



W
28
(9625)

Documento de Trabajo 9625

TANGENT MEASURES
AND Lp ESTIMATION
OF TANGENT MAPS.

. M. Eugenia Mera

. Manuel Morán

Noviembre 1.996

FACULTAD DE CIENCIAS ECONOMICAS Y EMPRESARIALES
UNIVERSIDAD COMPLUTENSE DE MADRID
VICEDECANATO
Campus de Somosaguas. 28223 MADRID. ESPAÑA.

TANGENT MEASURES AND L^p ESTIMATION OF TANGENT MAPS.

Maria Eugenia Mera. Manuel Morán.

November 8, 1996

Abstract

We analyze under what conditions the best L^p -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 underlying measure 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 \mathbb{R}^n and from finite orbits of smooth dynamical systems.

Keywords and Phrases: L^p -estimation, tangent maps, Hausdorff measures, tangent measures.

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

En este artículo se analiza bajo qué condiciones las mejores estimaciones lineales en norma L^p para la acción de una función f sobre bolas de radio pequeño, proporcionan estimaciones fiables de la aplicación tangente Df . Se comprueba que existe una relación inversa entre las condiciones de regularidad, en términos de densidades locales, de la medida subyacente y la suavidad que se requiere a la transformación f para asegurar la bondad de las estimaciones. Los resultados anteriores pueden aplicarse para estimar la aplicación tangente en dos situaciones que se presentan en el trabajo empírico: a partir de muestras finitas de una distribución de probabilidad en \mathbb{R}^n y a partir de órbitas finitas de sistemas dinámicos diferenciables.

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 6). 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 research springs from a study of the convergence of the Eckmann and Ruelle algorithm (see [4]) for the computation of the Liapunov exponents of chaotic dynamical systems. This algorithm 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 prove that Liapunov exponents can be approximated, up to an arbitrary degree of accuracy, using a version of the mentioned algorithm (see [7]). We now formulate the problem.

Problem.

Assume that f is a smooth real function on $M \subset \mathbb{R}^n$. 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 $\mathcal{L}_n(\mathbb{R}^n, \mathbb{R}) \equiv \mathcal{L}_n$ of linear forms from \mathbb{R}^n on \mathbb{R} , the functional

$$\mathcal{A}_{p,r}(\beta) = \left[\frac{1}{\mu(B(a, r))} \int_{B(a, r)} |f(y) - f(a) - \beta(y - a)|^p d\mu(y) \right]^{1/p}. \quad (1.1)$$

We ask under what conditions on p , f and μ

- A) There exists a unique linear form $\beta_r^{(p)} \in \mathcal{L}_n$ which minimizes $\mathcal{A}_{p,r}$ and
- B) $\beta_r^{(p)}$ 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.

We provide an answer to these questions in theorem 2.1 (section two), where we obtain the required convergence for pointwise differentiable functions under

an assumption of strong local regularity for the measure μ and in theorem 3.3 (section three), where we relax the local regularity assumption on the measure by requiring a greater degree of smoothness of f . 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 $M \subset \mathbb{R}^n$ 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 $\mathcal{A}: \mathcal{L}_n \rightarrow \mathbb{R}$ and $h: \mathcal{L}_n \rightarrow \mathbb{R}$ by

$$\mathcal{A}(\beta) = \|f - \beta - (f - \beta)(a)\|_p, \quad (1.2)$$

$$h(\beta) = \|\beta - \beta(a)\|_p. \quad (1.3)$$

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 (1.2) coincides with (1.1) 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 $M \subset \mathbb{R}^n$. Let us see how this also allows us to solve the problem for a vectorial field $f: M \rightarrow \mathbb{R}^m$. In this case we estimate the tangent map of f at a as the linear mapping β which minimizes the functional*

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

defined now on the set $\mathcal{L}_{n,m}$ of linear maps from \mathbb{R}^n into \mathbb{R}^m where $|\cdot|_p$ denotes the p -norm in \mathbb{R}^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 $1 \leq i \leq m$,

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

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

Lemma 1.1. *Let M be a bounded subset of \mathbb{R}^n , $a \in M$, $\mu \in \mathbf{P}(M)$, $p \in (1, \infty)$ and $\mathcal{S} = \{\beta \in \mathcal{L}_n : \|\beta\|_\infty = 1\}$. 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) *$\|\alpha\|_\infty \leq \frac{h(\alpha)}{h(T)}$, 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) is obvious. The assumption $\mu \in \mathbf{P}(M)$ guarantees that $h(T) > 0$, which easily gives statement (ii). Since M is a bounded set and $f \in L^p(\mu)$, $\mathcal{A}(\alpha) < \infty$ for every $\alpha \in \mathcal{L}_n$. From this it follows the existence of a minimum in \mathcal{L}_n for the continuous functional \mathcal{A} . 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], [3]) and from the fact that $\mu \in \mathbf{P}(M)$. We leave the details of the proof to the reader. ■

In section 2, we will need the following lemma.

Lemma 1.2. *Let M be a bounded subset of \mathbb{R}^n and let $\{\mu_n\}$ be a sequence of measures in $\mathbf{P}(M)$ which is weakly convergent to the measure μ ($\mu_n \xrightarrow{w} \mu$ 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 (1.3) 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*

- (i) $\lim_{n \rightarrow \infty} h_n(T_n) = h(T)$.

Let $f \in L^p(\mu)$ be a continuous μ -a.e real function, and let $\{\mathcal{A}_n\}$ and \mathcal{A} be the functionals given by (1.2) for the measures $\{\mu_n\}$ and μ respectively. Then

- (ii) $\lim_{n \rightarrow \infty} \mathcal{A}_n(\beta_n) = \mathcal{A}(\beta)$, and
- (iii) $\lim_{n \rightarrow \infty} \beta_n = \beta$,

where β_n is the best linear estimate in $L^p(\mu_n)$ -norm of f at a , and β is the best linear estimate in $L^p(\mu)$ -norm of f at a .

Proof. We only give the proof of part (i) which will be needed in section 2. Parts (ii) and (iii) are included for completeness. For a proof of these properties see [7]. The existence of $\{T_n\}$ and T is guaranteed by lemma 1.1. 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 $\limsup_{n \rightarrow \infty} h_n(T_n) \leq h(T)$. 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 $\liminf_{n \rightarrow \infty} h_n(T_n) \geq h(T)$. Therefore $\lim_{n \rightarrow \infty} h_n(T_n) = h(T)$. ■

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

In this section we use tangent measures to obtain the convergence of the best L^p -linear fittings to the differential under a strong regularity assumption on the local behaviour of μ (see theorem 2.1 below). We start by recalling several definitions used in the proof of theorem 2.1.

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

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

and the Hausdorff dimension of a measure μ is defined by $\dim \mu = \inf\{\dim(A) : \mu(A) > 0\}$. We denote by $\text{spt}(\mu)$ the support of the measure μ .

Let μ be a Radon measure on \mathbb{R}^n . We say that ν is a *tangent measure* of μ at $a \in \mathbb{R}^n$ if ν is a non-zero Radon measure on \mathbb{R}^n and if there exist sequences $\{r_i\}$ and $\{c_i\}$ of positive numbers such that $r_i \downarrow 0$ and

$$c_i T_{a, r_i} \# \mu \xrightarrow{w} \nu \text{ as } i \rightarrow \infty$$

where the mappings T_{a, r_i} are the homotheties given by $T_{a, r_i}(x) = \frac{(x-a)}{r_i}$ and $T_{a, r_i} \# \mu$ is the measure induced by T_{a, r_i} , that is $T_{a, r_i} \# \mu(A) = \mu(r_i A + a)$, $A \subset \mathbb{R}^n$. The

set of all such tangent measures is denoted by $Tan(\mu, a)$ (see [6] and [9] for details on tangent measures).

Let $0 \leq s < \infty$ and let μ be a Radon measure on \mathbb{R}^n . The *upper* and *lower s -densities* of the measure μ at a point $a \in \mathbb{R}^n$ are defined by

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

Let $f : M \subset \mathbb{R}^n \rightarrow \mathbb{R}$ 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

$$|f(y) - f(a) - Df(a)(y - a)| \leq \varepsilon |y - a|_2 \quad (2.1)$$

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 3).

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 2.1. *Let μ be a Radon probability measure on $M \subset \mathbb{R}^n$ such that*

$$0 < \Theta_*^s(\mu, a) \leq \Theta^{*s}(\mu, a) < \infty \quad (2.2)$$

for μ -almost every $a \in M$ with $s > n-1$. Let $p \in (1, \infty)$ and let f be a real function defined on M , differentiable μ -almost every $a \in M$. Let $\nu_r = \frac{1}{\mu(B(a, r))} \mu|_{B(a, r)}$ 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 (2.1) and

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

for μ -almost every $a \in M$.

Proof. Let A be the set of points where (2.2) holds. It is easy to see that the doubling condition

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

holds for all $a \in A$. Then (see [6]), for every sequence $\{r_i\} \downarrow 0$, there is a subsequence, which for simplicity we also denote by $\{r_i\}$, such that

$$\frac{1}{\mu(B(a, r_i))} T_{a, r_i \# \mu} \xrightarrow{w} \nu \in \text{Tan}(\mu, a). \quad (2.3)$$

Furthermore (see [6]), for μ -almost every $a \in \mathbb{R}^n$ and all $\nu \in \text{Tan}(\mu, a)$, $tcr^s \leq \nu(B(x, r)) \leq cr^s$ holds for $x \in \text{spt}(\nu)$, $0 < r < \infty$, and $t = \frac{\Theta_s^*(\mu, a)}{\Theta^{*s}(\mu, a)}$. Then we have $\liminf_{r \downarrow 0} \frac{\log \nu(B(x, r))}{\log r} \geq s$ for $x \in \text{spt}(\nu)$, which shows ([14]) that $\dim(\text{spt}(\nu)) \geq s > n - 1$ and therefore that $\dim \nu > n - 1$. Thus $\nu(\partial B(0, 1)) = 0$ which, together with (2.3), easily gives

$$\left(\frac{1}{\mu(B(a, r_i))} T_{a, r_i \# \mu} \right) \Big| B(0, 1) \xrightarrow{w} \nu \Big| B(0, 1), \quad (2.4)$$

and hence $\nu \Big| B(0, 1) \in \mathcal{P}(B(0, 1))$. By lemma 1.1 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 $h(T) > 0$.

By arguments similar to those given above (see [10]) for ν , it can be shown that (2.2) implies $\dim \mu \geq s > n - 1$. This proves that for $a \in A$, $\nu_{r_i} = \frac{1}{\mu(B(a, r_i))} \mu \Big| B(a, r_i) \in \mathcal{P}(B(a, r_i))$. Let C be the set of points at which f is differentiable. For $a \in A \cap C$, $f \in L^p(\nu_{r_i})$ for i large enough, and lemma 1.1 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} . Also, for such i , there is a $\{T_{r_i}\} \in \mathcal{S}$ which minimizes on \mathcal{S} the functional $\{h_i\}$ given by

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

By lemma 1.2, together with (2.4), we obtain that $\lim_{i \rightarrow \infty} \frac{1}{r_i} h_i(T_{r_i}) = h(T)$, so that there is an i_0 such that

$$h_i(T_{r_i}) \geq r_i h(T)/2, \text{ for } i > i_0. \quad (2.5)$$

Let $Df(a) \in \mathcal{L}_n$ satisfy (2.1). We now see that $\lim_{i \rightarrow \infty} \beta_{r_i} = Df(a)$. By part (ii) of lemma 1.1 we have

$$\|\beta_{r_i} - Df(a)\|_\infty \leq \frac{1}{h_i(T_{r_i})} \left[\frac{1}{\mu(B(a, r_i))} \int_{B(a, r_i)} |(\beta_{r_i} - Df(a))(y - a)|^p d\mu(y) \right]^{1/p} =$$

$$C_i \left[\int_{B(a, r_i)} |f(y) - f(a) - Df(a)(y - a) - (f(y) - f(a) - \beta_{r_i}(y - a))|^p d\mu(y) \right]^{1/p},$$

if we set $C_i = \left(h_i(T_{r_i}) (\mu(B(a, r_i)))^{1/p} \right)^{-1}$. Now, taking into account that β_{r_i} is the best linear estimate in $L^p(\nu_{r_i})$ -norm of f at a and (2.5), we obtain

$$\|\beta_{r_i} - Df(a)\|_\infty < \frac{4}{r_i h(T)} \left[\frac{1}{\mu(B(a, r_i))} \int_{B(a, r_i)} |f(y) - f(a) - Df(a)(y - a)|^p d\mu(y) \right]^{1/p} \quad (2.6)$$

for $i > i_0$. Using (2.1) we see that for any ε there is an i_1 such that

$$|f(y) - f(a) - Df(a)(y - a)| \leq \frac{\varepsilon h(T) |y - a|_2}{4}, \text{ for } y \in M \cap B(a, r_{i_1}). \quad (2.7)$$

Let i^* be an integer such that $i^* > i_0$ and $r_i < r_{i_1}$ for all $i > i^*$. Then, using (2.6) and (2.7), $\|\beta_{r_i} - Df(a)\|_\infty \leq \varepsilon$ holds for $i > i^*$, which proves that $\lim_{i \rightarrow \infty} \beta_{r_i} = Df(a)$.

We have proved that, given a sequence $\{r_i\} \rightarrow 0$, there exists a subsequence $\{r_{i_k}\}$ such that the result holds for this subsequence. We now prove that any subsequence of $\{r_i\}$ has the same property. Suppose that there is an $\varepsilon > 0$ and a subsequence $\{\beta_{r_{i_j}}\}$ of $\{\beta_{r_i}\}$ such that

$$\|\beta_{r_{i_j}} - Df(a)\|_\infty > \varepsilon \text{ for all } j. \quad (2.8)$$

Applying to the sequence $\{r_{i_j}\}$ the same argument we used above, there exists a subsequence $\{r_{i_{j_k}}\}$ such that $(\frac{1}{\mu(B(a, r_{i_{j_k}})}) T_{a, r_{i_{j_k}}} \# \mu) \mid B(0, 1) \xrightarrow{w} \nu \mid B(0, 1)$ and $\lim_{k \rightarrow \infty} \beta_{r_{i_{j_k}}} = Df(a)$ which contradicts (2.8). Notice that this proves the uniqueness of the mapping $Df(a)$ satisfying (2.1). ■

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

$\nu \in \text{Tan}(\mu, a)$ 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 $\nu_r = \frac{1}{\mu(B(a,r))} \mu|_{B(a,r)}$. 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^H 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 3.1. *Let μ be a Radon probability measure on $M \subset \mathbb{R}^n$ and $a \in M$ such that there are positive constants r_0, δ and d with the property that for every hyperplane H*

$$\mu(a + r_0 W_\delta^H) > d\mu(B(a, r_0)) \quad (3.1)$$

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

$$\|\beta\|_\infty < \frac{1}{d^{1/p} r_0 \delta} \left[\frac{1}{\mu(B(a, r_0))} \int_{B(a, r_0)} |\beta(y - a)|^p d\mu(y) \right]^{1/p}. \quad (3.2)$$

Proof. Let $\beta \in \mathcal{L}_n$ with $\beta \neq 0$ and $H = \text{Ker}(\beta)$. Let $\{e_1, \dots, e_{n-1}\}$ be a basis of H and take $e_n \in \mathbb{R}^n$ such that $|e_n|_2 = 1$ and $|\beta e_n| = \|\beta\|_\infty$. For all $x \in W_\delta^H$, let (x_1, x_2, \dots, x_n) be the coordinates of x in the basis $\{e_1, e_2, \dots, e_n\}$ of \mathbb{R}^n . Then $|\beta x| = |x_n| |\beta e_n| = |x_n| \|\beta\|_\infty > \delta \|\beta\|_\infty$ holds. For $\nu = T_{a, r_0} \# \mu$ we get

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

and from this it follows

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

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

and using (3.1) we see that (3.2) holds. ■

We now prove that the condition given by (3.1) 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 3.2. *Let μ be a Radon probability measure on $M \subset \mathbb{R}^n$ such that*

$$n - 1 < \alpha_1 < \liminf_{r \downarrow 0} \frac{\log \mu(B(x, r))}{\log r} \leq \limsup_{r \downarrow 0} \frac{\log \mu(B(x, r))}{\log r} < \alpha_2 \quad (3.3)$$

for μ -a.e. $x \in M$ (see Remark 5). Let $\sigma > 0$ and $C_\sigma = \{a \in M : \text{there are constants } r_0, K \text{ and } d, \text{ all of them in the interval } (0, 1] \text{ such that for each hyperplane } H \text{ and for } r < r_0, \frac{\mu(a + r W_{K, \sigma}^H)}{\mu(B(a, r))} > d \text{ holds}\}$. 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 (3.3), we know that $\dim \mu > n - 1$. From this, for any hyperplane H , it follows that $\frac{\mu(a + r W_{K, \sigma}^H)}{\mu(B(a, r))}$ is a continuous function of a and r , for μ -almost every a and all $r > 0$. 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

$$\frac{\mu(a + r W_{2K, \sigma}^H)}{\mu(B(a, r))} > d \quad (3.4)$$

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 (3.4) 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 (3.4) the value of the constant K . Hence, the set $C_{r_0, K, d}$ where the inequality $\frac{\mu(a + r W_{K, \sigma}^H)}{\mu(B(a, r))} > d$ 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 (3.3) holds. Then, for all $x \in E$, there is an r_x such that

$$r^{\alpha_2} \leq \mu(B(x, r)) \leq r^{\alpha_1}, \text{ for } r < r_x. \quad (3.5)$$

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 $\mu(E_j \setminus C_\sigma) > 0$. For μ -a.e. $x \in E_j \setminus C_\sigma$

$$\lim_{r \rightarrow 0} \frac{\mu(E_j \cap B(x, r))}{\mu(B(x, r))} = 1 \quad (3.6)$$

holds (see [5]). Let $x \in E_j \setminus C_\sigma$ satisfying (3.6). Then, there is an r_1 such that

$$\mu(E_j \cap B(x, r)) > \frac{\mu(B(x, r))}{2} \text{ for } r < r_1. \quad (3.7)$$

Since $x \notin C_\sigma$, taking $K = 1$, $d = 1/4$ and $r_0 < \min\{r_1, 1/j, (4C)^{1/a}\}$ where C and a are constants defined below, there is a hyperplane H and an $r_2 < r_0$ such that

$$\mu(x + r_2 W_{r_2^H}) \leq \mu(B(x, r_2))/4. \quad (3.8)$$

Now, using (3.7), (3.8) and (3.5)

$$\begin{aligned} \mu(E_j \cap (x + r_2 H_{r_2^\sigma})) &= \mu(E_j \cap B(x, r_2)) - \mu(E_j \cap (x + r_2 W_{r_2^H})) \geq \\ &\mu(E_j \cap B(x, r_2)) - \mu(x + r_2 W_{r_2^H}) > \mu(B(x, r_2))/4 \geq \frac{r_2^{\alpha_2}}{4}. \end{aligned} \quad (3.9)$$

We now use the fact that $E_j \cap (x + r_2 H_{r_2^\sigma})$ can be covered with K^* balls centered at points $x_k \in E_j \cap (x + r_2 H_{r_2^\sigma})$ with radius $r_2^{1+\sigma}$ where

$$K^* = C r_2^{-\sigma(n-1)} \quad (3.10)$$

and C is a constant which depends only on n . Then, using (3.9), (3.10) and (3.5)

$$\frac{r_2^{\alpha_2}}{4} < \mu(E_j \cap (x + r_2 H_{r_2^\sigma})) \leq \sum_{k=1}^{K^*} \mu(B(x_k, r_2^{1+\sigma})) \leq C r_2^{-\sigma(n-1) + (1+\sigma)\alpha_1}$$

so that $r_2^a < 4C$ with $a = \alpha_2 - \alpha_1 - \sigma(\alpha_1 - n + 1) < 0$, which contradicts that $r_2 < \min\{r_1, 1/j, (4C)^{1/a}\}$. ■

Remark 2. Notice that lemmas 3.1 and 3.2 together imply that given $\sigma > \frac{\alpha_2 - \alpha_1}{\alpha_1 - n + 1}$, there are constants r_0 and d such that for all $\beta \in \mathcal{L}_n$ and $r < r_0$

$$\|\beta\|_\infty < \frac{1}{d^{1/p} r^{1+\sigma}} \left[\frac{1}{\mu(B(a,r))} \int_{B(a,r)} |\beta(y-a)|_p d\mu(y) \right]^{1/p}$$

for μ -a.e. $a \in M$.

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 $f : M \subset \mathbb{R}^n \rightarrow \mathbb{R}$ 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

$$|f(y) - f(x) - Df(x)(y-x)| \leq L(|y-x|_2)^{1+\varepsilon}, \text{ for all } y \in B(x, r_x) \cap M. \quad (3.11)$$

Remark 3. 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 $C^{1+\varepsilon}(\mathbb{R}^n)$ (i.e. F is $C^1(\mathbb{R}^n)$ 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 3.3. Let μ be a Radon probability measure on $M \subset \mathbb{R}^n$ satisfying (3.3). Let f be a real valued function defined on M satisfying condition **D)** for a constant $\varepsilon > \frac{\alpha_2 - \alpha_1}{\alpha_1 - n + 1}$. Let $\nu_r = \frac{1}{\mu(B(a,r))} \mu|_{B(a,r)}$, $p \in (1, \infty)$, 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 (3.11) for $x = a$, and

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

where σ is any constant with $\frac{\alpha_2 - \alpha_1}{\alpha_1 - n + 1} < \sigma < \varepsilon$.

Proof. Let E be the set where (3.3) holds and let $a \in E \cap A \cap C_\sigma$ (see lemma 3.2 and condition **D)** above for the definition of the sets E , C_σ and A). Then, $\nu_r \in \mathcal{P}(B(a,r))$, the hypotheses of lemma 1.1 are satisfied and the existence and uniqueness of β_r is guaranteed for $r < r_a$. Applying lemmas 3.1 and 3.2 (see

remark 2) to the linear maps $\beta_r - Df(a)$, where $Df(a)$ is a linear map satisfying (3.11) for $x = a$, there exist constants r_0 and d such that

$$\|\beta_r - Df(a)\|_\infty < \frac{1}{d^{1/p}r^{1+\sigma}} \left[\frac{1}{\mu(B(a,r))} \int_{B(a,r)} |(\beta_r - Df(a))(y-a)|^p d\mu(y) \right]^{1/p}, \quad (3.12)$$

for $r < r_0$. But the right hand expression in the inequality (3.12) is equal to

$$C \left[\int_{B(a,r)} |f(y) - f(a) - Df(a)(y-a) - (\beta_r(y-a))|^p d\mu(y) \right]^{1/p}$$

if we take $C = \frac{(d^{1/p}r^{1+\sigma})^{-1}}{(\mu(B(a,r)))^{1/p}}$. Taking into account that β_r is the best linear estimate in $L^p(\nu_r)$ -norm of f at a , we obtain

$$\|\beta_r - Df(a)\|_\infty < \frac{2(d^{1/p}r^{1+\sigma})^{-1}}{(\mu(B(a,r)))^{1/p}} \left[\int_{B(a,r)} |f(y) - f(a) - Df(a)(y-a)|^p d\mu(y) \right]^{1/p}, \quad (3.13)$$

for $r < r_0$. But f satisfies (3.11) for $x = a$ which, together with (3.13), gives

$$\|\beta_r - Df(a)\|_\infty \leq \frac{2L}{d^{1/p}} r^{\varepsilon-\sigma}, \text{ for } r < \min\{r_0, r_a\}$$

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

Remark 4. Notice that the assumption (3.3) over the measure μ implies that $\dim(\mu) > \alpha_1$ and $Dim(\mu) < \alpha_2$, where we denote by $Dim(\mu)$ the packing dimension of the measure μ (see [13]). Theorem 3.3 is then proved by imposing on f condition **D**) with $\varepsilon > \frac{\alpha_2 - \alpha_1}{\alpha_1 - n + 1}$, 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 .

Remark 5. In the case when the upper and lower logarithmic densities given in (3.3) coincide μ -a.e., the measure μ is said to be regular and exact dimensional (see [2]). Measures which are invariant under smooth dynamical systems with hyperbolic behavior often turn out to be exact dimensional (see [8]). In this case, theorem 3.3 shows the convergence to the tangent map of the best L^p -estimates of any function $f \in C^{1+\varepsilon}(U)$ with $\mu(U) = 1$ and with ε arbitrarily small.

Remark 6. 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 \mathbb{R}^k .*

Let X_1, X_2, \dots, 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 \mathbb{R}^k . Let f be a real valued function on \mathbb{R}^k and assume that f and P satisfy the hypotheses of theorems 2.1 or 3.1. For $\omega \in \Omega$, let $P_{n,\omega}$ be the empirical probability measure of $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ given by

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

For $a \in \text{spt}(P)$ and $r > 0$, let $\mu_n = \frac{1}{P_{n,\omega}(B(a,r))} P_{n,\omega}|_{B(a,r)}$ and $\mu = \frac{1}{P(B(a,r))} P|_{B(a,r)}$. Then ([1]) $P_{n,\omega} \xrightarrow{w} P$ for \mathcal{P} -almost every ω , and also $\mu_n \xrightarrow{w} \mu$ for \mathcal{P} -almost every ω . Using (iii) in lemma 1.2 we get $\lim_{n \rightarrow \infty} \beta_{n,r} = \beta_r$ 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 2.1 or 3.3, $\lim_{r \rightarrow 0} \beta_r = Df(a)$ at P -almost every a , and then $\lim_{r \rightarrow 0} \lim_{n \rightarrow \infty} \beta_{n,r} = Df(a)$ 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 $M \subset \mathbb{R}^k$, a dynamical law $f : M \rightarrow M$ 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 2.1 or 3.3, $\lim_{r \rightarrow 0} \lim_{n \rightarrow \infty} \beta_{n,r} = Df(a)$ holds at ν -almost every a for ν -a.e. x_0 , where $\beta_{n,r}$ is the best linear estimate in $L^p(\left(\frac{1}{\nu_{n,x_0}(B(a,r))} \nu_{n,x_0}\right)|_{B(a,r)})$ -norm of f at a .

Acknowledgments: We are indebted to Professor Pertti Mattila for a shortening of the previous longer proof of lemma 3.2.

References

- [1] **P. Billingsley**, *Convergence of Probability Measures*, John Wiley (1968).
- [2] **C.D. Cutler**, Connecting Ergodicity and Dimension in Dynamical Systems, *Ergodic Th. Dynam. Syst.*, **10**, (1990), 451-462.
- [3] **P.J. Davis**, *Interpolation and Approximation*. Blaisdell Publishing Company (1963).
- [4] **J.P. Eckmann, S.O. Kamphorst, D. Ruelle and S. Ciliberto**, Liapunov Exponents from Time Series, *Physical Review, A*, **34,6**, (1989), 4971-4979.
- [5] **H. Federer**, *Geometric Measure Theory*, Springer-Verlag, Berlín-Heidelberg-New York, 1969.
- [6] **P. Mattila**, *Geometry of Sets and Measures in Euclidean Spaces. Fractals and Rectifiability*, Cambridge University Press (1995).
- [7] **M.E.Mera and M. Morán**, Convergence of the Eckmann and Ruelle Algorithm for the Estimation of Liapunov Exponents, manuscript in preparation.
- [8] **Y. Pesin and H. Weiss**, A Multifractal Analysis of Equilibrium Measures for Conformal Expanding Maps and Markov Moran Geometric Constructions, preprint.
- [9] **D. Preiss**, Geometry of Measures in \mathbb{R}^n : Distribution, Rectifiability and Densities, *Annals of Mathematics*, **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] **Stein E.M.** *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. Theor. & Dyn. Sys.*, **2** (1982), 109.