Optimal maps for the L^2-Wasserstein distance*

J.A. Cuesta-Albertos[1], C. Matrán[2] and A. Tuero-Díaz[1]

January 22, 1996

[1] Departamento de Matemáticas, Estadística y Computación. Universidad de Cantabria. Spain.
[2] Departamento de Estadística e Investigación Operativa. Universidad de Valladolid. Spain.

Abstract

In this paper we analize properties of the maps that give optimal transportation plans for the $L\text{-}^2$-Wasserstein distance. The main topics treated include existence, measurability, continuity, uniqueness and convergence with respect to the weak convergence of probability measures. Moreover, we obtain the continuity of the optimal maps with respect ot the weak convergence what includes, in particular, the consistency of the empirical optimal maps.

Key words and phrases: Wasserstein distance, Monge-Kantorovich transportation problem, optimal transportation plans, existence, uniqueness, consistency, Skorohod representation theorem, dependence structure, maximal monotone operators, continuity.
A.M.S. 1980 subject classification: 60E05, 60B10.
Running title: Optimal maps. Wasserstein distance.

1 Introduction

The Monge-Kantorovich mass-transportation problem (MTP) consists in to minimize the total cost of transportation of a mass placed initially in a location given by a probability measure P to a final location given by another probability measure Q.

*Research partially supported by the Spanish DGICYT under grants PB91-0306-C02-00, 01 and 02

To be more precise, let (B, β) be a measurable space and P and Q be two probability measures on β. Let $c : B \times B \to R^+$ be a function which through the value $c(x, y)$ measures the cost of transportation of a unit mass from x to y. We will denote

$$C(P, Q) = \inf \left\{ \int c(x, y)\mu(dx, dy), \mu \in M(P, Q) \right\},$$

where $M(P, Q)$ is the set of probability measures on β^2 with marginal distributions P and Q.

Under reasonable conditions on c and B the functional $C(P, Q)$ induces a metric on the family of all probability measures defined on β. This idea has been used from several viewpoints. A good reference which exhaustively treats properties and applications of these functionals and metrics is [14].

Here we are interested in the case in which B is the euclidean space $\Re^p$ and $c(x, y) = \|x - y\|^2$ where $\| \ \|$ denotes for the usual norm in $\Re^p$. However, as do many authors on this subject, given P and Q two probability measures on $\Re^p$, with $\int \|x\|^2 dP$ and $\int \|x\|^2 dQ$ finite, we use the terminology L^2-Wasserstein distance between P and Q, $W(P, Q)$, to denote the value defined by means of

$$W^2(P, Q) = \inf \left\{ \int \|x - y\|^2 \mu(dx, dy), \quad \mu \in M(P, Q) \right\}. \tag{1}$$

The infimum in (1) is attained, so that to find $W(P, Q)$ it is necessary and sufficient to obtain, in some (rich enough) probability space (Ω, σ, ν), a pair (X, Y) of random vectors whith distributions laws $\mathcal{L}(X) = P$ and $\mathcal{L}(Y) = Q$ and verifying

$$\int \|X - Y\|^2 d\nu = \inf \left\{ \int \|S - T\|^2 d\nu, \quad \mathcal{L}(S) = P, \quad \mathcal{L}(T) = Q \right\} \quad (= W^2(P, Q)).$$

Such a pair (X, Y) is called an (L^2-)optimal transportation plan (in short, o.t.p.) for (P, Q). ((L^2-)Optimal coupling for (P, Q) is an alternative, sometimes used, terminology).

In this paper we are mainly concerned with the basic theoretical properties of optimal maps giving L^2-Wasserstein distances: In [6] it was proved that, under continuity assumptions on the probability P, the L^2-optimal transportation plan (X, Y) can be represented as

$(X, \psi(X))$ for some suitable "optimal map" ψ. Moreover, as noted in [8], an interesting consequence of the characterization of optimal transportation plans in [16] is the fact that the optimality of a map ψ is essentially independent of the distribution of X.

Results of this kind were pioneered by Knott and Smith in [11, 17] by considering the opposite point of view of handling mappings ψ, possibly multivalued, such that $(X, \psi(X))$ is an o.t.p. for some pair of probability measures.

However, in spite of the large number of papers on this topic which have appeared in the last years, only the one-dimensional situation has at present a completely satisfactory treatment. In fact, there exist very simple questions, arising from the consideration of well-known relevant properties of the optimal maps in $\Re$, which have been studied only partially. Ironically such a type of property very often gives fruitful results. For example the monotonicity of optimal maps established in [6] has been the main tool to prove there a Central Limit Theorem in Hilbert spaces. It was also the key tool in [18] to prove that "representations" based on the L^2-Wasserstein distance give Skorohod a.s. representations in $\Re^p$ for weak convergence of probability measures.

In this paper we present new results arising from the precedent considerations. Three different problems are considered in as many Sections.

In Section 3 we study the uniqueness of the optimal map. This problem is implicit in many papers on the subject and partial answers have been given in [13], when both probabilities are Gaussian, or in [9], where the cost functions and the spaces under consideration are more general, but a finite support of the probability Q is assumed. Here we present an elementary proof arising from the consideration of random convex linear combinations (notice that the proof covers the problem in separable Hilbert spaces). In this framework we also show the "strict convexity" of the L^2-Wasserstein distance, a fact that allows us to consider Gelbrich's bound (see [10] and [7]) from new viewpoints, by considering probabilities with the same structure of dependence in some basis (see [8] and [7]). Our analysis will show that the existence of two of such bases implies a kind of linear relationship between the probabilities.

Section 4 is devoted to studying the topological nature of optimal maps. The Appendix in [18] showed the measurability of increasing functions in the σ-algebra of Lebesgue sets. With somewhat related techniques we will even prove the a.e. continuity of increasing functions (and hence that of optimal maps).

In Section 5 we will consider the problem of a.s. convergence of representations based on the L^2-Wasserstein distance. In particular we obtain that if $(X, H_n(X))$ are optimal transportation plans between P and Q_n, $n = 1, 2, ...$, and $(Q_n)_n$ converges weakly to P, then $(H_n(X))_n$ converges a.s. to X. We also obtain, as a corollary, the consistency of the empirical optimal maps.

These results justify the use of approximations to an optimal map by means of optimal maps for approximate distributions. The interest of this fact comes from a general expression for the o.t.p. is not known. However in [1] a algorithm to find an o.t.p. between P and Q is provided under the assumption that P is absolutely continuous with respect to the Lebesgue measure and the support of Q is finite. So the results in this section with a Montecarlo approximation to Q yield optimal maps in an approximate way for general Q.

On the other hand, in [18] it was proved that, under continuity assumptions on P, o.t.p.'s $(X, H_n(X))$ between P and Q_n, $n = 1, 2, ...$, provide an a.s. Skorohod representation: If $(Q_n)_n$ converges weakly to P, then $(H_n(X))_n$ converges a.s. to X. Our Theorem 5.6 provides a natural non-trivial generalization of this result.

Finally we want to remark that in the next section we will give some notation and preliminary results, but we will also provide technical results which complete and justify some non-obvious steps of the proofs in [6].

2 Some preliminary results

In this section we summarize, for the sake of completeness, some notation, definitions and results which can be found mainly in [6, 8, 18]. We also fill a gap which has not been noticed in the literature till now.

The set of probability measures, P, on the euclidean space $\Re^p$ verifying $\int \|x\|^2 dP < \infty$ will be denoted by $\mathcal{P}_2$.

We assume through the paper that all the random variables under consideration are $\Re^p$-valued and defined on the same, rich enough, space (Ω, σ, ν). The symbol λ_p will denote the Lebesgue measure in $\Re^p$, and absolute continuity of a measure μ will be assumed to be w.r.t. λ_p and denoted by $\mu \ll \lambda_p$. Given a vector $v \in \Re^p$, $< v >$ denotes the linear space spanned by v, and the superscript $^\perp$ denotes the orthogonal subspace.

It is showed in [12] that in the real case, optimal maps coincide with increasing arrangements. Even though increasing maps also play an important role in the multivariate case, in $\Re^p$ the appropiate concept of increasing map to be handled must be specified. It turns out to be that of monotone operator in the sense of Zarantonello (see, e.g., [3]): Let us denote by $< \cdot, \cdot >$ the inner product in $\Re^p$. A mapping $H : D \subset \Re^p \to \Re^p$ is said to be Z-increasing when for every x, x' in D it holds that $< H(x) - H(x'), x - x' > \geq 0$. While that condition is not sufficient to characterize optimal maps, in [6] it was proved the necessity also in $\Re^p$. A precise formulation of this fact is provided by the following statement. It has been proved in [6] (Theorem 2.3 and Corollary 2.4) in separable Hilbert spaces.

Theorem 2.1 *Assume that for the probability P there exists an orthogonal basis, $\{v_n\}_{n=1}^p$, such that, for each n and almost everywhere ω in $< v_n >^\perp$, the conditional distribution function on $< v_n >$ given ω is continuous. Let (X, Y) be an o.t.p. for (P,Q). Then we have:*

a) $P\{x : \#S_x = 1\} = 1$, where $S_x = S_x(Y)$ is the support of a regular conditional distribution of Y given $X = x$ and $\#$ denotes the cardinal of a set.

b) $\nu \otimes \nu\{(\omega, \omega') :< X(\omega) - X(\omega'), Y(\omega) - Y(\omega') >< 0\} = 0$.

From statement a) it is evident that if (X, Y) is an o.t.p. for (P,Q) and we consider the ν-a.e. defined choice function $x \to H(x)$, where $H(x)$ is the only element in S_x, then we obtain that $Y = H(X)$, ν-a.s. and that, from b),

$$P \otimes P\{(x, x') :< H(x) - H(x'), x - x' >< 0\} = 0. \qquad (2)$$

Two open questions remain about H regarding measurability and the existence of a P-probability one set, D, in which H behaves as a Z-increasing map.

With respect to the first question, let $D = \{x : \#S_x = 1\}$. Measurability of D when Y is a real random variable can be easily proved by handling a conditional distribution function $F(y/X = x)$ of Y given $X = x$ and noting that

$$\{x : \#S_x = 1\} = \bigcap_{k \in N} \{x : \sup_{\substack{y,y' \in Q \\ 0<y-y'<\frac{1}{k}}} [F(y/X = x) - F(y'/X = x)] = 1\}.$$

The argument does not extend to random variables with values in abstract spaces, but this difficulty can be circumvented in the context of Hilbert spaces treated in [6] by considering the class of real functions $\{f_n\}_n$ defined by $f_n(y) =< y, e_n >$ for some complete orthonormal system $\{e_n\}_n$. If $P(\cdot/X = x)$ is a regular conditional distribution of Y given $X = x$, then $F_n(z/X = x) \stackrel{\text{def}}{=} P(f_n(Y) \leq z/X = x)$ are conditional distribution functions of $f_n(Y)$. Moreover the corresponding supports verify

$$\{x : \#S_x(Y) = 1\} = \bigcap_n \{x : \#S_x(f_n(Y)) = 1\}$$

and, if we denote this set by A, we have that, if $B \in \beta$, then

$$\{x \in A : H(x) \in B\} = \{x \in A : P(B/X = x) = 1\}.$$

Therefore we have proved that there exists a P-probability one measurable set, A, and a map H whose restriction to A is measurable and such that $(X, H(X))$ is an o.t.p. (Note that both measurability properties hold even if the σ-algebra under consideration is not complete).

The conditions assumed in the preceding results are trivially satisfied for probability measures absolutely continuous with respect to λ_p. Therefore, for simplicity, we will generally assume absolute continuity for the probability measure P. However more precise results involving the conditions above will be considered. The relevant conditions for the arguments to be valid will be given.

With respect to the second question, it is not trivial that (2) implies the existence of a P-probability one set on which H is Z-increasing. This is in spite of the fact that this has been taken sometimes as guaranteed [6, 18]. The gap can be circumvented making use of the characterization of optimal maps in [16], which we adapt here for Hilbert spaces:

Let P, Q be probability measures on a real Hilbert space E. In order that the pair (X, Y) of E-valued random variables be an o.c for (P, Q) it is necessary and sufficient that there exists a lower semicontinuous proper convex function f on E such that $Y \in \partial f(X)$ ν-a.s. *(here ∂f denotes the subdifferential of the function f, see [3, 15] for a detailed study).*

An equivalent condition to $Y \in \partial f(X)$ ν-a.s. is $< X, Y >= f(X) + f^*(Y)$ ν-a.s. (where f^* denotes the conjugate function of f), so taking into account the measurability of the involved functions (f^* becomes also a lower-semicontinuous proper convex function), and the previous result, we have that $\mathcal{U} = \{x \in A : H(x) \in \partial f(x)\}$ is a P-probability one measurable set. Finally, since the subdifferentials of lower-semicontinuous proper convex functions are the "maximal cyclically monotone", hence Z-increasing, operators (see theorem 2.3 in [3]), it becomes proved (even in Hilbert spaces) the following theorem which summarize the conclusions in this section.

Theorem 2.2 *Let (X,Y) be an o.t.p. for (P,Q) and assume that $P \ll \lambda_p$. Then there exists a P-probability one set D and a map $H : \Re^p \to \Re^p$ such that*

a) $Y = H(X)$, ν-a.s.

b) H is Borel-measurable.

c) H is cyclically monotone (hence Z-increasing) on D.

However the aroused question: *Does the condition (2) imply the existence of a P-probability one set on which H is Z-increasing?* has also an affirmative answer through the following statement, whose proof will be given in the appendix (Theorem 6.4).

Theorem 2.3 *Let P be a probability measure absolutely continuous with respect to λ_p. Let T be a set such that P(T)=1 and suppose that $H : T \subset \Re^p \to \Re^p$ is a map which verifies that*

$$P \otimes P\{(x, x') :< x - x', H(x) - H(x') >< 0\} = 0.$$

Then there exists a P-probability one set, D, such that H is Z-increasing on D.

3 Geometric properties

Let P be an absolutely continuous probability measure in $\mathcal{P}_2$, and let X be a $\Re^p$-valued r.v. with law $\mathcal{L}(\mathrm{X}) = \mathrm{P}$. In this section we are primarily concerned with the problem of uniqueness of the L_2-Wasserstein-based representation in the form $\psi(X)$ of a probability Q on $\Re^p$ through an optimal map ψ. However our proof involves considerations which are better presented in a more general set-up.

Consider Q_1, Q_2 in $\mathcal{P}_2$ and let (X, Y_1) and (X, Y_2) be two o.t.p.'s for (P, Q_1) and (P, Q_2), respectively. Moreover assume that $B(= B(\lambda))$ is a Bernoulli random variable with parameter λ, $0 < \lambda < 1$, independent of the random variables above, and define the random convex linear combination $Z = BY_1 + (1 - B)Y_2$.

It is obvious that $\mathcal{L}(Z) = \lambda Q_1 + (1 - \lambda)Q_2$ and also that, if $P_1(\cdot/\mathrm{X{=}x})$ and $P_2(\cdot/\mathrm{X{=}x})$ are regular conditional distributions of Y_1 and Y_2 given X=x, then

$$P_Z(\cdot/\mathrm{X{=}x}) \overset{\mathrm{def}}{=} \lambda P_1(\cdot/\mathrm{X{=}x}) + (1 - \lambda)P_2(\cdot/\mathrm{X{=}x})$$

is a regular conditional distribution of Z given X=x.

Hence $S_x(Z) = S_x(Y_1) \cup S_x(Y_2)$ for every x, so that $\#S_x(Z) = 1$ is possible only for those x for which $S_x(Y_1) = S_x(Y_2)$ and $\#S_x(Y_1) = 1$. Therefore (recall property a) in Theorem 2.1) (X,Z) can be an o.t.p. for $(P, \lambda Q_1 + (1 - \lambda)Q_2)$ only if $Z = Y_1 = Y_2$ a.s. and, on the contrary, we have that

$$W^2(P, \lambda Q_1 + (1 - \lambda)Q_2) < \int \|X - Z\|^2 d\nu = \int_{\{B=1\}} \|X - Y_1\|^2 d\nu + \int_{\{B=0\}} \|X - Y_2\|^2 d\nu = \lambda W^2(P, Q_1) + (1 - \lambda)W^2(P, Q_2). \quad (3)$$

This shows the strict convexity of the mapping $W^2(P, \cdot)$ and the uniqueness of the representation based on X (it suffices to consider $Q_1 = Q_2$ but $\nu(Y_1 \neq Y_2) > 0$ in the reasoning to get an absurd) which are collected in the following theorem.

Theorem 3.1 *Consider P and Q with or without affixes in $\mathcal{P}_2$ and assume that $P \ll \lambda_p$. Then we have:*

a) If $Q_1 \neq Q_2$ then, for every λ in $(0,1)$,

$W^2(P, \lambda Q_1 + (1-\lambda)Q_2) < \lambda W^2(P, Q_1) + (1-\lambda) W^2(P, Q_2)$.

b) If (X, Y_1) and (X, Y_2) are o.t.p.'s for (P,Q), then $Y_1 = Y_2$ a.s.

Remark 3.1.1

We would like to note that the argument in the proof is purely probabilistic and works as long as item a) in Theorem 2.1 holds. In particular, this extends to situations somewhat more general than absolutely continuous probabilities and to cover Hilbert spaces (see [6] for details). On the other hand, the following trivial examples show that it is necessary to impose on P some regularity condition.

If P is a probability degenerated on 0 and Q_1 and Q_2 are two probability measures, then inequality in (3) is trivially an equality and we have a counterexample to strict convexity.

To give a counterexample to the uniqueness of the representation, note that for the probability measures P and Q defined on $\Re^2$ by P[(1,-1)] = P[(-1,1)] = Q[(1,1)] = Q[(-1,-1)] = 1/2, more than one o.t.p. based on the same random variable X is obviously possible.

For those probabilities P for which strict convexity of $W^2(P, \cdot)$ holds, it is possible to define their "metric projection" onto every closed convex set, C, of probabilities in $\mathcal{P}_2$.

As an example, let Σ be a positive semidefinite matrix and let us consider the set, $\mathcal{P}_\Sigma$, of probabilities in $\mathcal{P}_2$ with covariance operator Σ and zero mean and consider P in $\mathcal{P}_2$ with nondegenerate covariance operator Σ_0 and zero mean. Gelbrich's bound (see [10] and [7]) gives a common lower bound for the L_2-Wasserstein distance between P and every Q in $\mathcal{P}_\Sigma$:

$$W^2(P,Q) \geq \text{ trace } \left(\Sigma_0 + \Sigma - 2 \left(\Sigma_0^{1/2} \Sigma \Sigma_0^{1/2} \right)^{1/2} \right).$$

Moreover the value of the distance between P and $P \circ T^{-1}$ is well known if T is a non-negative definite linear transformation, so by considering $T = \left(\Sigma_0^{1/2}\right)^{-}\left(\Sigma_0^{1/2}\Sigma\Sigma_0^{1/2}\right)^{1/2}\left(\Sigma_0^{1/2}\right)^{-}$ (the superscript $^{-}$ denotes inverse) we have that $P^* \stackrel{\text{def}}{=} P \circ T^{-1}$ belongs to $\mathcal{P}_\Sigma$ and

$$W^2(P, P^*) = \text{ trace } \left(\Sigma_0 + \Sigma - 2\left(\Sigma_0^{1/2}\Sigma\Sigma_0^{1/2}\right)^{1/2}\right).$$

Hence

$$W^2(P, P^*) = W^2(P, \mathcal{P}_\Sigma) = \text{ inf }_{Q\in\mathcal{P}_\Sigma} W^2(P, Q).$$

Since $\mathcal{P}_\Sigma$ is obviously convex and closed with respect to the L_2-Wasserstein distance, when $W^2(P, \cdot)$ is strictly convex (in particular e.g. if $P \ll \lambda_p$), there exists a unique probability P_0 in $\mathcal{P}_\Sigma$ such that $W^2(P, P_0) = W^2(P, \mathcal{P}_\Sigma)$ which gives even a more relevant role to the linear transformation T, because then necessarily $P_0 = P^* = P \circ T^{-1}$. Moreover, under hypotheses ensuring uniqueness of the representation (if $P_0 \ll \lambda_p$ e.g.), if $\mathcal{L}(X) = P$, then the L_2-nearest random vector to X with covariance operator Σ and zero mean is T(X).

In connection with these considerations, it is interesting to recall that the only linear operators which are optimal maps are those non-negative definite. Then, when T is a linear operator, the fact that (X,T(X)) is an o.t.p. for (P,Q) is equivalent to the existence of an orthonormal basis $\{e_1, \ldots, e_p\}$ in which T can be written as

$$T(x) = T\left(\sum_{i=1}^{p} x_i e_i\right) = \sum_{i=1}^{p}(\lambda_i x_i)e_i$$

for some non-negative $\lambda_1, \ldots, \lambda_p$. In other words, optimal maps based on linear transformations are homotecies on each component for some basis and, in particular, this implies that P and Q have the same structure of dependence (see [8]) in such basis.

The general formulation of the optimal maps related to probabilities P and Q with the same structure of dependence in the basis $\{e_1, \ldots, e_p\}$ is

$$T(x) = T\left(\sum_{i=1}^{p} x_i e_i\right) = \sum_{i=1}^{p} f_i(x_i)e_i,$$

for some non-decreasing real functions $\{f_1, \dots, f_p\}$. Therefore let us say that such an optimal map "relates structures of dependence". We finalize this section by showing that such a representation is essentially unique and that the optimal maps are intrinsically related to special orthogonal decompositions of the space.

First note the trivial fact that T is an optimal map if and only if T+a is an optimal map for every a in $\Re^p$. So let us consider the reduced case in which $f_i(0) = 0$, $i = 1, ..., p$. To be more precise, let us consider that $H : \Re^p \to \Re^p$ is a map which can be expresed as

$$H\left(\sum_{i=1}^{p} x_i e_i\right) = \sum_{i=1}^{p} f_i(x_i) e_i,$$

and as

$$H\left(\sum_{i=1}^{p} \hat{x}_i \hat{e}_i\right) = \sum_{i=1}^{p} g_i(\hat{x}_i) \hat{e}_i,$$

where $\{e_1, ..., e_p\}$ and $\{\hat{e}_1, ..., \hat{e}_p\}$ are orthonormal bases and $f_1, ..., f_p$ are real, non-decreasing functions such that $f_i(0) = 0$. Also, to avoid trivialities, let us assume that no pair of vectors $e_i, \hat{e}_j$ is linearly related (i.e. $e_i \neq \pm \hat{e}_j$, for all i, j $(i \neq j)$). Then we have the following proposition (whose proof will be deferred to the Appendix).

Proposition 3.2 *Under the hypotheses above, we have that H is a linear map which can be expressed as $H\left(\sum_{i=1}^{p} x_i e_i\right) = \sum_{i=1}^{p} (\beta_i x_i) e_i$, where the eigenvalues β_i are positive with multiplicities $m_i \geq 2$, $i = 1, ..., p$.*

The conclusion of the proposition can be summarized by noting that every optimal map which relates structures of dependence uniquely determines an orthogonal decomposition of the space: Let us say that a vector v in $\Re^p$ constitutes an invariant direction for H if $H(< v >) \subset < v >$ and let H be an optimal map which relates structures of dependence such that $H(0) = 0$, then the following theorem is a simple consequence of the preceding proposition.

Theorem 3.3 *There exists a unique decomposition of $\Re^p$ as a direct sum of orthogonal subspaces $\Re^p = V_1 \oplus ... \oplus V_k$ such that:*

i) Every vector in a subspace provides an invariant direction for H.

ii) If dim(V_i) > 1 for some i, then there exists a $\beta \geq 0$ such that $H(v) = \beta v$ for every v in V_i.

iii) If $i \neq j$ and $H/_{V_i}$ and $H/_{V_j}$ are linear, then the associated eigenvalues are different.

iv) If $v = v_1 + \cdots + v_k$, with $v_i \in V_i$, then $H(v) = H(v_1) + \cdots + H(v_k)$

To end this section, we want to note that the converse of $i)$ does not hold in general. For instance, let us suppose that $p = 2$ and define

$$H(x_1e_1 + x_2e_2) = x_1^2\mathrm{sign}(\mathrm{x}_1)\mathrm{e}_1 + \frac{\mathrm{x}_2^2}{2}\mathrm{sign}(\mathrm{x}_2)\mathrm{e}_2$$

The decomposition in the theorem is obviously given by $< e_1 > \oplus < e_2 >$. But the vector $\hat{e}_1 = e_1 + 2e_2$ provides an invariant direction for H. Of course, if $\hat{e}_2$ is an orthogonal vector to $\hat{e}_1$, then $\hat{e}_2$ does not yield an invariant direction.

4 Continuity of Z-monotone maps

In this section we study the a.e. continuity of Z-increasing maps. To this we need to state some technical lemmas. For these results we use the following notation.

The symbol $H : D \subset \Re^p \to \Re^p$ denotes a Z-increasing map defined on a measurable set D; $\{e_1, ..., e_p\}$ is a(n fixed) orthonormal basis in $\Re^p$. Our proof will be based on the study of the maps:

$$\mathrm{H}^j(x) =< H(x), e_j > \text{ and } \vec{H}^j(x) = \mathrm{H}^j(x)e_j, \; j = 1, ..., p.$$

Given $x \in \Re^p$, we denote $x^j =< x, e_j >$, $j = 1, ..., p$.

Lemma 4.1 *Let $x \in \Re^p$ and $j \in \{1, ..., p\}$. Let us set $V_j(x) = \{x + \alpha e_j : \alpha \in \Re\}$. Then H^j increases with α on $V_j(x)$.*

PROOF.- Let $x \in \Re^p$ and $y = x + \alpha e_j$ and $y' = x + \alpha' e_j$ such that y and y' belong to D. Then

$$(\mathrm{H}^j(y) - \mathrm{H}^j(y'))\,(\alpha - \alpha') =< \vec{H}^j(y) - \vec{H}^j(y'), y - y' >=$$

$$< H(y) - H(y'), y - y' > \geq 0$$

because H is Z-increasing.

The following lemma is included in Theorem 1.5 in [3].

Lemma 4.2 *Assume that D= $\Re^p$ and let $\{x_n\}$ be a convergent sequence in $\Re^p$. Then the sequences $\{H^j(x_n)\}$, j = 1,2,...,p, are bounded.*

Now we are ready to prove the λ_p-a.e. continuity of H.

Proposition 4.3 *Let $H : \Re^p \longrightarrow \Re^p$ be a Z-increasing map. Then H is a.e. continuous with respect to the Lebesgue measure.*

PROOF.- Since $H = \sum_j \vec{H}^j(x)$, the proposition will be proved if we show that every H^j , $j = 1, ..., p$, is λ_p-a.e. continuous.

We only analize the case $j = 1$ because the others are identical. For that let us suppose that H^1 is not λ_p-a.e. continuous and let $D(H^1)$ be the set of discontinuities of H^1. It is well known (see e.g. the footnote in p. 343 of [5]) that the set of discontinuities is Borel-measurable for each arbitrary function.

On the other hand, the previous lemma shows that H^1 is bounded on every compact set. Then, by handling a sequence $\{K_n\}_n$ of compact sets with $K_n \uparrow \Re^p$, the hypothesis implies that $\lambda_p\,(D(H^1) \cap K_n) > 0$ for some n, hence in the measurable non-null set $M = D(H^1) \cap K_n$ we can assure that the function H^1 is bounded by a constant K ($|H^1(x)| \leq K,\ x \in M$).

Now, with the same notation as in Lemma 4.1, it turns out that there exists x_0 orthogonal to e_1 such that $V_1(x_0) \cap M$ is not denumerable (if this set where denumerable for every x_0 orthogonal to e_1 then $\lambda_p(M) > 0$ would not be possible).

Since H^1 is bounded on any compact set, we can define

$$H^1(x+) \stackrel{\text{def}}{=} \limsup_{y \to x} H^1(y) \text{ and } H^1(x-) \stackrel{\text{def}}{=} \liminf_{y \to x} H^1(y).$$

Also, for every $\delta > 0$, let us define the sets

$B^1_\delta = \{x \ : \ H^1(x+) > H^1(x) + \delta\}$, $B^2_\delta = \{x \ : \ H^1(x+) < H^1(x) - \delta\}$,

$C^1_\delta = \{x \ : \ H^1(x-) < H^1(x) - \delta\}$ and $C^2_\delta = \{x \ : \ H^1(x-) > H^1(x) + \delta\}$.

Since $B^1_{\delta_n} \cup B^2_{\delta_n} \cup C^1_{\delta_n} \cup C^2_{\delta_n} \uparrow D(H^1)$ as $\delta_n \downarrow 0$, the nondenumerability of the set $V_1(x_0) \cap M$ implies that there exists $\delta > 0$ such that one of the sets $B^1_\delta \cap V_1(x_0) \cap M$, $B^2_\delta \cap V_1(x_0) \cap M$, $C^1_\delta \cap V_1(x_0) \cap M$, $C^2_\delta \cap V_1(x_0) \cap M$ is not denumerable. Assume that e.g. the set $B^1_\delta \cap V_1(x_0) \cap M$ is not denumerable. Then there exists a sequence $\{y_n\}$ in this set such that

$$< y_n, e_1 >< < y_{n+1}, e_1 >, \ \forall n \in N.$$

If we show that $H^1(y_{n+1}) \geq H^1(y_n+)$ the proposition will be proved because by definition of B^1_δ, we would then have that

$$H^1(y_n) > H^1(y_1) + (n-1)\delta,$$

and this would contradict that $H^1(y_n) \leq K$, $\forall n \in N$.

To simplify the notation we only consider the case $n = 1$.

Let $\{x_m\}$ be a sequence which converges to y_1 and such that $\lim_m \mathrm{H}^1(x_m) = \mathrm{H}^1(y_1+)$. Then

$$0 \ \leq < H(y_2) - H(x_m), y_2 - x_m >= \sum_j \left(\mathrm{H}^j(y_2) - \mathrm{H}^j(x_m)\right)(y_2^j - x_m^j).$$

But $\{\mathrm{H}^j(x_m)\}$ is bounded by Lemma 4.2 and $\lim_m (y_2^j - x_m^j) = 0$ if $j \geq 2$; therefore we have that

$$0 \leq \liminf_m \left(\mathrm{H}^1(y_2) - \mathrm{H}^1(x_m)\right)(y_2^1 - x_m^1) =$$

$$\liminf_m \Big[\left(\mathrm{H}^1(y_2) - \mathrm{H}^1(x_m)\right)(y_2^1 - y_1^1) + \left(\mathrm{H}^1(y_2) - \mathrm{H}^1(x_m)\right)(y_1^1 - x_m^1)\Big] =$$

$$\liminf_m \left(\mathrm{H}^1(y_2) - \mathrm{H}^1(x_m)\right)(y_2^1 - y_1^1)$$

because $\lim_m (y_1^1 - x_m^1) = 0$ and $\{\mathrm{H}^1(x_m)\}$ is bounded. From these inequalities it is evident that $\mathrm{H}^1(y_1+) \leq \mathrm{H}^1(y_2)$.

The next step is the extension of the previous result to cover more general situations in which H is defined on a proper subset of

$\Re^p$ and we consider a probability measure P instead of the Lebesgue measure.

Some regularity condition on P is needed to avoid simple counterexamples which could be easily obtained by considering, e.g., a probability P on $\Re^2$ such that $\text{supp}(P) \subset \{\beta e_1 : \beta \in \Re\}$ and the map $H(x_1, x_2) = (H^1(x_1), H^2(x_2))$ defined through the expresions

$$H^1(x_1) \equiv 0 \text{ and } H^2(x_2) = \left| \begin{array}{ll} 1, & \text{if } x_2 \geq 0; \\ 0, & \text{if } x_2 < 0. \end{array} \right.$$

It is not difficult to prove that H is Z-increasing but it is continuous at P-a.e. no point.

The only specific tools used in the previous proof rely in Lemma 4.2 and in the fact that for a non-null (w.r.t. Lebesgue measure) set B and every hyperplane, there is at least a non-denumerable section of B on a point in this hyperplane. Every probability measure absolutely continuous with respect to λ_p possesses this property.

The generalization of Lemma 4.2 is not trivial because Theorem 1.5 in [3] is stated for sequences which converge to an interior point in the domain of H. Nevertheless we are going to generalize Lemma 4.2 by assuming that the probability P verifies certain additional property. Also we will show in the appendix (Proposition 6.1) that absolute continuity w.r.t. λ_p implies that condition.

On the other hand, this condition is a little bit more restrictive than that needed to carry out our proof of Lemma 4.2 but we state it as follows, to be used in Section 6.

Given $x \in \Re^p$ and $\delta > 0$ we denote by $B(x, \delta)$ the open ball with radious δ centered at x and, if $x, z \in \Re^p$ and $\delta, \alpha > 0$, we denote

$$S(x, z, \delta, \alpha) = B(x, \delta) \cap \{y \neq x : \text{ ang}(y - x, z) < \alpha\}.$$

Given x in the support of a probability P, let us say that P possesses property $\mathcal{C}$ at x if for every $z \in \Re^p$ and $\delta, \alpha > 0$, we have that $P[S(x, z, \delta, \alpha)] \neq 0$. We will say that P possesses property $\mathcal{C}$, if

$$P\{x : P[S(x, z, \delta, \alpha)] > 0, \forall z \in \Re^p, \delta \text{ and } \alpha > 0\} = 1.$$

Property $\mathcal{C}$ is satisfied by some probabilities which are not absolutely continuous with respect to λ_p (for instance, consider a probability measure, on $\Re^2$, which gives positive probability to every point whose coordinates are both rational numbers). However, even in the real case, this property is not sufficient to guarantee the P-a.s. continuity of an increasing function.

We generalize Lemma 4.2 as follows.

Lemma 4.4 *Let $\{x_n\} \subset D$ be a sequence that converges to a point x where P verifies $\mathcal{C}$. Then the sequences $\{H^j(x_n)\}, j = 1, 2, ..., p$ are bounded.*

PROOF.- If we suppose that $\{\mathrm{H}^j(x_n)\}$ is not bounded for some j then there exists a subsequence (which we can w.l.o.g. identify with the original one) such that if we set

$$J^+ = \{j : \mathrm{H}^j(x_n) \to +\infty\} \text{ and } J^- = \{j : \mathrm{H}^j(x_n) \to -\infty\},$$

then $J^+ \cup J^-$ is not empty and the sequences $\{\mathrm{H}^j(x_n)\}$ are bounded if $j \notin J^+ \cup J^-$.

Given $\epsilon > 0$, let us define

$$A_\epsilon := \{y \neq x : y^j > x^j + \epsilon, j \in J^+ \text{ and } y^j < x^j - \epsilon, j \in J^-\}.$$

P possesses property $\mathcal{C}$ at x, so that there exists $\epsilon > 0$ such that $P(A_\epsilon) > 0$. Consider $x_0 \in A_\epsilon$ and set $K = \sup_j |H^j(x_0)|$ and $K^* = \sup\left(K, \sup_{j \notin J^+ \cup J^-} \sup_n |\mathrm{H}^j(x_n)|\right)$.

Trivially $K^* < \infty$ and we have that

$$< H(x_0) - H(x_n), x_0 - x_n >= \sum_j \left(\mathrm{H}^j(x_0) - \mathrm{H}^j(x_n)\right)(x_0^j - x_n^j) \leq$$

$$2K^* \sum_{j \notin J^+ \cup J^-} |x_0^j - x_n^j| + \sum_{j \in J^+} \left(H^j(x_0) - \mathrm{H}^j(x_n)\right)(x_0^j - x_n^j) - \sum_{j \in J^-} \left(\mathrm{H}^j(x_n) - H^j(x_0)\right)(x_0^j - x_n^j)$$

which converges to $-\infty$ because the differences $\{|x_0^j - x_n^j|\}$ are bounded for every j and they are greater than $\epsilon/2$ from an index onward if $j \in J^+ \cup J^-$. But this is not possible because H is an increasing map.

Now, taking into account Proposition 6.1 and the preceding argument we can conclude with:

Theorem 4.5 *Let $H : D \subset \Re^p \to \Re^p$ be a Z-increasing map and let $P \ll \lambda_p$ be a probability measure such that $P(D) > 0$. Then*

$$P\{x \in D : H \text{ is not continuous at } x\} = 0.$$

Corollary 4.6 *Let $H : D \subset \Re^p \to \Re^p$ be a Z-increasing map and let P be a probability measure absolutely continuous with respect to the Lebesgue measure such that $P(D) = 1$. Then H is P-a.e. continuous.*

5 Convergence of representations

In this section we give a theorem which relates convergence in law with the a.s.-convergence of the representations given through the o.t.p.'s. The proof for this result uses ideas from [18]. The following auxiliary results are stated here for completeness. The proposition is a consequence of Theorem 2 in [16] and the lemma follows from Lemma 8.3 in [2].

Proposition 5.1 *Let P,Q be probability measures in $\mathcal{P}_2$ such that $P \ll \lambda_p$. Suppose that $H : D \subset \Re^p \to \Re^p$ is a Z-increasing map such that $(X,H(X))$ is an o.t.p. for (P,Q). Let A be such that $P(A) > 0$ and let P^* be the P-conditional probability measure given A. Let Y be a r.v. such that $P_Y = P^*$ and let $Q^* = \nu_{H(Y)}$. Then $(Y,H(Y))$ is an o.t.p. for (P^*, Q^*) and*

$$W(P^*, Q^*) = \frac{1}{P(A)} \int_A \|x - H(x)\|^2 d\nu.$$

Lemma 5.2 *Let P be a probability measure in $\mathcal{P}_2$. Let $\{B_n\}$ be a sequence of measurable sets such that $\lim_n P(B_n) = 1$. Let Q_n be the P-conditional probability given B_n, $n \in \mathcal{N}$. Then*

$$\lim_{n\to\infty} W(Q_n, P) = 0.$$

In the following technical lemmas, the notation $X_n \stackrel{L}{\rightarrow} X$ means that the sequence $\{X_n\}$ converges in law to the r.v. X.

Lemma 5.3 *Let P be a probability measure in $\Re^p$ absolutely continuous with respect to λ_p. Let $\{H_n\}$ and H be Z-increasing functions such that $P[H^{-1}(\Re^p)] = P[H_n^{-1}(\Re^p)] = 1$ and that $(X, H_n(X)) \stackrel{L}{\rightarrow} (X, H(X))$, where X verifies that $\nu_X = P$. Then*

$$lim_n P\left[H^{-1}(C) \cap H_n^{-1}(C^c)\right] = 0.$$

for every open set C.

PROOF.- According to Corollary 4.6, let A be the P-probability one set in which H is continuous. Then, if C is an open set, $H^{-1}(C) \cap A$ is an open set with respect to the induced topology in A and there exists an open set in $\Re^p$, Γ, such that

$$H^{-1}(C) \cap A = \Gamma \cap A.$$

Moreover P(A) = 1 and then

$$P[H^{-1}(C) \cap H_n^{-1}(C)] = P[H^{-1}(C) \cap A \cap H_n^{-1}(C)] =$$

$$P[\Gamma \cap H_n^{-1}(C)] = \nu_{(X,H_n(X))}[\Gamma \times C].$$

Since $(X, H_n(X)) \stackrel{L}{\rightarrow} (X, H(X))$ and $\Gamma \times C$ is an open set in $\Re^p \times \Re^p$ we have

$$P[H^{-1}(C)] \geq \limsup P[H^1(C) \bigcap H_n^{-1}(C)] \geq \liminf P\left[H^{-1}(C) \cap H_n^{-1}(C)\right] =$$

$$\liminf \nu_{(X,H_n(X))}(\Gamma \times C) \geq \nu_{(X,H(X))}(\Gamma \times C) = P[H^{-1}(C)],$$

and we have proved that $\{P\left[H^{-1}(C) \cap H_n^{-1}(C)\right]\}$ converges to $P\left[H^{-1}(C)\right]$ or, in other words, that $\{P\left[H^{-1}(C) \cap H_n^{-1}(C^c)\right]\}$ converges to zero.

Lemma 5.4 *Let $\delta > 0, \epsilon \in (0, \frac{\pi}{2}), a \in \Re^p$. Then there exists $h > 0$ such that*

$$B(a, h) \subset \{y : ang(y - u, u - a) > \pi - \epsilon\} \text{ as } \|u - a\| > \delta$$

PROOF.- Consider $u \in \Re^p$ such that $\|u - a\| > \delta$ and let

$$\hat{u} = a + \frac{\delta}{\|\mu - a\|}(u - a).$$

It suffices to prove that

$$\{y : ang(y - \hat{u}, \hat{u} - a) > \pi - \epsilon\} \subset \{y : ang(y - u, u - a) > \pi - \epsilon\}$$

and then take $h < \delta \sin \epsilon$. The computations to prove these two steps are the same as those in Lemma 4.6 in [18]

Proposition 5.5 *Let P be a probability measure on $\Re^p$, absolutely continuous with respect to λ_p. Let $H : D \subset \Re^p \to \Re^p$, $H_n : D_n \subset \Re^p \to \Re^p$, $n \in \mathcal{N}$ be Z-increasing functions such that $P(D) = P(D_n) = 1$. Let X be a r.v. with $\nu_X = P$ and such that $(X, H_n(X)) \stackrel{L}{\to} (X, H(X))$. Then*

$$H_n(X) \stackrel{a.s.}{\to} H(X).$$

PROOF.- Let A:$= \{x \in D :\ P$ holds $\mathcal{C}$ on x and H is continuous on $x\}$. By Proposition 6.1 and Corollary 4.6 we have that P(A) $= 1$ and if we prove that $H_n(x) \underset{n\to\infty}{\to} H(x)$, for every x in $A \cap D \cap \liminf D_n$ the proposition will result.

Let $x \in A \cap D \cap (\liminf D_n)$ and suppose that $\{H_n(x)\}$ does not converge to $H(x)$. Then there exists a positive real number δ and a subsequence (which we identify with the original one to simplify the notation) which verify

$$\|H_n(x) - H(x)\| > \delta, \text{ for every n, and } \frac{H_n(x) - H(x)}{\|H_n(x) - H(x)\|} \underset{n\to\infty}{\to} z,$$

where $\|z\| = 1$.

Let $\epsilon \in (0, \frac{\pi}{2})$. By Lemma 5.4 there exists $h > 0$ such that

$$B(H(x), h)_{\subset}\{y : ang(y - u, u - H(x)) > \pi - \epsilon\} \tag{4}$$

if $\|u - H(x)\|_{>}\delta$.

Note that $x \in H^{-1}[B(H(x),h)] \cap A$ and that this set is open on A, therefore, there exists $\delta^* > 0$ such that

$$B(x,\delta^*) \cap A \cap D \cap (\liminf D_n) \subset H^{-1}[B(H(x),h)] \cap A \cap D \cap (\liminf D_n).$$

Let $\theta > 0$ such that $\theta + \frac{\pi}{2} < \pi - \epsilon$. As P holds $\mathcal{C}$ on x we have that

$$P\left[S(x,z,\delta^*,\theta) \cap H^{-1}\left(B(H(x),h)\right)\right] = P[S(x,z,\delta^*,\theta)] > 0.$$

Let y be an element in $S(x,z,\delta^*,\theta) \cap (\liminf D_n)$. The increasing character of H_n gives us that

$$ang(H_n(y)-H_n(x), H_n(x)-H(x)) \leq ang(H_n(y)-H_n(x), z) + ang(z, H_n(x)-H(x)) \leq$$

$$ang(y-x,z) + \frac{\pi}{2} + ang(z, H_n(x)-H(x)) < \theta + \frac{\pi}{2} + ang(z, H_n(x)-H(x))$$

and therefore

$$\liminf_n ang(H_n(y)-H_n(x), H_n(x)-H(x)) \leq \theta + \frac{\pi}{2} < \pi - \epsilon.$$

From this and (4) we have that from an index onward $H_n(y) \notin B(H(x),h)$ and then the sequence $\{P\left[H^{-1}\left(B(H(x),h)\right) \cap H_n^{-1}\left(B^c(H(x),h)\right)\right]\}$ does not converge to zero, which contradicts Lemma 5.3.

Theorem 5.6 *Let* $\{Q_n\}$, Q *and* P *be probability measures in* $\mathcal{P}_2$ *such that* $P \ll \lambda_p$ *and* $Q_n \xrightarrow{\mathcal{L}} Q$.

Let us assume that $(X, H_n(X))$ *are o.t.p.'s for* (P, Q_n), $n \in \mathcal{N}$. *If* $H : D \subset \Re^p \to \Re^p$, *is a Z-increasing map such that (X,H(X)) is an o.t.p. for* (P,Q), *then*

$$H_n(X) \xrightarrow{a.s.} H(X).$$

PROOF.- Let r be a natural number and let A_r be a bounded Q-continuity set, such that $Q(A_r) > 1 - \frac{1}{r}$. We define the probability measures

$$Q^r(B) = \frac{Q[B \cap A_r]}{Q[A_r]},$$

$$P^{n,r}(B) = \frac{P[B \cap H_n^{-1}(A_r)]}{P[H_n^{-1}(A_r)]},$$
$$Q^{n,r}(B) = \frac{P[H_n^{-1}(B) \cap H_n^{-1}(A_r)]}{P[H_n^{-1}(A_r)]}.$$

Let $X_{n,r}$ be a r.v. with probability distribution $P^{n,r}$. Then it is evident that $\nu_{H_n(X_{n,r})} = Q^{n,r}$ so the support of $Q^{n,r}$ is a.s. contained in A_r and it is bounded. Trivially, the same happens with Q^r. These facts imply (taking into account that A_r is a Q-continuity set and that $Q_n = P \circ H_n$) that

$$\lim_n W(Q^{n,r}, Q^r) = 0, \text{ for every } r \in \mathcal{N}.$$

On the other hand from Lemma 5.2 we obtain that $\lim_r W(Q^r, Q) = 0$ and then there exists a subsequence such that

$$\lim_r W(Q^{n_r,r}, Q) = 0,$$

and that

$$\lim_r P[H_{n_r}^{-1}(A_r)] = 1.$$

So, once again, Lemma 5.2 implies that

$$\lim_r W(P^{n_r,r}, P) = 0$$

and we have proved that

$$\lim_r W(P^{n_r,r}, Q^{n_r,r}) = W(P, Q),$$

and also that $P^{n_r,r} \stackrel{\text{Ł}}{\to} P$ and $Q^{n_r,r} \stackrel{\text{Ł}}{\to} Q$. Therefore we have that the sequence $\{\nu_{(X_{n_r,r};H_{n_r}(X_{n_r,r}))}\}_r$ is tight.

Moreover the Proposition 5.1 implies that $(X_{n_r,r}; H_{n_r}(X_{n_r,r}))$ is an o.t.p. for $(P^{n_r,r}, Q^{n_r,r})$. So if we consider a weakly convergent subsequence, $\{\nu_{(X_{n_{r^*},r^*};H_{n_{r^*}}(X_{n_{r^*},r^*}))}\}_{r^*}$, and μ is its limit, we have that

$$W(P, Q) = \lim_r \int \|x - y\|^2 \nu_{(X_{n_{r^*},r}\ H_{n_{r^*}}(X_{n_{r^*},r}))}(dx, dy) \geq$$

$$\int \|x - y\|^2 \mu(dx, dy) \geq W(P, Q),$$

where the first inequality follows from a well-known property of the weak convergence (see, for instance, Theorem 5.3 in [4]) and the second one holds because the marginal distributions of μ are P and Q.

Therefore μ gives an o.t.p. between P and Q and $\mu = \nu_{(X,H(X))}$, whence we have proved that every weakly-convergent subsequence of $\{(X_{n_r,r}; H_{n_r}(X_{n_r,r}))\}_r$, converges to $(X, H(X))$ and the subsequence $\{(X_{n_r,r}; H_{n_r}(X_{n_r,r}))\}_r$ converges weakly to $(X, H(X))$.

Now note that X is any r.v. with distribution P, and that $P^{n_r,r}$ is the P-conditional distribution given $H_{n_r}^{-1}(A_r)$. Therefore, if Y_{n_r} is a r.v. whose distribution is the P-conditional distribution given $\left[H_{n_r}^{-1}(A_r)\right]^c$ and B is a Bernoulli r.v., independent of $X_{n_r,r}$ and of Y_{n_r}, such that $\nu[B=1] = P[H_{n_r}^{-1}(A_r)]$, we can assume that

$$X = BX_{n_r,r} + (1-B)Y_{n_r}$$

and then, given $\epsilon > 0$,

$$\nu[\|(X_{n_r}, H_{n_r}(X_{n_r})) - (X, H_{n_r}(X))\| \geq \epsilon] \leq \nu[X_{n_r} \neq X] = \nu[B=0],$$

which converges to zero because $\{P[H_{n_r}^{-1}(A_r)]\}$ converges to one.

So we have proved that

$$(X, H_{n_r}(X)) \xrightarrow{\mathcal{L}} (X, H(X)). \tag{5}$$

Finally, if we apply the preceeding argument, not to the original sequence $\{(X, H_n(X))\}_n$, but to any of its subsequences, we would have that every subsequence of the whole sequence contains a second subsequence which satisfies (5) and therefore, the sequence $\{(X, H_n(X))\}_n$ also satisfies (5).

Now the result is obtained by applying Proposition 5.5.

To apply the previous scheme to the usual statistical framework, let us consider an absolutely continuous probability measure P in $\mathcal{P}_2$ and a sequence of (pairwise) independent identically distributed random variables $X_1, \ldots, X_n, \ldots$ defined in (Ω, σ, ν), with $\mathcal{L}(X_n) = Q \in \mathcal{P}_2$. Let Q_n^ω be the sample distribution based on $X_1(\omega), \ldots, X_n(\omega)$, $\omega \in \Omega$.

Theorem 5.7 (Consistency of the sample optimal transportation polans) *With the previous hypothesis, let* $(X, H_n^\omega(X))$ *be a sample o.t.p. for* (P, Q_n^ω) *and let* $H : D \subset \Re^p \to \Re^p$ *be an increasing map such that* $(X, H(X))$ *is an o.t.p. for (P,Q), then* $H_n^\omega(X) \overset{a.s.}{\to} H(X)$ *for* ν*-a.e.*$\omega \in \Omega$.

PROOF.- Obvious from Theorem 5.6 taking into account the available results of Glivenko-Cantelli type.

Theorems 5.6 and 5.7 can be very useful when computing the o.t.p. between two probability measures when it is possible to find an expression for the functions H_n^ω on it. This can be done in some cases. For instance in [1] a simple way is suggested for the construction of these functions if the density of P is continuous.

6 Appendix

We begin with the proof that the absolutely continuous probability measures satisfy property $\mathcal{C}$.

Proposition 6.1 *Let P be a probability measure on* $\Re^p$ *such that* $P \ll \lambda_p$*. Then P satisfies property* $\mathcal{C}$*.*

PROOF.- Suppose that P does not possesses $\mathcal{C}$. If we denote

$A = \{x :$ there exist $z \in \Re^p, \delta$ and $\alpha > 0$ such that $P[S(x, z, \delta, \alpha)] = 0\}$,

this means that $P(A) > 0$. In fact the measurability of A can be easily obtained:

Let $\{z_n\}$ be a denumerable dense subset of the unit sphere and let $x, z \in \Re^p$ and $\delta, \alpha > 0$. Trivially there exist $n \in \mathcal{N}$ and δ^* and α^* rational numbers such that

$$S(x, z, \delta, \alpha) \supset S(x, z_n, \delta^*, \alpha^*)$$

and therefore

$$A = \bigcup_{\substack{n \in \mathcal{N} \\ \delta, \alpha \in Q}} \{x : P[S(x, z_n, \delta, \alpha)] = 0\}.$$

Note that, if we fix z, δ and α then the map $x \to P[S(x, z, \delta, \alpha)]$ is continuous and hence measurable. Therefore $\{x : P[S(x, z, \delta, \alpha)] = 0\}$ is measurable for every z, δ, α fixed. So A is measurable because it is a denumerable union of measurable sets.

Indeed $P(A) > 0$ implies that there exist z_n in the unit sphere, δ^* and $\alpha^* > 0$ such that $P\{x : P[S(x, z_n, \delta^*, \alpha^*)] = 0\} > 0$ and we can conclude that there exists $K \in \Re$ such that, if we denote

$$A^* = [-K, K]^p \cap \{x : P[S(x, z_n, \delta^*, \alpha^*)] = 0\}$$

then $P(A^*) > 0$.

From now on, the values z_n, δ^* and α^* will remain fixed and we will use the notation $S(x) \equiv S(x, z_n, \delta^*, \alpha^*)$.

We prove next that $P\left[\cup_{x \in A^*} S(x)\right] = 0$. This set is measurable because it is an union of open sets. Let $\{B_n\}$ be a denumerable basis for the topology in $\Re^p$. Therefore, given $x \in A^*$, there exists a denumerable family $\{B_{n_k(x)}\}$ such that $S(x) = \cup_k B_{n_k(x)}$.

$S(x)$ is a P-probability zero set, so that $P[B_{n_k(x)}] = 0$, $\forall k$ and then $\cup_{x \in A^*} S(x)$ can be written as a denumerable union of P-probability zero sets and $P\left[\cup_{x \in A^*} S(x)\right] = 0$ as we wanted to prove.

Note that without loss of generality we can assume that $A^* \cap (\cup_{x \in A^*} S(x)) = \emptyset$. In particular, given the point $x \in A^*$ consider the set $V_{z_n}^{\delta^*}(x) = \{x + h z_n : h \in (0, \delta^*)\}$. By definition of $S(x)$, it is evident that $V_{z_n}^{\delta^*}(x) \subset S(x)$, and then $A^* \cap V_{z_n}^{\delta^*}(x) = \emptyset$.

Given the real number t denote $A_t^* := A^* \cap \{y \in \Re^p :< y, z_n >= t\}$. Therefore Fubini's theorem implies

$$\lambda_p(A^*) = \int_{-\infty}^{\infty} \lambda_{p-1}(A_t^*) dt,$$

where λ_{p-1} denotes the Lebesgue measure on $\Re^{p-1}$.

The absolute continuity of P guarantees that $\lambda_p(A^*) > 0$ also. Then there exists a positive number ϵ such that the set $C_\epsilon = \{t \in \Re : \lambda_{p-1}(A_t^*) \geq \epsilon\}$ is infinite. For each point t in C_ϵ we consider the new set

$$H_t = \cup_{y \in A_t^*} V_{z_n}^{\delta^*}(y).$$

$H_t \cap A^* = \emptyset$ because $A_t^* \subset A^*$ and then $V_{z_n}^{\delta^*}(y) \cap A^* = \emptyset, \forall y \in A_t^*$.

The key property of $\{H_t : t \in C_\epsilon\}$ is that

$$\text{if } t, t' \in C_\epsilon,\ t \neq t' \text{ then } H_t \cap H_{t'} = \emptyset. \tag{6}$$

.

This is evident if $|t - t'| \geq \delta^*$ by construction of these sets. So suppose that $|t - t'| < \delta^*$ and that $t < t'$. If $y \in A^*_{t'}$, as $A^*_{t'} \subset A^*$, then $y \notin H_t$ and then $V^{\delta^*}_{z_n}(y) \cap H_t = \emptyset$.

Finally, let $t \in C_\epsilon$. By construction of H_t we have that

$$H_t \subset [-K, K + \delta^*]^p$$

and that $\lambda_p(H_t) \geq \epsilon \times \delta^*$ but this contradicts (6) and that C_ϵ is an infinite set. Then $\lambda_p(A^*) = 0$.

Next we prove the Theorem 2.3. We need two technical lemmas.

Lemma 6.2 *Let $H : T \subset \Re^p \to \Re^p$ and let $x_0,\ x_1 \in T$ such that $ang(x_1$ - x_0, $H(x_1)$ - $H(x_0)) > \pi/2$. Then*

$$lim_{\delta \to 0} inf \left\{ \|z\| : \begin{array}{l} ang(x_1 - x_0, z - H(x_0)) \leq \frac{\pi}{2} + \delta, and \\ ang(x_0 - x_1, z - H(x_1)) \leq \frac{\pi}{2} + \delta \end{array} \right\} = \infty.$$

PROOF.- Let us denote

$$S_\delta = \{z : \sup\ (ang(x_1 - x_0, z - H(x_0)), ang(x_0 - x_1, z - H(x_1))) \leq \frac{\pi}{2} + \delta\}.$$

The sets $\{S_\delta, \delta > 0\}$ decrease with δ. Therefore, if we suppose that the lemma does not hold, there exists $K > 0$ such that

$$S^K_\delta = S_\delta \cap \{z : \|z\| \leq K\}$$

is not empty for every δ.

Now note that $\{S^K_\delta : \delta > 0\}$ is a family of compact sets such that every intersection of a finite number of them is not empty. So $\bigcap_\delta S^K_\delta$ is also non empty. But it is evident that

$$\bigcap_\delta S^K_\delta \subset \left\{ z : \sup\ (ang(x_1 - x_0, z - H(x_0)), ang(x_0 - x_1, z - H(x_1))) \leq \frac{\pi}{2} \right\}$$

which is an empty set by hypothesis.

Lemma 6.3 *Given x, z* $\in \Re^p$ *and* $\alpha > 0$*, we denote by C(x,z,*α*) the set* $S(x, z, \infty, \alpha)$*. Let P be a probability measure absolutely continuous with respect to* λ_p*.*

Fix $\alpha_0 > 0$ *and consider* $z_1, ..., z_m \in \Re^p$ *. Let* $x_0 \in \Re^p$ *such that P possesses property* $\mathcal{C}$ *at it. Then there exist* $h, \delta > 0$ *such that if* $\|x - x_0\| < \delta$*,*

$$inf_i P[C(x, z_i, \alpha_0)] \geq h.$$

PROOF.- Fix $\alpha_0 > 0$, let $z_1, ..., z_m$ be in the unit sphere with center at 0, S(0,1), and let x_0 be such that P holds property $\mathcal{C}$ on x_0. Then $P[C(x_0, z, \alpha_0)] > 0$ for every z. Moreover, the absolute continuity of P guarantees that the map $z \to P[C(x_0, z, \alpha_0)]$ is continuous. So there exists $z_0 \in S(0, 1)$ such that

$$0 < h = P[C(x_0, z_0, \alpha_0)] = \inf_{z \in S(0,1)} P[C(x_0, z, \alpha_0)].$$

This and the fact that all maps $x \to P[C(x, z_i, \alpha_0)]$, $i = 1, ..., m$ are also continuous prove the lemma.

Theorem 6.4 *Let P be a probability measure absolutely continuous with respect to* λ_p*. Let T be a set such that P(T)=1 and suppose that* $H : T \subset \Re^p \to \Re^p$ *is a map that satisfies*

$$P \otimes P\{(x, x') :< x - x', H(x) - H(x') >< 0\} = 0.$$

Then there exists a P-probability one set, D, such that H is Z-increasing on D.

PROOF.- Let A $\stackrel{\text{def}}{=} \{(x, x') :< x - x', H(x) - H(x') >\geq 0\}$.

Let x $\in$T and $A_x \stackrel{\text{def}}{=} \{x' : (x, x') \in A\}$ and $D \stackrel{\text{def}}{=} \{x : P(A_x) = 1$ and P holds $\mathcal{C}$ on $x\}$.

By hypothesis and Proposition 6.1, we have that P(D) = 1.

Let $x_0 \in D, x_1 \in T$ such that $< x_0 - x_1, H(x_0) - H(x_1) >< 0$. If we prove that $x_1 \notin D$ the theorem will be proved. Hence let us suppose that, on the contrary, $x_1 \in D$.

Take $\alpha > 0$ such that

$$ang(x_1 - x_0, H(x_1) - H(x_0)) > \frac{\pi}{2} + \alpha. \tag{7}$$

Choose $\alpha_0 \in (0, \frac{\pi}{4})$ and consider a family $\{z_1, ..., z_m\}$ in S(0,1) such that, for every z in S(0,1), $\inf_i |ang(z, z_i)| < \alpha_0$. Let h and δ be the numbers that we obtain by applying Lemma 6.3 to x_0, α_0 and $\{z_1, ..., z_m\}$.

Trivially there exists $\delta^* \leq \delta$ such that

$$S(x_0, x_1 - x_0, \delta^*, \alpha) \subset C(x_1, x_0 - x_1, \alpha).$$

P possesses $\mathcal{C}$ at x_0, so there exists a point y_α in $S(x_0, x_1 - x_0, \delta^*, \alpha) \cap A_{x_0} \cap A_{x_1} \cap D$. By definition of these sets, y_α satisfies that

$$ang(x_1 - x_0, H(y_\alpha) - H(x_0)) < \frac{\pi}{2} + \alpha,$$

and

$$ang(x_0 - x_1, H(y_\alpha) - H(x_1)) < \frac{\pi}{2} + \alpha.$$

Let $u = \frac{1}{2}(H(x_0)+H(x_1))$ and consider the vector $v_\alpha = H(y_\alpha)-u$. (7) implies that $ang(x_1 - x_0, u - H(x_0)) > \frac{\pi}{2} + \alpha$ and then $v_\alpha \neq 0$.

On the other hand, we have that $u \notin C(H(y_\alpha), v_\alpha, \frac{\pi}{2} + 2\alpha_0)$. So there exists $i(\equiv i(\alpha))$ such that $u \notin C(H(y_\alpha), z_i, \frac{\pi}{2} + \alpha_0)$.

Now by definition of A_{y_α},

$$H\left[C(y_\alpha, z_i, \alpha_0) \cap A_{y_\alpha}\right] \subset C(H(y_\alpha), z_i, \frac{\pi}{2} + \alpha_0)$$

and, as $y_\alpha \in D$ and $\|x_0 - y_\alpha\| < \delta$, we obtain that

$$P[C(y_\alpha, z_i, \alpha_0) \cap A_{y_\alpha}] \geq h$$

and therefore

$$P \circ H^{-1}\left[C\left(H(y_\alpha), z_i, \tfrac{\pi}{2} + \alpha_0\right)\right] \geq h. \quad (8)$$

The application of Lemma 6.2 gives then $\lim_{\alpha\to 0}\|H(y_\alpha)\| = \infty$, and, by construction, $ang(u - H(y_\alpha), z_{i(\alpha)}) > \frac{\pi}{2} + \alpha_0$. So, if we denote

$$r_\alpha = \sup\left\{r : B(u, r) \cap C\left(H(y_\alpha), z_{i(\alpha)}, \frac{\pi}{2} + \alpha_0\right) = \emptyset\right\},$$

we have that $\lim_{\alpha\to 0} r_\alpha = \infty$ which contradicts (8).

We finalize with the proof of the Proposition 3.2. (Note that in the proof we do not need to assume that the maps g_i are increasing.)

PROOF of Proposition 3.2.- First we show that $f_n(t) = \beta_n t$, $t \in \Re$ for some $\beta_n \geq 0$ and every $n = 1, ..., p$.

Since $\{\hat{e}_1, ..., \hat{e}_p\}$ is an orthonormal basis, each e_i can be written as $e_i = \sum_k a_k^i \hat{e}_k$. Hence

$$H\left(\sum_i x_i e_i\right) = H\left(\sum_i x_i \left(\sum_k a_k^i \hat{e}_k\right)\right) = H\left(\sum_k \left(\sum_i x_i a_k^i\right) \hat{e}_k\right) = \sum_k g_k \left(\sum_i x_i a_k^i\right) \hat{e}_k.$$

On the other hand, it holds that

$$H\left(\sum_i x_i e_i\right) = \sum_i f_i(x_i) e_i = \sum_i f_i(x_i) \left[\sum_k a_k^i \hat{e}_k\right] = \sum_k \left(\sum_i a_k^i f_i(x_i)\right) \hat{e}_k;$$

and it follows that

$$g_k \left(\sum_i x_i a_k^i\right) = \sum_i a_k^i f_i(x_i), \text{ for every } k \text{ and } x_1, ..., x_p. \qquad (9)$$

Now let us consider any $n \in \{1, ..., p\}$ and let k be such that $a_k^n \neq 0$ (if such k would not exist, then $e_n = 0$ and $\{e_1, ..., e_p\}$ could not be a basis).

Consider t, z in $\Re$ and take two sets of real numbers $\{x_1, ..., x_p\}$ and $\{\hat{x}_1, ..., \hat{x}_p\}$ such that $\sum_i a_k^i x_i = \sum_i a_k^i \hat{x}_i$, with $x_n = t$ and $\hat{x}_n = z$. Such a choice is possible because $a_k^i \neq 0$ for some $i \neq n$ (note that in other case we would have $< e_j,\, j \neq n \;\; > \subset < \hat{e}_j,\, j \neq k >$ so $e_n = \pm \hat{e}_k$). This and (9) imply

$$\sum_i a_k^i f_i(x_i) = \sum_i a_k^i f_i(\hat{x}_i).$$

Also note that with the choice above, for every r in $\Re$ it holds that

$$\sum_{i \neq n} a_k^i x_i + a_k^n (x_n + r) = \sum_{i \neq n} a_k^i \hat{x}_i + a_k^n (\hat{x}_n + r),$$

and by (9)

$$\sum_{i \neq n} a_k^i f_i(x_i) + a_k^n f_n(x_n + r) = \sum_{i \neq n} a_k^i f_i(\hat{x}_i) + a_k^n f_n(\hat{x}_n + r).$$

Therefore

$$a_k^n(f_n(x_n + r) - f_n(x_n)) = a_k^n(f_n(\hat{x}_n + r) - f_n(\hat{x}_n))$$

and

$$f_n(t + r) - f_n(t) = f_n(z + r) - f_n(z), \text{ for every } t, z, r \in \Re$$

then, in particular,

$$f_n(t + r) - f_n(t) = f_n(r) - f_n(0) = f_n(r), \text{ for every } t, r \in \Re.$$

Finally the facts that $f_n(t + r) = f_n(t) + f_n(r)$ for every $t, r \in \Re$ and the monotone nondecreasing character of f_n suffice to ensure that $f_n(t) = \beta_n t$ for some $\beta_n \geq 0$ and every $t \in \Re$.

Now the Proposition is a consequence of the double diagonalization of the linear map H in the bases $\{e_1, ..., e_p\}$ and $\{\hat{e}_1, ...\hat{e}_p\}$ with the imposed restrictions $e_i \neq \pm\hat{e}_j$, for all i, j $(i \neq j)$.

References

[1] AURENHAMMER, F. , HOFFMANN, F. and ARONOV, B. Minkowski-type theorems and least-squares partitioning. *Report B-92-09^T from "Institute for Computer Science". Freie Universität Berlin.*

[2] BICKEL, P.J. and FREEDMAN, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* 9, 1196-1217.

[3] BARBU, V. (1976). *Nonlinear semigroups and differential equations in Banach spaces.* Noordhoff International Publishing: Leiden.

[4] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley: New York.

[5] BILLINGSLEY, P. (1986). *Probability and measure.* Wiley: New York.

[6] CUESTA, J.A. and MATRAN, C. (1989). Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.* 17, 1264-1276.

[7] CUESTA, J.A., MATRAN, C. and TUERO-DIAZ, A. (1992). On the lower bound for the L^2-Wasserstein metric in a Hilbert space. Preprint.

[8] CUESTA, J.A., RUSCHENDORF, L. and TUERO-DIAZ, A. (1993). Optimal coupling of multivariate distributions and stochastic processes. *J. Multivariate Anal.* 46, 355-361.

[9] CUESTA-ALBERTOS, J.A. and TUERO-DIAZ, A. (1993). A characterization for the solutions of the Monge-Kantorovich mas transference problem. *Stat. and Prob. Letters* 16, 147-152.

[10] GELBRICH, M. (1990). On a formula for the L^2-Wasserstein Metric between Measures on Euclidean and Hilbert Spaces. *Match. Nachr.* 147, 185-203.

[11] KNOTT, M. and SMITH, C.S. (1984). On the optimal mapping of distributions. *J. Optim. Th. Appl.* 52, 323-329.

[12] LORENTZ, G. G. (1953) An inequality for rearrangements *Am. Math. Monthly,* 60, 176-179.

[13] OLKIN, I. and PUKELSHEIM, F. (1982). The distance between two random vectors with given dispersion matrices. *Lin. Algebra Appl.* 48, 257-263.

[14] RACHEV, S.T. (1991). *Probability Metrics and the Stability of Stochastic Models.* Wiley.

[15] ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton University Press.

[16] RUSCHENDORF, L. and RACHEV, S.T. (1990). A characterization of random variables with minimun L^2-distance. *J. Multivariate Anal.* 32, 48-54.

[17] SMITH, C.S. and KNOTT, M. (1987). Note on the optimal transportation of distributions. *J. Optim. Th. Appl.* 52, 323-329.

[18] TUERO, A. (1991). On the stochastic convergence of representations based on Wasserstein metrics. *Ann. Probab.* To appear.