$L^P(\mu)-$ ESTIMATION OF TANGENT MAPS.

M. Eugenia Mera.   Manuel Morán.

Departamento de Análisis Económico.
Facultad de Económicas.
Campus de Somosaguas.
Universidad Complutense.
28223 Madrid. España.
E-Mail Address: ececo06@sis.ucm.es

ABSTRACT

We analyze under what conditions the best $L^p(\mu)$-linear fittings of the action of a mapping f on small balls give reliable estimates of the tangent map Df. We show that there is an inverse relationship between the conditions on the regularity, in terms of local densities, of the measure μ and the smoothness of the mapping f which are required to ensure the goodness of the estimates. The above results can be applied to the estimation of tangent maps in two empirical settings: from finite samples of a given probability distribution on IR^n and from finite orbits of smooth dynamical systems. As an application of the results of this paper we obtain sufficient conditions on the measure μ to ensure the convergence of Eckmann and Ruelle algorithm for computing the Liapunov exponents of smooth dynamical systems.

Mathematics Subject Classification: $28A78$, $28A33$, $28A25$, $46N10$.

Partially supported by the Dirección General de Enseñanza Superior e Investigación. $PS95-093$.

1 Introduction.

In this paper we provide a rigorous basis to a standard method used in numerical analysis for estimating tangent maps from data sets distributed according to a given probability measure (see Remark 7). This method is based upon the estimates of the tangent map $Df(a)$ of a mapping f at a point a by best L^p-linear estimates of the action of the mapping f on small balls centered at a.

This is a relevant problem for the theory of differentiation with respect to measures in IR^n, and also from the point of view of smooth dynamical systems. In the later case a crucial question is how to determine the Liapunov exponents from an orbit of the system. The Liapunov exponents are the asymptotic exponential rates of convergence or divergence of orbits with nearby initial conditions. They characterize and quantify the chaotic behaviour. The Eckmann and Ruelle algorithm (see [5]) is one of the algorithms most often used for the numerical estimation of the Liapunov exponents. It is based upon the L^p-estimation of the tangent maps along a given orbit of the system. As an application of the results in this article, we are able to solve the open problem of finding under what conditions the Liapunov exponents can be approximated, up to an arbitrary degree of accuracy, using the mentioned algorithm (see [8]). These conditions (see Theorem 2) are quite natural in smooth dynamical systems theory and they cover many interesting cases (see Remark 6 and [1]).

We now formulate the problem solved in this article.

Problem.

Assume that f is a smooth real function on $M \subset \mathrm{IR}^n$. Assume also that μ is a probability Radon measure on M, and let a be a given point in M. Let $B(a,r)$ denote the closed ball, in the Euclidean metric, of radius r centered at a. We define on the set $\mathcal{L}_n(\mathrm{IR}^n, \mathrm{IR}) \equiv \mathcal{L}_n$ of linear forms from IR^n on IR, the functional

$$\mathcal{A}_{p,r}(\beta) = \left[\frac{1}{\mu(B(a,r))} \int_{B(a,r)} |f(y) - f(a) - \beta(y-a)|^p \ d\mu(y) \right]^{1/p} . \quad (1)$$

We adopt the notation $\|\beta\|_2$ for the usual norm of linear maps, i.e. $\|\beta\|_2 = \max\{|\beta v| : |v|_2 = 1, v \in \mathrm{IR}^n\}$ where $|\cdot|_2$ denotes the Euclidean norm.

We ask under what conditions on p, f and μ
A) There exists a unique linear form $\beta_r \in \mathcal{L}_n$ which minimizes $\mathcal{A}_{p,r}$ and
B) β_r tends to the tangent map $Df(a)$ when r tends to zero.

The answer to these questions, in particular to question **B)**, turns out to be non trivial, due to the fact that the measure μ might exhibit a complex local structure, as it is the case when we think of μ as the invariant measure of a dynamical system. Consider, for instance, the case when the measure μ is concentrated on a hyperplane. Then the functional $\mathcal{A}_{p,r}$ does not give any information on how alike the action of f and of linear maps out of the hyperplane are, and the restriction of a linear map to a hyperplane does not determine the linear map. As we will see below, difficulties also arise when the measure μ is concentrated near hyperplanes on arbitrarily small balls, making possible the existence of tangent measures (see section 2 for a definition) of μ at a concentrated on hyperplanes. Notice that this case is relevant for the invariant measure at a dynamics in a smooth submanifold of IR^n.

We show below that the key idea to establish the convergence of β_r to $Df(a)$ when r tends to zero is to obtain a relationship between the usual norm and the $L^p(\mu\,|B(a,r))$-norm of the linear maps $(Df(a)-\beta_r)$. We prove that under suitable conditions, there is a constant $\sigma \in [0,1)$ such that μ-*a.e* $a \in M$, and any $\beta \in \mathcal{L}_n$,

$$\| \beta \|_2 \leq \frac{K}{r^{1+\sigma}} \left[\frac{1}{\mu(B(a,r))} \int_{B(a,r)} |\beta(y-a)|^p \, d\mu(y) \right]^{1/p} \tag{2}$$

holds for small r, where K is a constant dependent on a. Then, applying the last inequality to the linear map $(Df(a)-\beta_r)$, and using the fact that β_r minimizes the functional $\mathcal{A}_{p,r}$, we get

$$\| Df(a) - \beta_r \|_2 \leq \frac{K}{r^{1+\sigma}} \left[\frac{1}{\mu(B(a,r))} \int_{B(a,r)} |(Df(a)-\beta_r)(y-a)|^p \, d\mu(y) \right]^{1/p} \leq$$

$$\frac{K}{r^{1+\sigma}}\left[\frac{1}{\mu(B(a,r))}\int\limits_{B(a,r)}|f(y)-f(a)-\beta_r(y-a)|^p\,d\mu(y)\right]^{1/p}+$$

$$\frac{K}{r^{1+\sigma}}\left[\frac{1}{\mu(B(a,r))}\int\limits_{B(a,r)}|f(y)-f(a)-Df(a)(y-a)|^p\,d\mu(y)\right]^{1/p}\leq$$

$$\frac{2K}{r^{1+\sigma}}\left[\frac{1}{\mu(B(a,r))}\int\limits_{B(a,r)}|f(y)-f(a)-Df(a)(y-a)|^p\,d\mu(y)\right]^{1/p}, \quad (3)$$

and the convergence can be obtained if the degree of differentiability of f is higher than $1+\sigma$.

In Theorem 1 (section two) we show that, under an assumption of strong local regularity of the measure μ, for any sequence $\{r_i\}\downarrow 0$ there is a subsequence $\{r_{i_j}\}$ such that (2) holds for $\sigma=0$. This fact allows us to obtain the required convergence for pointwise differentiable functions. In Theorem 2 (section three), we relax the assumption of local regularity on the measure and find that (2) holds for a positive σ and any $r<r_0$, where r_0 is a constant dependent on a, and we also get the convergence for $f\in C^{1+\varepsilon}$ provided $\varepsilon>\sigma$. In the statements of Theorems 1 and 2 we stress the role played by inequality (2), which we think useful in its own right.

In the remaining part of this section we analyze the problem of existence and uniqueness of the best L^p-linear fittings and prove two lemmas needed later.

Existence and uniqueness of the best L^p-linear estimate.

We now consider a slightly more general problem than the one we will treat later on. We are concerned with the existence and uniqueness of the best L^p-linear fitting of a real function $f\in L^p(\mu)$, where μ is a Radon probability measure on a bounded subset $M\subset \mathrm{IR}^n$ and $p\in(1,\infty)$. We denote by $\|f\|_p$ the norm of f in the metric space $L^p(\mu)$. For $a\in M$ we define the functionals $\mathcal{A}:\mathcal{L}_n\rightarrow \mathrm{IR}$ and $h:\mathcal{L}_n\rightarrow \mathrm{IR}$ by

$$\mathcal{A}(\beta)=\|f-\beta-(f-\beta)(a)\|_p\,, \quad (4)$$

$$h(\beta) = \|\beta - \beta(a)\|_p . \tag{5}$$

If there exists a unique $\beta \in \mathcal{L}_n$ which minimizes $\mathcal{A}$ we say that β is the *best linear estimate in $L^p(\mu)$-norm of f at a.*

Notice that (4) coincides with (1) when the considered measure is $\nu = \frac{1}{\mu(B(a,r))}\mu\,|B(a,r)$ (throughout the text $\mu\,|B(a,r)$ denotes the restriction of the measure μ to the ball $B(a,r)$).

Remark 1 *In this paper we solve problems* ***A)*** *and* ***B)*** *above for real functions defined on $M \subset \mathrm{IR}^n$. Let us see how this also allows us to solve the problem for a vectorial field $f : M \to \mathrm{IR}^m$. In this case we estimate the tangent map of f at a as the linear mapping β which minimizes the functional*

$$\mathcal{A}(\beta) = \left[\int_M \left(|f(y) - f(a) - \beta(y-a)|_p\right)^p \; d\mu(y)\right]^{1/p}$$

defined now on the set $\mathcal{L}_{n,m}$ of linear maps from IR^n into IR^m where $|\cdot|_p$ denotes the p-norm in IR^m. We assume that $|f|_p \in L^p(\mu)$. If f_i and β_i denote the i-th coordinate of f and β respectively, then $(\mathcal{A}(\beta))^p = \sum_{i=1}^m (\mathcal{A}_i(\beta))^p$, where for $1 \le i \le m$,

$$(\mathcal{A}_i(\beta))^p = \int_M |f_i(y) - f_i(a) - \beta_i(y-a)|^p \; d\mu(y).$$

Since the minimum of $\mathcal{A}$ is attained at the linear map that minimizes $\mathcal{A}^p$ and this minimum is clearly attained by a linear mapping β whose i-th coordinate β_i minimizes $(\mathcal{A}_i)^p$, or equivalently $\mathcal{A}_i$, it follows that the problem for vectorial fields can be decomposed into the corresponding problems for their coordinate real functions.

In the next lemma we obtain the existence and uniqueness of the best L^p-linear fitting. We restrict our attention to the set $\mathbf{P}(M)$ of Radon probability measures such that $\mu(H) < 1$ for all hyperplanes H.

Lemma 1 *Let M be a bounded subset of IR^n, $a \in M$, $\mu \in \mathbf{P}(M)$, $p \in (1,\infty)$ and $\mathcal{S} = \{\beta \in \mathcal{L}_n : \|\beta\|_2 = 1\}$. Then*
(i) *There is a $T \in \mathcal{S}$ where the minimum value of h on $\mathcal{S}$ is attained and $h(T) > 0$.*
(ii) $\|\alpha\|_2 \le \frac{h(\alpha)}{h(T)}$*, for all $\alpha \in \mathcal{L}_n$.*
(iii) *If $f \in L^p(\mu)$, there is a unique $\beta \in \mathcal{L}_n$ where the minimum of $\mathcal{A}$ on $\mathcal{L}_n$ is attained.*

Proof. The first part of statement **(i)** follows from the continuity of the functional h on the compact set $\mathcal{S}$. The assumption $\mu \in \mathbf{P}(M)$ guarantees that $h\,(T) > 0$, which together with the fact $h\,(\alpha) \geq \|\alpha\|_2\; h(T)$ for any $\alpha \in \mathcal{L}_n$, give statement **(ii)**. Let $\tau := \inf_{\alpha\in\mathcal{L}_n} \mathcal{A}(\alpha)$ and $R := \frac{\tau+\mathcal{A}(0)}{h(T)}$. Then $\mathcal{A}(\alpha) > \tau$ if $\|\alpha\|_2 > R$, so that the continuous functional $\mathcal{A}$ attains its minimum on the compact set $\{\alpha \in \mathcal{L}_n : \|\ \alpha\ \|_2 \leq R\}$. The uniqueness of such minimum can be obtained from the strict convexity of the normed space $L^p(\mu)$ for $p \in (1, \infty)$ (see [11] and [4]) and from the fact that $\mu \in \mathbf{P}(M)$. ■

In section 2, we will need the following lemma.

Lemma 2 *Let M be a bounded subset of IR^n and let $\{\mu_n\}$ be a sequence of measures in $\mathbf{P}(M)$ which is weakly convergent to the measure μ ($\mu_n \overset{w}{\to} \mu$ for the sequel) with $\mu \in \mathbf{P}(M)$. For $a \in M$ and $p \in (1, \infty)$, let $\{h_n\}$ and h be the functionals defined by (5) for the measures $\{\mu_n\}$ and μ respectively, and let T_n and T be the linear forms of $\mathcal{S}$ where the minima of h_n and h are attained. Then $\lim_{n\to\infty} h_n(T_n) = h(T)$.*

Proof.

The existence of $\{T_n\}$ and T is guaranteed by Lemma 1. Since T_n minimizes h_n on $\mathcal{S}$, we have that $h_n(T_n) \leq h_n(T)$ which, together with the weak convergence, gives $\limsup_{n\to\infty} h_n(T_n) \leq h(T)$. Using the definition of weak convergence, we see that the sequence $\{h_n\}$ is pointwise convergent to h on $\mathcal{S}$. Furthermore, it is easy to prove that $\{h_n\}$ is also an equicontinuous sequence on $\mathcal{S}$ which proves the uniform convergence of $\{h_n\}$ to h on $\mathcal{S}$. Hence, for arbitrarily small ε and sufficiently large n, $h_n(T_n) > h(T_n) - \varepsilon \geq h(T) - \varepsilon$. This shows that $\liminf_{n\to\infty} h_n(T_n) \geq h(T)$. Therefore $\lim_{n\to\infty} h_n(T_n) = h(T)$. ■

2 Tangent measures and the convergence of the best L^p-linear estimates.

Let μ be a Radon probability measure on $M \subset \mathrm{IR}^n$, $a \in \mathrm{IR}^n$, $r > 0$, and let $\nu_r := \frac{1}{\mu(B(a,r))}\mu\,|B(a,r)$. In this section we obtain the convergence of the best linear fittings in $L^p(\nu_r)$-norm of f at a to $Df(a)$ under a strong regularity assumption on the local behaviour of μ (see Theorem 1). The tangent measures of μ at the point a are one of the most useful tools for the study of the local structure of μ at a. They are weak limits of sequences of measures defined as suitable normalizations of measures obtained by blowing up the measure μ by sequences of expansive homotheties centered at a. That is, ν is a *tangent measure* of μ at $a \in \mathrm{IR}^n$ if ν is a non-zero Radon measure on IR^n and if there exist sequences $\{r_i\}$ and $\{c_i\}$ of positive numbers such that $r_i \downarrow 0$ and

$$c_i \varphi_{a,r_i\#}\mu \overset{w}{\to} \nu \text{ as } i \to \infty$$

where φ_{a,r_i} is the homothecy given by $\varphi_{a,r_i}(x) = \frac{(x-a)}{r_i}$ and $\varphi_{a,r_i\#}\mu$ is the measure induced by φ_{a,r_i}, that is $\varphi_{a,r_i\#}\mu(A) = \mu(r_i A + a)$, $A \subset \mathrm{IR}^n$. The set of all such tangent measures is denoted by $Tan(\mu, a)$ (see [7] and [9] for details on tangent measures). In Theorem 1 we use the properties $(P1)$ and $(P2)$ given below (see [7]).

(P1) Let $a \in \mathrm{IR}^n$. If the doubling condition

$$\limsup_{r\downarrow 0} \frac{\mu(B(a,2r))}{\mu(B(a,r))} = K < \infty \tag{6}$$

holds, then every sequence $\{r_i\} \downarrow 0$ contains a subsequence $\{r_{i_j}\}$ such that the measures $\frac{1}{\mu(B(a,r_{i_j}))}\varphi_{\alpha,r_{i_j}\#}\mu$ converge weakly to a tangent measure of μ at a.

Let $0 \le s < \infty$, the *upper* and *lower s-densities* of the measure μ at a point $a \in \mathrm{IR}^n$ are respectively defined by

$$\Theta^{*s}(\mu, a) = \limsup_{r\downarrow 0} \frac{\mu(B(a,r))}{(2r)^s} \text{ and } \Theta^s_*(\mu, a) = \liminf_{r\downarrow 0} \frac{\mu(B(a,r))}{(2r)^s}.$$

(P2) Let s be a positive number, and let A be the set of points $a \in \mathrm{IR}^n$ such that

$$0 < \Theta^s_*(\mu, a) \le \Theta^{*s}(\mu, a) < \infty \tag{7}$$

holds. Then, for μ-*a.e.* $a \in A$ and every $\nu \in Tan(\mu, a)$ there is a positive number c such that

$$tcr^s \le \nu(B(x,r)) \le cr^s, \text{ for } x \in spt(\nu), 0 < r < \infty \tag{8}$$

where $t = \frac{\Theta^s_*(\mu,a)}{\Theta^{*s}(\mu,a)}$, and $spt(\nu)$ denotes the support of the measure ν.

We now recall several definitions used in the proof of Theorem 1.

Given $X \subset \mathrm{IR}^n$ and $\delta > 0$, a collection of balls $\{B_i : i \in \mathrm{IN}\}$ is a δ-covering of the set X if $X \subset \bigcup_{i=1}^{\infty} B_i$ and $d(B_i) \leq \delta$ where $d(\cdot)$ stands for diameter. We define the s-dimensional outer Hausdorff measure $\mathcal{H}_\delta^s$ of a set X by $\mathcal{H}_\delta^s(X) = \inf\{\sum_{i=1}^{\infty}(d(B_i))^s\}$ where the infimum is taken over the set of δ-coverings of X. The s-dimensional Hausdorff measure of X is given by $\mathcal{H}^s(X) = \lim_{\delta\downarrow 0} \mathcal{H}_\delta^s(X)$. The Hausdorff dimension of X is the threshold value

$$\dim(X) = \sup\{t : \mathcal{H}^t(X) > 0\} = \inf\{t : \mathcal{H}^t(X) < +\infty\},$$

and the Hausdorff dimension of a measure μ is defined by $\dim \mu = \inf\{\dim(X) : \mu(X) > 0\}$.

Let $f : M \subset \mathrm{IR}^n \to \mathrm{IR}$ and $a \in M$. We say that f is differentiable at a if there is a linear map $Df(a) \in \mathcal{L}_n$ such that for any $\varepsilon > 0$ there is a $\delta > 0$ satisfying

$$|f(y) - f(a) - Df(a)(y - a)| \leq \varepsilon\, |y - a|_2 \tag{9}$$

for all $y \in M \cap B(a, \delta)$. Notice that this condition holds at every point of the domain of a differentiable function defined on an open set (see also Remark 2).

The next theorem gives sufficient conditions for the convergence to the differential of the best L^p-linear fittings on small balls in terms of the above local densities.

Theorem 1 *Let μ be a Radon probability measure on $M \subset \mathrm{IR}^n$ such that (7) holds for μ-almost every $a \in M$ with $s > n - 1$ and let $p \in (1, \infty)$. Then*
(i) *For μ-a.e. $a \in M$, and any sequence $\{r_i\} \downarrow 0$, there are a subsequence $\{r_{i_j}\}$ and a positive constant K such that for any $\beta \in \mathcal{L}_n$*

$$\| \beta \|_2 \leq \frac{K}{r_{i_j}} \left[\frac{1}{\mu(B(a, r_{i_j}))} \int_{B(a, r_{i_j})} |\beta(y - a)|^p \, d\mu(y) \right]^{1/p} \tag{10}$$

holds for any $j \in \mathrm{IN}$.
(ii) *Let f be a real function defined on M, differentiable μ-almost every*

$a \in M$. *Let* $\nu_r = \frac{1}{\mu(B(a,r))}\mu\lfloor B(a,r)$ *and let* β_r *be the best linear estimate in* $L^p(\nu_r)$*-norm of* f *at* a*. Then there exists a unique* $Df(a)$ *satisfying (9) and*

$$\lim_{r\downarrow 0}\beta_r = Df(a)$$

for μ*-almost every* $a \in M$*.*

Proof.
(i) Let A be the set of points where (7) holds. It is easy to see that (7) implies (6). Then, property $(P1)$ ensures that for every $a \in A$ and for every sequence $\{r_i\} \downarrow 0$, there is a subsequence, which for simplicity we also denote by $\{r_i\}$, such that

$$\frac{1}{\mu(B(a,r_i))}\varphi_{a,r_i\#}\mu \overset{w}{\to} \nu \in Tan(\mu,a). \tag{11}$$

Property $(P2)$ gives a set $B \subset A$, with $\mu(B) = 1$, such that for $a \in B$ the inequalities in (8) hold for the measure ν given in (11). Then we have that $\liminf_{r\downarrow 0}\frac{\log\nu(B(x,r))}{\log r} \geq s$ for $x \in spt(\nu)$, which shows (see [14]) that $\dim\nu > n-1$. Thus $\nu(\partial B(0,1)) = 0$ which, together with (11), easily gives

$$\left(\frac{1}{\mu(B(a,r_i))}\varphi_{a,r_i\#}\mu\right) \mid B(0,1) \overset{w}{\to} \nu \mid B(0,1), \tag{12}$$

and hence $\nu\lfloor B(0,1) \in \mathbf{P}(B(0,1))$. By Lemma 1 there is a $T \in \mathcal{S}$ which minimizes on $\mathcal{S}$ the functional given by

$$h(\alpha) = \left[\int_{B(0,1)} |\alpha y|^p \ d\nu(y)\right]^{1/p},$$

and $h(T) > 0$ holds for such T.

By arguments similar to those given above (see [10]) for ν, it can be shown that (7) implies $\dim\mu \geq s > n-1$. This proves that for $a \in B$, $\nu_{r_i} = \frac{1}{\mu(B(a,r_i))}\mu\lfloor B(a,r_i) \in \mathbf{P}(B(a,r_i))$. Then, by Lemma 1, there is a $\{T_{r_i}\} \in \mathcal{S}$ which minimizes on $\mathcal{S}$ the functional $\{h_i\}$ given by

$$h_i(\alpha) = \left[\frac{1}{\mu(B(a,r_i))}\int_{B(a,r_i)} |\alpha(y-a)|^p \ d\mu(y)\right]^{1/p} = r_i\left[\frac{1}{\mu(B(a,r_i))}\int_{B(0,1)} |\alpha y|^p \ d\varphi_{a,r_i\#}\mu(y)\right]^{1/p}.$$

By Lemma 2, together with (12), we obtain that $\lim_{i\to\infty} \frac{1}{r_i} h_i(T_{r_i}) = h(T)$, so that there is an i_0 such that

$$h_i(T_{r_i}) \geq r_i h(T)/2, \text{ for } i > i_0. \tag{13}$$

By part **(ii)** of Lemma 1, together with (13), we have that for any $\beta \in \mathcal{L}_n$,

$$\|\beta\|_2 \leq \frac{h_i(\beta)}{h_i(T_{r_i})} \leq \frac{2}{h(T) r_i} h_i(\beta)$$

holds for $i > i_0$, which gives (10) for this subsequence taking $K = \frac{2}{h(T)}$.

(ii) Let C be the set of points at which f is differentiable, $a \in B \cap C$, and $\{r_i\} \downarrow 0$. Given **(i)** above, there is a subsequence, which for simplicity we also denote by $\{r_i\}$, such that (10) holds. Since we also have that $f \in L^p(\nu_{r_i})$ for i large enough, Lemma 1 can be applied to obtain the existence and uniqueness of the best linear fitting in $L^p(\nu_{r_i})$-norm of f at a. We denote it by β_{r_i}. Let $Df(a) \in \mathcal{L}_n$ satisfy (9). We now see that $\lim_{i\to\infty} \beta_{r_i} = Df(a)$. Using (10) for $\beta = \beta_{r_i} - Df(a)$, and taking into account equality (3) together with the fact that β_{r_i} is the best linear estimate in $L^p(\nu_{r_i})$-norm of f at a, we obtain

$$\left\|\beta_{r_i} - Df(a)\right\|_2 \leq \frac{K}{r_i} \left[\frac{1}{\mu(B(a,r_i))} \int\limits_{B(a,r_i)} \left|(\beta_{r_i} - Df(a))(y-a)\right|^p d\mu(y) \right]^{1/p} \leq$$

$$\frac{2K}{r_i} \left[\frac{1}{\mu(B(a,r_i))} \int\limits_{B(a,r_i)} |f(y) - f(a) - Df(a)(y-a)|^p \, d\mu(y) \right]^{1/p}. \tag{14}$$

Using (9) we see that for any ε there is an i_1 such that

$$|f(y) - f(a) - Df(a)(y-a)| \leq \frac{\varepsilon \, |y-a|_2}{2K}, \text{ for } y \in M \cap B(a, r_{i_1}). \tag{15}$$

Let i^* be an integer such that $r_i < r_{i_1}$ for all $i > i^*$. Then, using (14) and (15), $\left\|\beta_{r_i} - Df(a)\right\|_2 \leq \varepsilon$ holds for $i > i^*$, which proves that $\lim_{i\to\infty} \beta_{r_i} = Df(a)$.

We have proved that, given a sequence $\{r_i\} \downarrow 0$, there exists a subsequence $\{r_{i_k}\}$ such that the result holds for this subsequence. This proves that $\lim_{r\to 0} \beta_r = Df(a)$ and it also gives the uniqueness of the mapping $Df(a)$ satisfying (9). ■

3 Convergence of the best L^p-linear estimates for smoother functions.

In the previous section we have required a strong degree of local regularity in the measure. This implies that, for μ-almost every point $a \in M$, all tangent measures $\nu \in Tan(\mu, a)$ have a Hausdorff dimension greater than $n-1$, so that they are not concentrated on hyperplanes. The assumptions that we shall impose in this section permit the existence of tangent measures concentrated on hyperplanes. However, they imply a low speed of concentration of μ near any hyperplane on small balls. It allows us to obtain the convergence of the best L^p-linear fittings for smoother functions.

The next lemma states a relationship between the usual and the $L^p(\nu_r)$-norm of any linear mapping β with $\nu_r = \frac{1}{\mu(B(a,r))}\mu \lfloor B(a,r)$. In order to obtain it, we have to impose that there is a fixed proportion of the measure of the ball $B(a,r)$ outside a strip around any hyperplane H through a.

Let H be a hyperplane through the origin and let $0 < \delta < 1$. We denote by H_δ and W_δ the sets given by

$$H_\delta = B(0,1) \cap \bigcup_{x \in H} B(x,\delta) \text{ and } W_\delta^H = B(0,1) \setminus H_\delta.$$

Lemma 3 *Let μ be a Radon probability measure on $M \subset \mathrm{IR}^n$ and $a \in M$ such that there are positive constants r_0, δ and d with the property that for every hyperplane H*

$$\mu(a + r_0 W_\delta^H) > d\mu(B(a,r_0)) \tag{16}$$

holds. Then, for $p \in (1,\infty)$ and all $\beta \in \mathcal{L}_n$, $\beta \neq 0$,

$$\| \beta \|_2 < \frac{1}{d^{1/p} r_0 \delta} \left[\frac{1}{\mu(B(a,r_0))} \int_{B(a,r_0)} |\beta(y-a)|^p \, d\mu(y) \right]^{1/p}. \tag{17}$$

Proof. Let $\beta \in \mathcal{L}_n$ with $\beta \neq 0$ and $H = Ker(\beta)$. Let $\{e_1, ..., e_{n-1}\}$ be a basis of H and take $e_n \in \mathrm{IR}^n$ such that $|e_n|_2 = 1$ and $|\beta e_n| = \|\beta\|_2$. For all $x \in W_\delta^H$, let $(x_1, x_2, ..., x_n)$ be the coordinates of x in the basis $\{e_1, e_2, ..., e_n\}$

of IR^n. Then $|\beta x| = |x_n|\,|\beta e_n| = |x_n|\,\|\beta\|_2 > \delta\,\|\beta\|_2$ holds. For $\nu = \varphi_{a,r_0\#}\mu$ we get

$$\left[\int_{W_\delta^H} (\|\,\beta\,\|_2)^p\,d\nu(x)\right]^{1/p} < \frac{1}{\delta}\left[\int_{W_\delta^H} |\beta x|^p\,d\nu(x)\right]^{1/p},$$

and from this it follows

$$\|\,\beta\,\|_2 < \frac{1}{(\nu(W_\delta^H))^{1/p}\delta}\left[\int_{W_\delta^H} |\beta x|^p\,d\nu(x)\right]^{1/p} \leq$$

$$\frac{1}{(\nu(W_\delta^H))^{1/p}\delta}\left[\int_{B(0,1)} |\beta x|^p\,d\nu(x)\right]^{1/p} =$$

$$\frac{1}{(\mu(a+r_0W_\delta^H))^{1/p}r_0\delta}\left[\int_{B(a,r_0)} |\beta(x-a)|^p\,d\mu(x)\right]^{1/p},$$

and using (16) we see that (17) holds. ■

We now prove that the condition given by (16) holds for μ-almost every $a \in M$ and for any $r < r_0$ under a weak assumption on the logarithmic local densities of the measure μ.

Lemma 4 *Let μ be a Radon probability measure on $M \subset$ IR^n such that*

$$n-1 < \alpha_1 < \liminf_{r\downarrow 0} \frac{\log\mu(B(x,r))}{\log r} \leq \limsup_{r\downarrow 0} \frac{\log\mu(B(x,r))}{\log r} < \alpha_2 \qquad (18)$$

for μ-a.e. $x \in M$ (see Remark 6). Let $\sigma > 0$ and $C_\sigma = \{a \in M$: there are constants r_0, K and d, all of them in the interval $(0,1]$ such that for each hyperplane H and for $r < r_0$, $\frac{\mu(a+rW^H_{Kr^\sigma})}{\mu(B(a,r))} > d$ holds$\}$. Then, for $\sigma > \frac{\alpha_2-\alpha_1}{\alpha_1-n+1}$, $\mu(C_\sigma) = 1$.

Proof. We claim that C_σ is a μ-measurable set. By (18), we know that $\dim\mu > n-1$. From this, for any hyperplane H, it follows that $\frac{\mu(a+rW^H_{Kr^\sigma})}{\mu(B(a,r))}$ is a continuous function of a and r. Let r_0, K and d be fixed constants and let H be a given hyperplane. The set of points $C_{r_0,2K,d,H}$ for which the inequality

$$\frac{\mu(a+rW^H_{2Kr^\sigma})}{\mu(B(a,r))} > d \qquad (19)$$

holds for any $r < r_0$ can be expressed as a countable intersection of μ-measurable sets. Therefore, the set of points $C^*_{r_0,2K,d}$ at which inequality (19) holds for a countable and dense set of hyperplanes is also μ-measurable. This inequality also holds at the points of $C^*_{r_0,2K,d}$ for any hyperplane if we reduce in (19) the value of the constant K. Hence, the set $C_{r_0,K,d}$ where the inequality $\frac{\mu(a+rW^H_{Kr^\sigma})}{\mu(B(a,r))} > d$ holds for any hyperplane H and for every $r < r_0$ is μ-measurable. Lastly, we can express C_σ as a countable union of sets $C_{r_0,K,d}$, and the claim follows.

We now prove that $\mu(C_\sigma) = 1$. The following argument, due to Pertti Mattila, is a simplification of a previous and more involved argument we had given originally as proof.

Suppose that there is a $\sigma > \frac{\alpha_2-\alpha_1}{\alpha_1-n+1}$ such that $\mu(C_\sigma) < 1$. Let E be the set for which (18) holds. Then, for all $x \in E$, there is an r_x such that

$$r^{\alpha_2} \leq \mu(B(x,r)) \leq r^{\alpha_1}, \text{ for } r < r_x. \tag{20}$$

Let $E_j = \{x \in E : r_x > 1/j\}$. Then $E = \bigcup_{j=1}^{\infty} E_j$ and there is a j such that $\mu(E_j \backslash C_\sigma) > 0$. For μ-*a.e.* $x \in E_j \backslash C_\sigma$

$$\lim_{r\to 0} \frac{\mu(E_j \cap B(x,r))}{\mu(B(x,r))} = 1 \tag{21}$$

holds (see [6]). Let $x \in E_j \backslash C_\sigma$ satisfying (21). Then, there is an r_1 such that

$$\mu(E_j \cap B(x,r)) > \frac{\mu(B(x,r))}{2} \text{ for } r < r_1. \tag{22}$$

It is easy to see that for any r, the set $E_j \cap (x + rH_{r^\sigma})$ can be covered by K^* balls with radius $r^{1+\sigma}$, centered at points $x_1, ..., x_{K^*}$ in $E_j \cap (x + rH_{r^\sigma})$, where

$$K^* \leq Qr^{-\sigma(n-1)} \tag{23}$$

and Q is a constant depending only on n. Since $x \notin C_\sigma$, for any constants r_0, K and d in $(0,1]$, there exist a hyperplane H and a radius $r_2 < r_0$ such that $\mu(x + r_2 W^H_{Kr_2^\sigma}) \leq d\mu(B(x, r_2))$ holds. Taking $K = 1$, $d = 1/4$ and $r_0 < \min\{r_1, 1/j, (4Q)^{1/q}\}$ where $q = \alpha_2 - \alpha_1 - \sigma(\alpha_1 - n + 1)$, we get a hyperplane H and an $r_2 < r_0$ satisfying

$$\mu(x + r_2 W^H_{r_2^\sigma}) \leq \mu(B(x, r_2))/4. \tag{24}$$

Using (23) and (20)

$$\mu(E_j \cap (x + r_2 H_{r_2^\sigma})) \leq \sum_{k=1}^{K^*} \mu(B(x_k, r_2^{1+\sigma})) \leq Q r_2^{-\sigma(n-1)+(1+\sigma)\alpha_1} \tag{25}$$

holds, and inequalities (22), (24) and (20) give

$$\mu(E_j \cap (x + r_2 H_{r_2^\sigma})) = \mu(E_j \cap B(x, r_2)) - \mu(E_j \cap (x + r_2 W^H_{r_2^\sigma})) \geq$$

$$\mu(E_j \cap B(x, r_2)) - \mu(x + r_2 W^H_{r_2^\sigma}) > \mu(B(x, r_2))/4 \geq \frac{r_2^{\alpha_2}}{4}. \tag{26}$$

Therefore (25) and (26) implies $r_2^q < 4Q$, which contradicts that $r_2 < \min\{r_1, \frac{1}{j}, (4Q)^{\frac{1}{q}}\}$. ■

We now prove the convergence to the differential of the best L^p-linear fittings on small balls. In order to do this we consider the functions $f : M \subset \mathrm{IR}^n \to \mathrm{IR}$ satisfying the following condition:

D) There are constants ε and L with $0 < \varepsilon < 1$ and $L > 0$, and a set A with $\mu(A) = 1$, such that for all $x \in A$ there is a linear map $Df(x) \in \mathcal{L}_n$ and an r_x satisfying

$$|f(y) - f(x) - Df(x)(y - x)| \leq L\left(|y - x|_2\right)^{1+\varepsilon}, \tag{27}$$

for all $y \in B(x, r_x) \cap M$.

Remark 2 *Condition **D)** is satisfied for all functions f for which the Whitney extension theorem hypotheses hold for a set of full measure (see [12]). For such functions f, there is an extension F of f which is $C^{1+\varepsilon}(\mathrm{IR}^n)$ (i.e. F is $C^1(\mathrm{IR}^n)$ and it has Hölder continuous derivatives with exponent ε). Conversely, if $f \in C^{1+\varepsilon}(U)$, where U is an open set of full measure, then condition **D)** holds.*

Theorem 2 *Let μ be a Radon probability measure on $M \subset \mathrm{IR}^n$ satisfying (18) μ-a.e, let σ be a constant with $\sigma > \frac{\alpha_2 - \alpha_1}{\alpha_1 - n + 1}$ and $p \in (1, \infty)$. Then,*
***(i)** For μ-a.e $a \in M$, there are positive constants r_0 and K such that for all $\beta \in \mathcal{L}_n$*

$$\parallel \beta \parallel_2 \leq \frac{K}{r^{1+\sigma}} \left[\frac{1}{\mu(B(a,r))} \int_{B(a,r)} |\beta(y-a)|^p \, d\mu(y) \right]^{1/p}$$

holds for $r < r_0$.
(ii) *Let f be a real valued function defined on M satisfying condition* **D**) *for a constant $\varepsilon > \sigma$. Let $\nu_r = \frac{1}{\mu(B(a,r))}\mu\,|B(a,r)$, and let β_r be the best linear estimate in $L^p(\nu_r)$-norm of f at a. Then there exists a unique $Df(a)$ satisfying (27) for $x = a$, and*

$$\|\beta_r - Df(a)\|_2 = O(r^{\varepsilon-\sigma}) \ \mu\text{-a.e. } a.$$

Proof.
(i) The proof follows from Lemmas 3 and 4 for any $a \in C_\sigma$ (see Lemma 4 for the definition of this set).
(ii) Let E be the set where (18) holds and let $a \in E \cap A \cap C_\sigma$ (see condition **D)** above for the definition of the set A). Then, $\nu_r \in \mathbf{P}(B(a,r))$, the hypotheses of Lemma 1 are satisfied and the existence and uniqueness of β_r is guaranteed for $r < r_a$. Applying part **(i)** to the linear maps $\beta_r - Df(a)$, where $Df(a)$ is a linear map satisfying (27) for $x = a$, there exist constants r_0 and K such that

$$\|\beta_r - Df(a)\|_2 \ \leq \frac{K}{r^{1+\sigma}} \left[\frac{1}{\mu(B(a,r))} \int\limits_{B(a,r)} |(\beta_r - Df(a))(y-a)|^p \, d\mu(y) \right]^{1/p},$$

for $r < r_0$. This inequality, together with equality (3), and the fact that β_r is the best linear estimate in $L^p(\nu_r)$-norm of f at a, give

$$\|\beta_r - Df(a)\|_2 \ \leq$$

$$\frac{2K}{r^{1+\sigma}} \left[\frac{1}{\mu(B(a,r))} \int\limits_{B(a,r)} |f(y) - f(a) - Df(a)(y-a)|^p \, d\mu(y) \right]^{1/p}, \qquad (28)$$

for $r < r_0$. But f satisfies (27) for $x = a$ which, together with (28), gives

$$\|\beta_r - Df(a)\|_2 \leq 2KLr^{\varepsilon-\sigma}, \text{ for } r < \min\{r_0, r_a\}$$

and since $\varepsilon > \sigma$, we are done. This also proves that $Df(a)$ must be unique. ■

Remark 3 *Notice that (7) implies (18) for any α_1, α_2 with $n - 1 < \alpha_1 <$* $\dim \mu$ *$< \alpha_2$. Then part **(i)** of Theorem 2 follows for any $\sigma > 0$,* and *part **(ii)** holds for any $f \in C^{1+\varepsilon}(U)$ with $\mu(U) = 1$ and ε arbitrarily small.*

Remark 4 *Assumption (18) over the measure* μ *implies that* $\dim\mu \geq \alpha_1$ *and* $Dim\mu \leq \alpha_2$*, where we denote by* $Dim\mu$ *the packing dimension of the measure* μ *(see [13]). Conversely, if* μ *is an* f*-invariant and ergodic measure with* $\dim\mu > n-1$ *and* f *is differentiable, (18) holds for all constants* α_1 *and* α_2 *with* $n-1 < \alpha_1 < \dim\mu$ *and* $\alpha_2 > Dim\mu$*. Theorem 2 is then proved by imposing condition* **D**) *on* f *with* $\varepsilon > \frac{Dim\mu - \dim\mu}{\dim\mu - n+1}$*, thus linking the degree of differentiability of the functions for which the answer of the problem posed in the introduction is positive, with the difference between the Hausdorff and packing dimensions of the measure* μ*. Observe that the constraint* $\varepsilon < 1$ *in condition* ***D)*** *implies that the hypothesis of Theorem 2 does not hold for a measure such that* $Dim\mu - \dim\mu \geq \dim\mu - n + 1$.

Remark 5 *If the dynamics is defined on a smooth d-dimensional submanifold* M *of* $I\!R^n$*, condition (18) does not hold. However, taking a suitable atlas* $(U_i, \Psi_i)_{i\in I\!N}$ *of* M*, and the best linear approximation in* $L^p(\nu_a\,|B(\Psi_a(a), r))$*-norm for* $h := \Psi_{f(a)} \circ f \circ\Psi_a^{-1}$ *at* $\Psi_a(a)$ *(we are denoting by* (U_x, Ψ_x) *a chart of the atlas such that* $x \in U_x$*, and by* $\nu_x := \Psi_{x\#}\mu$*), an extension of Theorem 2 can be obtained for* $f \in C^{1+\varepsilon}$ *with* $\varepsilon > \frac{\alpha_2-\alpha_1}{\alpha_1-d+1} + \frac{\alpha_2-\alpha_1}{p}$ *if we replace the condition* $\alpha_1 > n-1$ *in (18) with* $\alpha_1 > d-1$ *(see [8] for details). This allows us to compute the Liapunov exponents of a dynamics in a smooth manifold, thus solving the issue of the so called spurious exponents.*

Remark 6 *In the case when the upper and lower logarithmic densities given in (18) coincide and are constant* μ*-a.e., the measure* μ *is said to be regular and exact dimensional (see [3]). Eckmann and Ruelle conjectured that any ergodic measure for a smooth dynamical system with hyperbolic behaviour turn out regular and exact dimensional. This conjecture has been proved in [1] for a compactly supported Borel probability measure, with non-zero Liapunov exponents, and invariant under a* $C^{1+\varepsilon}$ *diffeomorphism of a smooth Riemann manifold. In this case, Theorem 2 shows the convergence to the tangent map of the best* L^p*-estimates.*

Remark 7 *The above results can be applied to the estimation of tangent maps from data sets in two empirical settings:*
a) *Finite samples of a given probability distribution on* $I\!R^k$*.*
Let $X_1, X_2, ..., X_n$ *be independent random k-vectors defined on some probability space* $(\Omega, \mathcal{B}, \mathcal{P})$ *and with a common probability distribution* P *on* $I\!R^k$*. Let* f *be a real valued function on* $I\!R^k$ *and assume that* f *and* P *satisfy the hypotheses of Theorems 1 or 2. For* $\omega \in \Omega$*, let* $P_{n,\omega}$ *be the empirical probability*

measure of $X_1(\omega), X_2(\omega), ..., X_n(\omega)$ *given by*

$$P_{n,\omega}(A) = \frac{1}{n}\sum_{j=1}^{n} I_A(X_j(\omega)).$$

For $a \in spt(P)$ *and* $r > 0$, *let*

$$\mu_n = \frac{1}{P_{n,\omega}(B(a,r))} P_{n,\omega}\,\lfloor B(a,r) \quad \text{and } \mu = \frac{1}{P(B(a,r))} P\,\lfloor B(a,r)\ .$$

Then ([2]) $P_{n,\omega} \overset{w}{\to} P$ *for* $\mathcal{P}$*-almost every* ω, *and also* $\mu_n \overset{w}{\to} \mu$ *for* $\mathcal{P}$*-almost every* ω, *which easily gives that* $\lim_{n\to\infty}\beta_{n,r} = \beta_r$ *at* P*-almost every* a, *for* $\mathcal{P}$*-almost every* ω , *where* $\beta_{n,r}$ *is the best linear estimate in* $L^p(\mu_n)$*-norm of* f *at* a, *and* β_r *is the best linear estimate in* $L^p(\mu)$*-norm of* f *at* a. *Since* f *and* P *satisfy the hypotheses of Theorems 1 or 2,* $\lim_{r\to 0}\beta_r = Df(a)$ *at* P*-almost every* a, *and then* $\lim_{r\to 0}\lim_{n\to\infty}\beta_{n,r} = Df(a)$ *for* $\mathcal{P}$*-almost every* ω.

b) *Data sets from finite orbits of smooth dynamical systems.*

Let (M, f, ν) *be a probabilistic dynamical system composed of a state space* $M \subset \mathrm{IR}^k$, *a dynamical law* $f : M \to M$ *such that the state* x_k *of the system at time* k *evolves according to the equation* $x_{k+1} = f(x_k)$, *and an* f*-invariant and ergodic probability measure* ν *on* M. *For* $x \in M$, *let* ν_{n,x_0} *be the orbital measure, given by*

$$\nu_{n,x_0}(A) = \frac{1}{n}\sum_{j=0}^{n-1} I_A(x_j).$$

Using an argument similar to that given above and Remark 1, we see that if ν *and the coordinates of* f *satisfy the hypotheses of Theorems 1 or 2,* $\lim_{r\to 0}\lim_{n\to\infty}\beta_{n,r} = Df(a)$ *holds at* ν*-almost every* a *for* ν*-a.e.* x_0, *where* $\beta_{n,r}$ *is the best linear estimate in* $L^p((\frac{1}{\nu_{n,x_0}(B(a,r))}\nu_{n,x_0})\,\lfloor B(a,r))$*-norm of* f *at* a.

Acknowledgments: We are indebted to Professor Pertti Mattila for a shortening of the previous longer proof of Lemma 4.

References

[1] L. Barreira, Y. Pesin and J. Schmeling, On the Pointwise Dimension of Hyperbolic Measures. A Proof of the Eckmann-Ruelle Conjeture, *Elec-*

tronic Research Announcements of the American Mathematical Society **2** (1996), 69-72.

[2] P. Billingsley, "Convergence of Probability Measures," John Wiley, 1968.

[3] C.D. Cutler, Connecting Ergodicity and Dimension in Dynamical Systems, *Ergod. Theory & Dynam. Syst.* **10,** (1990), 451-462.

[4] P.J. Davis," Interpolation and Approximation,". Blaisdell Publishing Company, 1963.

[5] J.P. Eckmann, S.O. Kamphorst, D. Ruelle and S. Ciliberto, Liapunov Exponents from Time Series, *Physical Review A* **34**, No. 6 (1986), 4971-4979.

[6] H. Federer, "Geometric Measure Theory," Springer-Verlag, 1969.

[7] P. Mattila, "Geometry of Sets and Measures in Euclidean Spaces. Fractals and Rectifiability," Cambridge University Press, 1995.

[8] M.E. Mera and M. Morán, Convergence of the Eckmann and Ruelle Algorithm for the Estimation of Liapunov Exponents, to appear in *Ergod. Theory & Dynam. Syst.*

[9] D. Preiss, "Geometry of Measures in $\mathbb{R}^n$: Distribution, Rectifiability and Densities," *Ann. of Math.* **125,** (1987), 537-643.

[10] C.A. Rogers and S.J. Taylor, Functions Continuous and Singular with respect to a Hausdorff Measure, *Mathematika* **8,** (1961), 1-31.

[11] W. Rudin, " Análisis Real y Complejo," Alhambra, 1985.

[12] E.M. Stein, "Singular Integrals and Differentiability Properties of Functions," Princeton University Press, Princeton, New Jersey, 1970.

[13] C. Tricot, Two definitions of Fractional Dimension, *Math. Proc. Camb. Phil. Soc.* **91,** (1982), 57-74.

[14] L.S. Young, Dimension, Entropy and Lyapunov Exponents, *Ergod. Theory & Dynam. Syst.* **2**, (1982), 109.