If the function ψ is differentiable, we denote by $\mathrm{d}\psi(x)$ the differential of ψ at the point x and by $||\mathrm{d}\psi(x)||$ its linear operator norm:

$$||\mathrm{d}\psi(x)|| = \sup_{\{h \in \mathbb{R}^d : ||h||=1\}} ||\mathrm{d}\psi(x).h||.$$

As is well known, $\mathrm{d}\psi(x)$ does not depend on the particular point x when ψ is affine and, in this case, we shall write $\mathrm{d}\psi$ instead of $\mathrm{d}\psi(x)$.

As mentioned in the introduction, the rate function r_n used in the 1-dimensional context by Guerre and Maës [18] involves the order structure of the real line and is not directly extensible to higher dimensions. It is formally defined by

$$r_n(x) = \begin{cases} 1/S_0 & \text{if } x \in [0, X_{(1)}[\\ 2/S_i & \text{if } x \in [X_{(i)}, X_{(i+1)}[, \quad i = 1, \ldots, n-1, \\ 1/S_n & \text{if } x \in [X_{(n)}, 1], \end{cases} \tag{3.1}$$

where the S_i's are the spacings associated with the first n observations $X_1, \ldots, X_n$ of the system ϕ, namely

$$S_i = X_{(i+1)} - X_{(i)}, \quad i = 0, \ldots, n.$$

We shall now introduce the *multidimensional spacings* and the *multidimensional rate function* that will be used in the uniform norm of the estimation error of the system (1.1). For this, in view of Theorem 2.2, we triangulate the state space of the dynamical system (see also Alistair [2]).

Definition 3.1 *Let $X_1, \ldots, X_{n+1}$ be the first $n+1$ observations $(n \geq 1)$ of the dynamical system (1.1) and let $V_n(X_1), \ldots, V_n(X_n)$ be the Voronoi cells in K^d associated with $X_1, \ldots, X_n$. Further, let $\{T_{ij}^n : 1 \leq i \leq n, 1 \leq j \leq m_i\}$ be a set of d-simplices obtained by triangulating each Voronoi cell as in Theorem 2.2. Then the* (multidimensional) spacing *S_{ij}^n associated with T_{ij}^n is*

$$S_{ij}^n = \sup_{\{x \in T_{ij}^n\}} ||x - X_i||. \tag{3.2}$$

The (multidimensional) rate function *$r_n : K^d \to \mathbb{R}^{+*}$ is defined by*

$$r_n(x) = \frac{1}{S_n(x)},$$

where $S_n(x)$ is the spacing associated with the d-simplex $T_n(x) = T_{ij}^n$ containing x.

Convention For simplicity, and since no confusion is possible, the notation in the sequel ignores the dependence of T_{ij}^n, u_{ij}^n and S_{ij}^n on n. On the other hand, in order to underline the dependence of the observations on ϕ and x_1, we shall use when necessary the notation $X_i(\phi, x_1)$, $S_{ij}(\phi, x_1)$, $r_n(.\,;\phi, x_1)$, etc.

It is worth pointing out that the multidimensional rate r_n is a *piecewise constant function* which assigns the same value to each simplex T_{ij}. This property will prove extremely useful in the conjugation technique developed in the proofs of Theorem 3.1 and Proposition 3.2, see Section 5. Setting $T_{ij} = \text{conv}\ (X_i, t_2, \ldots, t_{d+1})$ and using the convexity of T_{ij}, we can rewrite the expression (3.2) defining the spacing S_{ij} as

$$S_{ij} = \max_{\{k=2,\ldots,d+1\}} ||t_k - X_i||.$$

In other words, the multidimensional spacing associated with T_{ij} is the length of the largest edge drawn from X_i. Whenever $d = 1$, the multidimensional rate function of Definition 3.1 coincides with the 1-dimensional rate function explicited in (3.1). Our geometrical viewpoint involves no order notion and provides a natural and intrinsic generalization of the results of Guerre and Maës [18]. Actually, many other definitions of multidimensional spacings have already been explored in the literature (see for example Deheuvels [14] and Janson [22]). We believe that definition (3.2) is the most adapted to our estimation framework (for related ideas, see Abdous and Berlinet [1]). The multidimensional results obtained in the present paper will strongly rely on the three following assumptions:

- **H1**(ϕ, x_1) The map ϕ and the initial condition x_1 are such that

$$\lim_{n\to\infty} \max_{\substack{1\le i\le n\\ 1\le j\le m_i}} S_{ij}(\phi, x_1) = 0.$$

- **H2**(ϕ, x_1) There exists a positive real constant $\beta = \beta(\phi, x_1)$ independent of n such that

$$\max_{\substack{1\le i\le n\\ 1\le j\le m_i}} ||\mathrm{d}u_{ij}||\,||\mathrm{d}u_{ij}^{-1}|| \le \beta, \quad \forall n \in \mathbb{N}^*.$$

- **H3**(ϕ) ϕ is of class C^1 (*i.e.*, continuously differentiable) on K^d. (Here and throughout, regularity properties of a function defined on K^d are understood with respect to the relative topology on K^d).

Let us now discuss the three hypotheses. Hypothesis **H1**(ϕ, x_1) requires that the maximal spacing tends to 0 as n tends to infinity. As a consequence, **H1**(ϕ, x_1) does not allow for ties among the X_i's, as such a situation would imply a stationary behavior of the process. Clearly, **H1**(ϕ, x_1) is equivalent to

$$\lim_{n\to\infty} \inf_{x\in K^d} r_n(x\,;\phi, x_1) = +\infty,$$

which means that the multidimensional rate function $r_n(.\,;\phi,x_1)$ asymptotically diverges. If we denote by diam $T_{ij}(\phi,x_1)$ the diameter of the d-simplex $T_{ij}(\phi,x_1)$, *i.e.*,

$$\text{diam } T_{ij}(\phi,x_1) = \sup_{\{x,y\,\in\,T_{ij}\}} ||x-y||,$$

we have the following results, whose proofs are clear:

Proposition 3.1

$$\mathbf{H1}(\phi,x_1) \text{ holds if, and only if, } \lim_{n\to\infty} \max_{\substack{1\le i\le n \\ 1\le j\le m_i}} \text{diam } T_{ij}(\phi,x_1) = 0.$$

Corollary 3.1

$$\mathbf{H1}(\phi,x_1) \text{ holds if, and only if, the sequence } \big\{X_i(\phi,x_1)\big\}_{i\ge 1} \text{ is dense in } K^d.$$

Necessary and sufficient conditions for the sequence $\big\{X_i(\phi,x_1)\big\}_{i\ge 1}$ to be dense in K^d can be found in Block and Coppel [5]. See also Delecroix, Guégan and Léorat [15] for sufficient conditions when the dynamical system (1.1) is ergodic and admits an invariant measure. Roughly speaking, $\mathbf{H1}(\phi,x_1)$ means that if we want to properly estimate ϕ, then the set of observations has to fill in asymptotically the space K^d. With this respect, observe that in the case $d=2$, setting $X_n=(X_{1n},X_{2n})$ and $\phi(.)=\big(\phi_1(.),\phi_2(.)\big)$, then $\mathbf{H1}(\phi,x_1)$ cannot be fulfilled if there exists a map $g(.)$ such that $X_{2n}-g(X_{1n})\to 0$ as n grows, that is when X_{2n} is a target for $g(X_{1n})$.

Setting $u_{ij}(x)=A_{ij}(x)+B_{ij}$ (A_{ij} matrix of size $d\times d$ and B_{ij} matrix of size $d\times 1$), it is well know that

$$||\mathrm{d}u_{ij}|| = \Big[\max\{\lambda\,:\,\lambda \text{ eigenvalue of } A'_{ij}A_{ij}\}\Big]^{1/2}$$

and

$$||\mathrm{d}u_{ij}^{-1}|| = \Big[\min\{\lambda\,:\,\lambda \text{ eigenvalue of } A'_{ij}A_{ij}\}\Big]^{-1/2},$$

where A'_{ij} denotes the transpose matrix of A_{ij}. Thus, condition $\mathbf{H2}(\phi,x_1)$ expresses that the largest and smallest eigenvalue of $A'_{ij}A_{ij}$ are of the same order. For example, when $d=2$ and when the triangle $T_{ij}=\text{conv }(X_i,t_1,t_2)$ is rectangle in X_i (A_{ij} diagonal), then $\mathbf{H2}(\phi,x_1)$ means that the edge lengths $||t_1-X_i||$ and $||t_2-X_i||$ are of the same order. Geometrically speaking, $\mathbf{H2}(\phi,x_1)$ means that the triangulation does not degenerate, *i.e.*, that any d-simplex of the triangulation collapses towards a $(d-1)$-dimensional one. It would be interesting to characterize $\mathbf{H2}(\phi,x_1)$ via the (topological or statistical) properties of the dynamical system ϕ. Observe however that $\mathbf{H2}(\phi,x_1)$ becomes trivial whenever $d=1$, since in this case

$$||\mathrm{d}u_{ij}^{-1}|| = \frac{1}{||\mathrm{d}u_{ij}||}.$$

Of course conditions $\mathbf{H1}(\phi,x_1)$ and $\mathbf{H2}(\phi,x_1)$ are not verifiable in practice with only a finite amount of observations. Condition $\mathbf{H3}(\phi)$ implies that ϕ is Lipschitz on K^d with

Lipschitz constant $||\mathrm{d}\phi||_\infty$ (as $\mathrm{d}\phi$ is continuous, $\mathrm{d}\phi$ is bounded on K^d and $||\mathrm{d}\phi||_\infty$ is well defined). Indeed, by the mean value theorem:

$$||\phi(x) - \phi(y)|| \leq ||\mathrm{d}\phi||_\infty \, ||x - y||, \quad \forall x, y \in K^d. \tag{3.3}$$

As ϕ is a continuous function from the compact and connected set K^d into itself, ϕ has a fixed point by the Brouwer fixed point theorem (Aubin [3]). Assuming that $||\mathrm{d}\phi||_\infty \leq 1$, one would get by (3.3) that any ball B centered at the fixed point would be invariant by ϕ (*i.e.*, $\phi(B) \subset B$). Therefore, any sequence entering B would be captured by B, and $\mathbf{H1}(\phi, x_1)$ would never hold by Corollary 3.1. Thus, under $\mathbf{H1}(\phi, x_1)$ and $\mathbf{H3}(\phi)$, we necessarily have $||\mathrm{d}\phi||_\infty > 1$.

3.2 Optimal multidimensional rate function

As stated in the introduction, we wish to compare estimators through a local risk minimax approach, the local risk of an estimator $\hat{\psi}_n$ being defined as (see Ibragimov and Has'minskii [21])

$$\sup_{\psi \in \mathcal{V}_\epsilon(\phi)} ||r_n(.\,; \psi, x_1)\big(\hat{\psi}_n(.\,; x_1) - \psi(.)\big)||_\infty,$$

where $\mathcal{V}_\epsilon(\phi)$ is the following neighborhood of ϕ:

$$\mathcal{V}_\epsilon(\phi) = \Big\{ \psi \in C^1(K^d) \ : \ ||\psi - \phi||_\infty \leq \epsilon, \, ||\mathrm{d}\psi||_\infty \leq 2\,||\mathrm{d}\phi||_\infty \Big\}, \tag{3.4}$$

and where $\hat{\psi}_n(.\,; x_1)$ is the estimation of ψ based on the observations $X_1, \ldots, X_{n+1}$ drawn from ψ with $X_1 = x_1$. We do not assume, as in Guerre and Maës [18], that $\mathbf{H1}(\psi, x_1)$ holds in (3.4) because this restricts unnecessarily the size of $\mathcal{V}_\epsilon(\phi)$. The next theorem establishes that the minimax risk remains asymptotically bounded away from 0 and infinity. Moreover, the lower bound shows that r_n is an optimal rate for estimation.

Theorem 3.1 *Assume that* $\mathbf{H1}(\phi, x_1)$, $\mathbf{H2}(\phi, x_1)$ *and* $\mathbf{H3}(\phi)$ *hold. Assume that there exists z in K^d such that $\mathrm{d}\phi(z)$ is an isomorphism and such that $||\mathrm{d}\phi^{-1}||_\infty < 1$, where ϕ^{-1} is a C^1 local inverse of ϕ defined on a neighborhood of $\phi(z)$. Then there exists a positive real constant A independent of ϕ such that, for large enough n,*

$$A \left(\frac{1}{||\mathrm{d}\phi^{-1}||_\infty} - 1 \right) \leq \inf_{\hat{\psi}_n} \sup_{\psi \in \mathcal{V}_\epsilon(\phi)} ||r_n(.\,; \psi, x_1)\big(\hat{\psi}_n(.\,; x_1) - \psi(.)\big)||_\infty \leq 2\,||\mathrm{d}\phi||_\infty, \tag{3.5}$$

where the infimum is taken over the class of all possible estimators $\hat{\psi}_n$ of ψ.

The existence of a C^1 local inverse of ϕ is guaranteed by the local inversion theorem, and $||\mathrm{d}\phi^{-1}||_\infty$ stands for the supremum of $\mathrm{d}\phi^{-1}$ on $\phi\big(\mathcal{V}(z)\big)$, where $\mathcal{V}(z)$ is a neighborhood of z.

Let us consider the 1-dimensional case. Whenever $d = 1$, by compactness of $K^1 = [0,1]$ and by continuity of ϕ', we can choose z in $[0,1]$ such that

$$|\phi'(z)| = ||\phi'||_\infty > 1$$

$\big($remember that $||\mathrm{d}\phi||_\infty > 1$ under $\mathbf{H1}(\phi, x_1)$ and $\mathbf{H3}(\phi)\big)$. Even if it means restricting the size of $\mathcal{V}(z)$, we can assume that $\mathcal{V}(z)$ is compact and that

$$|\phi'(x)| > 1 \text{ for all } x \text{ in } \mathcal{V}(z). \tag{3.6}$$

By compactness of $\phi\big(\mathcal{V}(z)\big)$ and by continuity of $(\phi^{-1})'$, we deduce the existence of a z' in $\mathcal{V}(z)$ such that

$$||(\phi^{-1})'||_\infty = |\big(\phi^{-1}\big)'\big(\phi(z')\big)|. \tag{3.7}$$

As we know that

$$\big(\phi^{-1}\big)'\big(\phi(x)\big) \times \phi'(x) = 1 \text{ for all } x \text{ in } \mathcal{V}(z),$$

equations (3.6) and (3.7) lead to

$$||(\phi^{-1})'||_\infty < 1.$$

We have just shown that the conditions on $\mathrm{d}\phi$ and $\mathrm{d}\phi^{-1}$ in Theorem 3.1 are automatically satisfied in the 1-dimensional context. Note that these conditions are fulfilled by expanding dynamical systems $\big($in the sense of Lasota and Mackey [23]$\big)$.

An immediate consequence of Theorem 3.1 is that the multidimensional rate function r_n is the fastest rate of convergence which can be achieved by any estimator in our framework.

3.3 A multidimensional rate-efficient estimator

In this paragraph, we explicit an estimator $\tilde{\psi}_n$ that achieves the optimal multidimensional rate r_n. Precisely, $\tilde{\psi}_n$ is the 1-nearest neighbor estimator $\big($Bosq and Lecoutre [7]$\big)$ previously used by Delecroix, Guégan and Léorat [15] in the estimation of the Lyapunov exponent $\big($see also Farmer and Sidorowich [17] and Casdagli [10], [11]$\big)$. Formally, the 1-nearest neighbor estimator $\tilde{\psi}_n$ of a dynamical system $\psi : K^d \to K^d$ is defined by

$$\tilde{\psi}_n(x) = X_{i+1} \text{ whenever } V_n(x, \psi, x_1) = V_n(X_i, \psi, x_1), \quad i = 1, \dots, n, \tag{3.8}$$

where $X_1, \dots, X_{n+1}$ are the first $n+1$ observations drawn from ψ with $X_1 = x_1$. The result is as follows:

Proposition 3.2 *Let $\psi : K^d \to K^d$ be such that $\mathbf{H3}(\psi)$ holds and let $\tilde{\psi}_n$ be the 1-nearest neighbor estimator defined in (3.8). Then, for $n \geq 1$,*

$$||r_n(\,.\,;\psi, x_1)\big(\tilde{\psi}_n(\,.\,;x_1) - \psi(.)\big)||_\infty \leq ||\mathrm{d}\psi||_\infty. \tag{3.9}$$

Inequality (3.9) shows that the 1-nearest neighbor estimator $\tilde{\psi}_n$ is rate-efficient. Since any ψ in $\mathcal{V}_\epsilon(\phi)$ satisfies $\mathbf{H3}(\psi)$ and verifies the inequality $||\mathrm{d}\psi||_\infty \leq 2\,||\mathrm{d}\phi||_\infty$, the upper bound in (3.5) is a straightforward consequence of (3.9). Under $\mathbf{H3}(\psi)$, (3.9) leads also to

$$||\tilde{\psi}_n(.\,;x_1) - \psi(.)||_\infty \leq \frac{||\mathrm{d}\psi||_\infty}{\inf\limits_{x\in K^d} r_n(x\,;\psi, x_1)},$$

which shows that $\tilde{\psi}_n$ approximates ψ up to an error bounded by $||\mathrm{d}\psi||_\infty / \inf\limits_{x\in K^d} r_n(x\,;\psi, x_1)$. In other words,

$$||\tilde{\psi}_n(.\,;x_1) - \psi(.)||_\infty = \mathrm{O}\left(\frac{1}{\inf\limits_{x\in K^d} r_n(x\,;\psi, x_1)}\right).$$

Moreover, if we set $\psi = \phi$ and if we assume further that $\mathbf{H1}(\phi, x_1)$ holds, we obtain that the estimator $\tilde{\phi}_n$ uniformly converges towards ϕ at the optimal rate r_n.

4 An example

Assume that $d = 2$ and consider the map $\phi : K^2 \to K^2$ defined by

$$\phi(x, y) = (3x + y, x + 3y) \pmod 1.$$

Here, the notation $ru \pmod 1$ means $ru - [ru]$, where $[ru]$ is the integer part of ru, *i.e.*, the largest integer q such that $ru - q \geq 0$. The map ϕ is defined on K^2 and continuously differentiable on $K^2_* = K^2 - (C_1 \cup \ldots \cup C_p)$, where the C_i's are non-overlapping convex subsets of K^2. The results obtained in the present paper hold for functions of class C^1 on the set K^2. Slight modifications of the proofs show however that our results are still valid with K^2 replaced by K^2_*.

The transformation ϕ is shown to be *exact* (Lasota and Mackey [23]). As an important consequence, ϕ is ergodic and preserves the Lebesgue measure. Thus, from Delecroix, Guégan and Léorat [15], we know that the sequence $\{X_i(\phi, z_1)\}_{i\geq 1}$ is dense in K^2_* for λ-almost z_1 in K^2_*. From Corollary 3.1, we deduce that $\mathbf{H1}(\phi, z_1)$ holds for λ-almost z_1 in K^2_*. As an illustration, Figure 3 depicts the first $n = 100$ and $n = 1000$ iterates of an initial point z_1, as well as the associated Voronoi polygons. The Jacobian matrix $\mathcal{J}_\phi$ of ϕ does not depend on the differentiation point z in K^2_* and takes the form

$$\mathcal{J}_\phi = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}.$$

Since $\det(\mathcal{J}_\phi) = 8 \neq 0$, ϕ is a C^1-diffeomorphism in the neighborhood of any z in K^2_*. Elementary linear algebra shows that

$$\mathcal{J}_{\phi^{-1}} = \frac{1}{8}\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix},$$

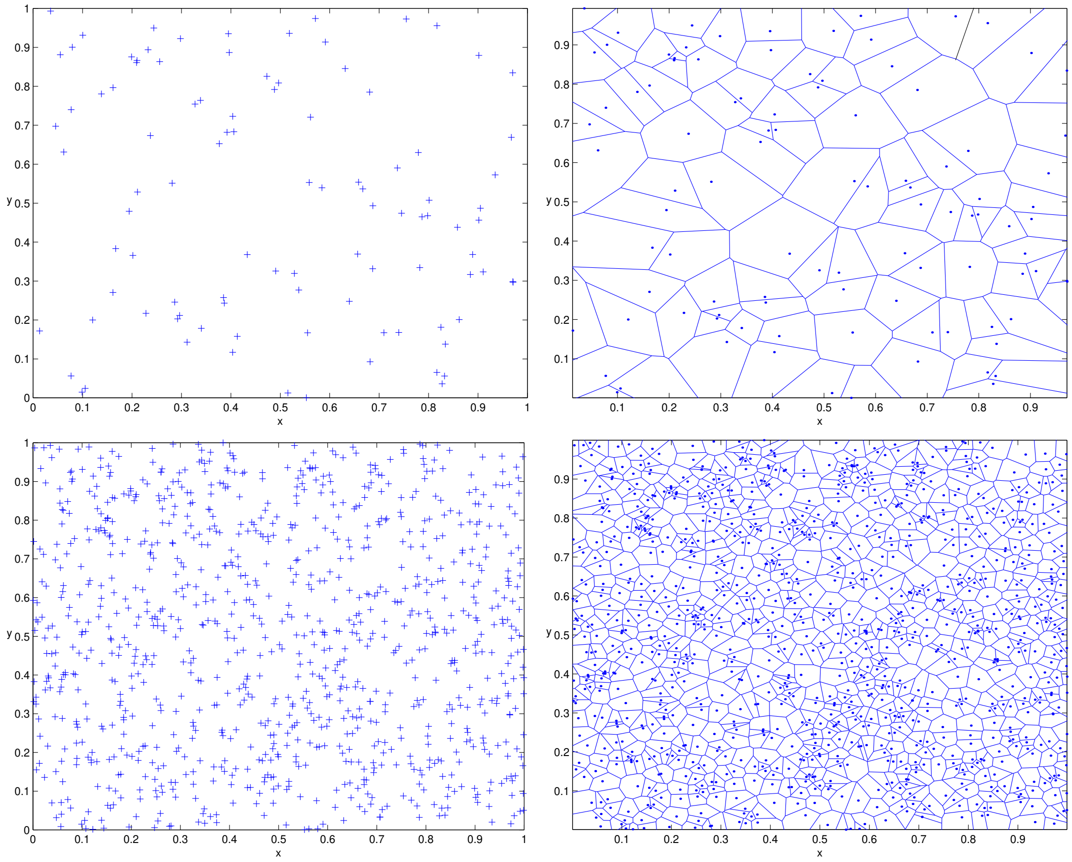

Figure 3: **TOP** First $\mathbf{n} = \mathbf{100}$ iterates of an initial point $\mathbf{z_1}$ (left) and associated Voronoi cells (right). **BOTTOM** First $\mathbf{n} = \mathbf{1000}$ iterates of $\mathbf{z_1}$ (left) and associated Voronoi cells (right).

and, consequently,

$$||\mathrm{d}\phi^{-1}||_\infty = \frac{1}{2} < 1.$$

Thus, provided $\mathbf{H2}(\phi, z_1)$ is satisfied, the minimax inequalities (3.5) hold with $||\mathrm{d}\phi||_\infty = 4$.

5 Proofs of the results

5.1 Proof of Theorem 3.1

In order to prove Theorem 3.1, we extend the conjugation technique used by Guerre and Maës [18]. The key idea is to study a family of functions embedded in the neighborhood $\mathcal{V}_\epsilon(\phi)$ of ϕ and which cannot be distinguished from ϕ on the basis of the first n observations.

For that purpose, let $\xi : \mathbb{R}^d \to \mathbb{R}^d$ be a non-constant function of class C^1 that vanishes outside the standard d-simplex Δ_d (note that this implies that $0 < ||\xi||_\infty, ||\mathrm{d}\xi||_\infty < \infty$). We recall that for each d-simplex T_{ij} $\big(= T_{ij}(\phi, x_1)\big)$ we denote by u_{ij} $\big(= u_{ij}(\phi, x_1)\big)$ an affine isomorphism such that $u_{ij}(T_{ij}) = \Delta_d$ and $u_{ij}(X_i) = 0$. We define for each i and j, $i = 1, \dots, n$, $j = 1, \dots, m_i$, a mapping $h_{ij} : \mathbb{R}^d \to \mathbb{R}^d$ by

$$h_{ij} = h_{ij}(\phi, x_1) = I + C_n S_{ij}\, \xi \circ u_{ij}, \tag{5.1}$$

where I is the identity map on $\mathbb{R}^d$, C_n is a positive real constant to be specified later on, and $\circ$ denotes the composition product of maps. Since ξ and u_{ij} are both of class C^1 on $\mathbb{R}^d$, h_{ij} is also of class C^1 and, from (5.1),

$$\mathrm{d}h_{ij}(x) = I + C_n S_{ij}\, \mathrm{d}\xi\big(u_{ij}(x)\big) \circ \mathrm{d}u_{ij}, \quad \forall x \in \mathbb{R}^d. \tag{5.2}$$

(Remember that $\mathrm{d}u_{ij}(x)$ does not depend on the particular point x and is abbreviated into $\mathrm{d}u_{ij}$). We state now a technical lemma.

Lemma 5.1 *Assume that* $\mathbf{H2}(\phi, x_1)$ *holds. Let* $i \in \{1, \dots, n\}$, *let* $j \in \{1, \dots, m_i\}$ *and let* $h_{ij} : \mathbb{R}^d \to \mathbb{R}^d$ *be defined as in (5.1). If*

$$C_n < \frac{1}{\beta\, ||\mathrm{d}\xi||_\infty}, \tag{5.3}$$

then

(a) *h_{ij} is a C^1-diffeomorphism from $\mathbb{R}^d$ onto $\mathbb{R}^d$.*

(b) *Any d-simplex T of the triangulation $\big\{T_{ij} : 1 \le i \le n,\, 1 \le j \le m_i\big\}$ is* globally *invariant under h_{ij},* i.e., *$h_{ij}(T) = T$.*

(c) *The unit hypercube K^d is* globally *invariant under h_{ij},* i.e., *$h_{ij}(K^d) = K^d$.*

Proof of Lemma 5.1

We already know from (5.1) that h_{ij} is a C^1 map from $\mathbb{R}^d$ into $\mathbb{R}^d$ whose differential is given by (5.2). Now, using the mean value theorem, for all x in T_{ij},

$$||x - X_i|| \le ||u_{ij}(x) - u_{ij}(X_i)||\, ||\mathrm{d}u_{ij}^{-1}||. \tag{5.4}$$

As $u_{ij}(X_i) = 0$ and $u_{ij}(x)$ belongs to Δ_d, we have $||u_{ij}(x) - u_{ij}(X_i)|| \le 1$ and therefore, from (5.4),

$$||x - X_i|| \le ||\mathrm{d}u_{ij}^{-1}||.$$

As a consequence,

$$S_{ij}\, ||\mathrm{d}u_{ij}|| \le ||\mathrm{d}u_{ij}||\, ||\mathrm{d}u_{ij}^{-1}||. \tag{5.5}$$

From (5.5), we deduce, under $\mathbf{H2}(\phi, x_1)$, that for all x in $\mathbb{R}^d$,

$$||C_n S_{ij}\, \mathrm{d}\xi\big(u_{ij}(x)\big) \circ \mathrm{d}u_{ij}|| \le C_n \beta\, ||\mathrm{d}\xi||_\infty, \tag{5.6}$$

so that, if (5.3) holds,

$$||C_n S_{ij}\, \mathrm{d}\xi\big(u_{ij}(x)\big) \circ \mathrm{d}u_{ij}|| < 1 \quad \text{for all } x \text{ in } \mathbb{R}^d.$$

As is well known, it follows from the last inequality that $\mathrm{d}h_{ij}(x)$ is an isomorphism and thus, by the local inversion theorem, that h_{ij} is a *local* C^1-diffeomorphism. Now, let x and y in $\mathbb{R}^d$ be such that

$$h_{ij}(x) = h_{ij}(y).$$

From (5.1),

$$||x-y|| = C_n S_{ij}\, ||\xi \circ u_{ij}(y) - \xi \circ u_{ij}(x)||,$$

so that, by the mean value theorem, (5.5) and $\mathbf{H2}(\phi, x_1)$,

$$||x-y|| \le C_n \beta\, ||\mathrm{d}\xi||_\infty\, ||x-y||.$$

This leads, under the hypothesis (5.3), to $||x-y|| = 0$, *i.e.*, $x = y$. The map h_{ij} is thus seen to be one to one on its image. As it is also a local C^1-diffeomorphism, it is actually a *global* C^1-diffeomorphism from $\mathbb{R}^d$ onto $h_{ij}(\mathbb{R}^d)$. To see that $h_{ij}(\mathbb{R}^d) = \mathbb{R}^d$, note that since ξ vanishes outside Δ_d and since $u_{ij}(T_{ij}) = \Delta_d$,

$$h_{ij} = I \ \text{ on } T_{ij}^c, \tag{5.7}$$

where the notation T_{ij}^c stands for the complementary set of T_{ij} in $\mathbb{R}^d$. Further, we know that $\mathbb{R}^d$ is simply connected and that h_{ij} is an homeomorphism. This implies that $h_{ij}(\mathbb{R}^d)$ is also simply connected. This remark and (5.7) lead to $h_{ij}(\mathbb{R}^d) = \mathbb{R}^d$ and (a) is proved. Points (b) and (c) are straightforward from (5.7).

■

The next step of the proof of Theorem 3.1 is contained in the following proposition:

Proposition 5.1 *Assume that* $\mathbf{H2}(\phi, x_1)$ *and* $\mathbf{H3}(\phi)$ *hold. Let* $\mathcal{V}_\epsilon(\phi)$ *and* h_{ij} *be defined as in (3.4) and (5.1). Assume that*

$$C_n \le \min\left\{ \frac{1}{3\beta\, ||\mathrm{d}\xi||_\infty}, \frac{\epsilon}{||\xi||_\infty\, (1 + ||\mathrm{d}\phi||_\infty) \max\limits_{\substack{1\le i\le n \\ 1\le j\le m_i}} S_{ij}} \right\}. \tag{5.8}$$

For $1 \le i \le n$ *and* $1 \le j \le m_i$, *define* $\phi_{ij} = h_{ij} \circ \phi \circ h_{ij}^{-1}$. *Then*

(a) *The first* n *observations of model (1.1) associated with the maps* ϕ *and* ϕ_{ij} *are the same,* i.e.,

$$X_k(\phi, x_1) = X_k(\phi_{ij}, x_1), \quad k = 1, \ldots, n.$$

(b) *We have*

$$||h_{ij} - I||_\infty = C_n S_{ij}\, ||\xi||_\infty.$$

(c) *The map ϕ_{ij} is in $\mathcal{V}_\epsilon(\phi)$.*

Proof of Proposition 5.1

First note that condition (5.8) implies condition (5.3), so that, under $\mathbf{H2}(\phi, x_1)$, h_{ij} is a C^1-diffeomorphism from $\mathbb{R}^d$ onto $\mathbb{R}^d$ by Lemma 5.1. From (c) of the former lemma, we deduce that ϕ_{ij} is well defined as a mapping from K^d to K^d. To show (a), observe simply that $h_{ij}(X_k) = X_k$ (for $k \neq i$ this is a direct consequence of (5.7), and for $k = i$, $u_{ij}(X_i) = 0$ and $\xi(0) = 0$ by continuity of ξ on the boundary of Δ_d). Consequently, the sequence $X_1(\phi, x_1), \ldots, X_n(\phi, x_1)$ generated by ϕ is identical to the sequence $X_1(\phi_{ij}, x_1), \ldots, X_n(\phi_{ij}, x_1)$ generated by ϕ_{ij}.

The equality $||h_{ij} - I||_\infty = C_n S_{ij}\, ||\xi||_\infty$ in (b) is a straightforward consequence of (5.1). The only non-trivial point is (c). Under $\mathbf{H3}(\phi)$, ϕ is of class C^1 on K^d, so that the same is true for ϕ_{ij}. We have now to prove that $||\phi_{ij} - \phi||_\infty \leq \epsilon$ and that $||\mathrm{d}\phi_{ij}||_\infty \leq 2\, ||\mathrm{d}\phi||_\infty$ (see (3.4) for the definition of the neighborhood $\mathcal{V}_\epsilon(\phi)$). Using the invariance of the uniform norm under a change of scale, the triangle inequality and the mean value theorem, we can write:

$$\begin{aligned} ||\phi_{ij} - \phi||_\infty = ||h_{ij} \circ \phi \circ h_{ij}^{-1} - \phi||_\infty &= ||h_{ij} \circ \phi - \phi \circ h_{ij}||_\infty \\ &\leq ||h_{ij} \circ \phi - \phi||_\infty + ||\phi - \phi \circ h_{ij}||_\infty \\ &\leq ||h_{ij} - I||_\infty + ||\mathrm{d}\phi||_\infty\, ||I - h_{ij}||_\infty, \end{aligned}$$

and thus, from point (b),

$$||\phi_{ij} - \phi||_\infty \leq \big(1 + ||\mathrm{d}\phi||_\infty\big) C_n S_{ij}\, ||\xi||_\infty \leq \epsilon, \tag{5.9}$$

where the last inequality follows from (5.8). Moreover, by composition rule for differentiation, for all x in K^d,

$$\mathrm{d}\phi_{ij}(x) = \mathrm{d}h_{ij}\big(\phi \circ h_{ij}^{-1}(x)\big) \circ \mathrm{d}\phi\big(h_{ij}^{-1}(x)\big) \circ \mathrm{d}h_{ij}^{-1}(x).$$

Consequently,

$$||\mathrm{d}\phi_{ij}||_\infty \leq ||\mathrm{d}h_{ij}||_\infty\, ||\mathrm{d}\phi||_\infty\, ||\mathrm{d}h_{ij}^{-1}||_\infty. \tag{5.10}$$

From (5.2), under (5.3), we obtain, for all x in $\mathbb{R}^d$,

$$\big(\mathrm{d}h_{ij}(x)\big)^{-1} = \sum_{p=0}^{\infty} (-1)^p \big[C_n S_{ij}\, \mathrm{d}\xi\big(u_{ij}(x)\big) \circ \mathrm{d}u_{ij}\big]^p.$$

Therefore, as $\big(\mathrm{d}h_{ij}(x)\big)^{-1} = \mathrm{d}h_{ij}^{-1}\big(h_{ij}(x)\big)$, from (5.6),

$$||\mathrm{d}h_{ij}^{-1}||_\infty \leq \frac{1}{1 - C_n \beta\, ||\mathrm{d}\xi||_\infty}. \tag{5.11}$$

Finally, (5.10), (5.2), (5.6) and (5.11) lead to

$$||\mathrm{d}\phi_{ij}||_\infty \leq \frac{1 + C_n\beta\,||\mathrm{d}\xi||_\infty}{1 - C_n\beta\,||\mathrm{d}\xi||_\infty}\,||\mathrm{d}\phi||_\infty.$$

Consequently, under assumption (5.8),

$$||\mathrm{d}\phi_{ij}||_\infty \leq 2\,||\mathrm{d}\phi||_\infty. \tag{5.12}$$

Inequalities (5.9) and (5.12) show that ϕ_{ij} is in the neighborhood $\mathcal{V}_\epsilon(\phi)$.

■

We shall now complete the proof of Theorem 3.1. From now on, we assume that $\mathbf{H1}(\phi, x_1)$, $\mathbf{H2}(\phi, x_1)$ and $\mathbf{H3}(\phi)$ are satisfied. Observe that, under $\mathbf{H1}(\phi, x_1)$, the assumption (5.8) is automatically satisfied if $C_n = 1/(3\beta\,||\mathrm{d}\xi||_\infty)$ and n is large enough. Thus, in the sequel, we assume that n is large enough so that we can take $C_n = C = 1/(3\beta\,||\mathrm{d}\xi||_\infty)$.

Consider a function h_{ij} as in (5.1) and define ϕ_{ij} as in Proposition 5.1. Appropriate indexes i and j will be chosen below. By point (a) of Proposition 5.1, $X_k(\phi, x_1) = X_k(\phi_{ij}, x_1)$, $k = 1, \ldots, n$. As a consequence, we have, for all $n \geq 1$,

$$r_n(\,.\,;\phi, x_1) = r_n(\,.\,;\phi_{ij}, x_1) \tag{5.13}$$

and, for any estimator $\hat{\phi}_n(\,.\,;x_1)$ of ϕ based on the observations $X_1, \ldots, X_{n+1}$ with $X_1 = x_1$,

$$\hat{\phi}_n(\,.\,;x_1) = \widehat{(\phi_{ij})}_n(\,.\,;x_1). \tag{5.14}$$

For clarity sake we denote by

$$\widehat{\mathcal{R}}_{n,\epsilon}(\phi, x_1)$$

the quantity

$$\sup_{\psi \in \mathcal{V}_\epsilon(\phi)} ||r_n(\,.\,;\psi, x_1)\big(\hat{\psi}_n(\,.\,;x_1) - \psi(.)\big)||_\infty.$$

We obtain, from (c) of Proposition 5.1,

$$\begin{aligned}
\widehat{\mathcal{R}}_{n,\epsilon}(\phi, x_1) &\geq \sup_{\psi \in \{\phi, \phi_{ij}\}} ||r_n(\,.\,;\psi, x_1)\big(\hat{\psi}_n(\,.\,;x_1) - \psi(.)\big)||_\infty \\
&\geq \frac{1}{2}\Big[||r_n(\,.\,;\phi, x_1)\big(\hat{\phi}_n(\,.\,;x_1) - \phi(.)\big)||_\infty + ||r_n(\,.\,;\phi_{ij}, x_1)\big(\widehat{(\phi_{ij})}_n(\,.\,;x_1) - \phi_{ij}(.)\big)||_\infty\Big] \\
&= \frac{1}{2}\Big[||r_n(\,.\,;\phi, x_1)\big(\hat{\phi}_n(\,.\,;x_1) - \phi(.)\big)||_\infty + ||r_n(\,.\,;\phi, x_1)\big(\hat{\phi}_n(\,.\,;x_1) - \phi_{ij}(.)\big)||_\infty\Big],
\end{aligned}$$

where the last equality follows from (5.13) and (5.14). Consequently, by the triangle inequality:

$$\widehat{\mathcal{R}}_{n,\epsilon}(\phi, x_1) \geq \frac{1}{2}\,||r_n(\,.\,;\phi, x_1)\big(\phi(.) - \phi_{ij}(.)\big)||_\infty,$$

and, by the invariance of the uniform norm under a change of scale:

$$\widehat{\mathcal{R}}_{n,\epsilon}(\phi, x_1) \geq \frac{1}{2}\,||r_n(.\,;\phi, x_1)\big(\phi \circ h_{ij}(.) - h_{ij} \circ \phi(.)\big)||_\infty, \tag{5.15}$$

observing that $r_n \circ h_{ij} = r_n$ by point (b) of Lemma 5.1. Let us now select an appropriate d-simplex T_{ij} for the definition of h_{ij}. Following the assumption in Theorem 3.1 there exists z in K^d such that $\mathrm{d}\phi(z)$ is an isomorphism. According to the local inversion theorem, there exists a neighborhood $\mathcal{V}(z)$ of z in K^d such that ϕ is a C^1-diffeomorphism from $\mathcal{V}(z)$ onto $\phi\big(\mathcal{V}(z)\big)$. Even if it means restricting the size of $\phi\big(\mathcal{V}(z)\big)$, we can assume that $\phi\big(\mathcal{V}(z)\big)$ is convex. For sake of clarity, we denote by $||\psi||_z$ the supremum on $\mathcal{V}(z)$ of any bounded function ψ on $\mathcal{V}(z)$. By construction of the triangulation (Theorem 2.2), for each n in $\mathbb{N}^*$, there exists a (uniquely determined, see the end of Section 2) d-simplex $T_n(z, \phi, x_1)$ that contains z. We denote by h_n the corresponding function h_{ij} and we observe that $r_n(x\,;\phi, x_1) = r_n(z\,;\phi, x_1)$ for all x in $T_n(z, \phi, x_1)$. Consequently, from inequality (5.15),

$$\begin{aligned}
\widehat{\mathcal{R}}_{n,\epsilon}(\phi, x_1) &\geq \frac{r_n(z\,;\phi, x_1)}{2}\,||\phi \circ h_n - h_n \circ \phi||_z \\
&\geq \frac{r_n(z\,;\phi, x_1)}{2}\,\big(||\phi \circ h_n - \phi||_z - ||\phi - h_n \circ \phi||_z\big) \\
&\geq \frac{r_n(z\,;\phi, x_1)}{2}\,\big(||\phi \circ h_n - \phi||_z - ||\phi - h_n \circ \phi||_\infty\big),
\end{aligned}$$

and thus

$$\widehat{\mathcal{R}}_{n,\epsilon}(\phi, x_1) \geq \frac{r_n(z\,;\phi, x_1)}{2}\,\big(||\phi \circ h_n - \phi||_z - ||I - h_n||_\infty\big). \tag{5.16}$$

The last step of the proof consists in finding a lower bound for the item $||\phi \circ h_n - \phi||_z$. By Proposition 3.1, we have $T_n(z, \phi, x_1) \subset \mathcal{V}(z)$ for large enough n, so that, by the mean value theorem ($\phi\big(\mathcal{V}(z)\big)$ is convex)

$$||\mathrm{d}\phi^{-1}||_\infty\,||\phi \circ h_n - \phi||_z \geq ||h_n - I||_z,$$

i.e.,

$$||\phi \circ h_n - \phi||_z \geq \frac{1}{||\mathrm{d}\phi^{-1}||_\infty}\,||h_n - I||_z.$$

Note that $\mathrm{d}\phi^{-1}\big(\phi(z)\big) = \big(\mathrm{d}\phi(z)\big)^{-1}$, so that $||\mathrm{d}\phi^{-1}||_\infty > 0$. As $T_n(z, \phi, x_1) \subset \mathcal{V}(z)$ for large enough n, and as $h_n - I$ vanishes outside $T_n(z, \phi, x_1)$ $\big($equality (5.7)$\big)$, we obtain, for large enough n,

$$||\phi \circ h_n - \phi||_z \geq \frac{1}{||\mathrm{d}\phi^{-1}||_\infty}\,||h_n - I||_\infty.$$

This, together with (5.16), gives

$$\widehat{\mathcal{R}}_{n,\epsilon}(\phi, x_1) \geq \frac{r_n(z\,;\phi, x_1)}{2}\left(\frac{1}{||\mathrm{d}\phi^{-1}||_\infty} - 1\right)||h_n - I||_\infty.$$

Consequently, from Proposition 5.1, point (b),

$$\widehat{\mathcal{R}}_{n,\epsilon}(\phi, x_1) \geq \frac{C||\xi||_\infty}{2}\left(\frac{1}{||\mathrm{d}\phi^{-1}||_\infty} - 1\right),$$

for large enough n. This proves that the lower bound of inequalities (3.5) holds. The upper bound is a consequence of Proposition 3.2 proven in the next paragraph.

5.2 Proof of Proposition 3.2

Let ψ be a dynamical system of K^d into K^d such that $\mathbf{H3}(\psi)$ holds. Let x in K^d and let $X_1, \ldots, X_{n+1}$ be the first $n+1$ observations drawn from ψ with $X_1 = x_1$. Assume that the Voronoi cell $V_n(x, \psi, x_1) = V_n(X_i, \psi, x_1)$ and that the d-simplex $T_n(x, \psi, x_1) = T_{ij}$. For $n \geq 1$, we have, by definition of r_n and $\tilde{\psi}_n(.\,; x_1)$,

$$||r_n(x\,; \psi, x_1)\big(\tilde{\psi}_n(x\,; x_1) - \psi(x)\big)|| = \frac{1}{S_{ij}}\,||\psi(X_i) - \psi(x)||,$$

so that, by the mean value theorem,

$$||r_n(x\,; \psi, x_1)\big(\tilde{\psi}_n(x\,; x_1) - \psi(x)\big)|| \leq \frac{||x - X_i||}{S_{ij}}\,||\mathrm{d}\psi||_\infty.$$

Consequently, by the very definition of the spacings,

$$||r_n(.\,; \psi, x_1)\big(\tilde{\psi}_n(.\,; x_1) - \psi(.)\big)||_\infty \leq ||\mathrm{d}\psi||_\infty.$$

Acknowledgments. The authors thank an anonymous referee for his careful reading of the paper. His suggestions to reorganize the manuscript led to a more readable version. Gérard Biau is indebted to Vincent Cavalier for many stimulating conversations on geometrical topics.

References

[1] Abdous, B. and Berlinet, A. (1986) Convergence uniforme presque sûre d'une classe d'estimateurs à noyau, pour la densité, *Comptes Rendus Mathématiques de l'Académie des Sciences de Paris*, **Vol. 303**, pp. 761–764.

[2] Alistair, I.M. (1992) Tessellations and dynamical systems, in *Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity, Proceedings*, ed. M. Casdagli and S. Eubank, **Vol. 12**, pp. 3–24, Addison–Wesley.

[3] Aubin, J.P. (1998) *Optima and Equilibria: an Introduction to Nonlinear Analysis*, Springer–Verlag, New York.

[4] Berliner, L.M. (1992) Statistics, probability and chaos (with discussion), *Statistical Science*, **Vol. 7**, pp. 69–122.

[5] Block, L.S. and Coppel, W.A. (1992) *Dynamics in One Dimension*, Springer–Verlag, Berlin.

[6] Bosq, D. and Guégan, D. (1995) Nonparametric estimation of the chaotic function and the invariant measure of a dynamical system, *Statistics and Probability Letters*, **Vol. 25**, pp. 201–212.

[7] Bosq, D. and Lecoutre, J.P. (1987) *Théorie de l'Estimation Fonctionnelle*, Economica, Paris.

[8] Boyarsky, A. and Góra, P. (1997) *Laws of Chaos – Invariant Measures and Dynamical Systems in One Dimension*, Birkhäuser, Boston.

[9] Brøndsted, A. (1983) *An Introduction to Convex Polytopes*, Springer–Verlag, New York.

[10] Casdagli, M. (1989) Nonlinear prediction of chaotic time series, *Physica D*, **Vol. 35**, pp. 335–356.

[11] Casdagli, M. (1991) Chaos and deterministic versus stochastic non-linear modelling, *Journal of the Royal Statistical Society*, **Vol. B54**, pp. 303–328.

[12] Chatterjee, S. and Yilmaz, M.R. (1992) Chaos, fractals and statistics (with discussion), *Statistical Science*, **Vol. 7**, pp. 49–122.

[13] Collet, P. (1996) Some ergodic properties of maps of the interval, in *Dynamical Systems* (*Temuco, 1991/1992*), Travaux en Cours, **Vol. 52**, pp. 55–91, Hermann, Paris.

[14] Deheuvels, P. (1983) Strong bounds for multidimensional spacings, *Zeitschrift für Warscheinlichkeitstheorie und Verwandte Gebiete*, **Vol. 64**, pp. 411–424.

[15] Delecroix, M., Guégan, D. and Léorat, G. (1997) Determinating Lyapunov exponents in deterministic dynamical systems, *Computational Statistics*, **Vol. 12**, pp. 93–107.

[16] Devroye, L. (1987) *A Course in Density Estimation*, Birkhäuser, Boston.

[17] Farmer, J.D. and Sidorowich, J.J. (1987) Predicting chaotic time series, *Physical Review Letters*, **Vol. 59**, pp. 845–848.

[18] Guerre, E. and Maës, J. (1998) Optimal rate for nonparametric estimation in deterministic dynamical systems, *Statistical Inference for Stochastic Processes*, **Vol. 1**, pp. 157–173.

[19] Guégan, D. (1994) *Séries Chronologiques Non Linéaires à Temps Discret*, Economica, Paris.

[20] Hall, P. and Wolff, R.C.L. (1995) On the strength of dependence of a time series generated by a chaotic map, *Journal of Time Series Analysis*, **Vol. 16**, pp. 571–583.

[21] Ibragimov, I.A. and Has'minskii, R.Z. (1981) *Statistical Estimation – Asymptotic Theory*, Springer–Verlag, New York.

[22] Janson, S. (1987) Maximal spacings in several dimensions, *The Annals of Probability*, **Vol. 15**, pp. 274–280.

[23] Lasota, A. and Mackey, M.C. (1994) *Chaos, Fractals, and Noise – Stochastic Aspects of Dynamics*, Springer–Verlag, New York.

[24] Maës, J. (1999) *Statistique Non Paramétrique des Processus Dynamiques Réels en Temps Discret*, PhD Thesis, Université Paris VI, France.

[25] Møller, J. (1994) *Lectures on Random Voronoi Tessellations*, Springer–Verlag, New York.

[26] Tong, H. (1995) A personal overview of non-linear time series analysis from a chaos perspective (with discussion), *Scandinavian Journal of Statistics*, **Vol. 22**, pp. 399–445.

[27] Ziegler, G.M. (1995) *Lectures on Polytopes*, Springer–Verlag, New York.