Approximation to probabilities through uniform laws on convex sets*

J.A. Cuesta-Albertos
Departamento de Matemáticas, Estadística y Computación,
Universidad de Cantabria, Spain

C. Matrán[a] and J. Rodríguez-Rodríguez[a]
Departamento de Estadística, Universidad de Valladolid, Spain

Abstract

Let P be a probability distribution on $I\!R^d$ and let $\mathcal{C}$ be the family of the uniform probabilities defined on compact convex sets of $I\!R^d$ with interior non-empty. We prove that there exists a best approximation to P in $\mathcal{C}$, based on the L_2-Wasserstein distance. The approximation can be considered as the best representation of P by a convex set in the minimum squares setting, improving on other existent representations for the shape of a distribution. As a by-product we obtain properties related to the limit behavior and marginals of uniform distributions on convex sets which can be of independent interest.

Key words and phrases: Wasserstein distance, uniform laws, convex sets, existence.
A.M.S. 1991 subject classification: Primary: 60E05, Secondary: 62E17

1 Introduction

Descriptive measures of the shape of a distribution P are very often based on values or figures related to the minimum squares setting. These are the cases of the mean, variance, covariance matrix, regression,... This is also the case of the concentration ellipsoid of P, defined through the property that the uniform distribution on it has the same mean and covariance matrix as P.

A different approach, also related to the minimum squares setting, can be adopted by considering the uniform distribution on every ellipsoid and choosing that nearest to P in the L_2 sense. This is the approach introduced in [**?**] and [**?**], making use of the natural version of the L_2-distance for probability measures: the L_2-Wasserstein distance. The procedure very often has the serious drawback of computational difficulties, but also the obvious advantages deriveing from the fact that the subsequent "analysis by eye" is based on a more accurate approximation to our distribution

*Research partially supported by DGESIC, grants PB98-0369-C02-00,01 and 02.
[a]These authors have also been supported by PAPIJCYL grantVA53/00B.

than that given by the concentration ellipsoid. Within a more general framework the process can be generalized by defining a kind of figure or pattern and obtaining the best approximation, based on the L_2-Wasserstein distance, to our distribution in terms of uniform distributions supported by figures of the pattern. An application of those ideas to the goodness of fit tests can be seen at [?], [?] and [?].

In this work we begin the exploration of the possibilities of this model applied to the family of compact convex sets. This family could give much more information about the shape of the distribution than that given, for example, by the ellipsoids. In contrast, its study has a notable complexity even for proving the existence of a best approximation, which is the main objective of the work. As a by-product which can be of independent interest we will obtain some properties related to the limit behavior and marginals of uniform distributions on convex sets.

The use of simulation procedures to obtain an approximation to the solution of this problem is justified by our results in this paper and those in [?].

We will assume that all the random vectors (r.v.'s) in this paper are defined on the same rich enough probability space (Ω, σ, μ) and, unless explicitly stated, all of them will be $I\!R^d$-valued, $d \geq 1$. We will consider the usual Borel σ-algebra, β^d, and norm, $\|\cdot\|$, in this space. Given the r.v. X we will denote by P_X its probability distribution.

We will denote by $\mathcal{P}_2$ the family of all probabilities defined on β^d with finite second order moment. Given $P, Q \in \mathcal{P}_2$, the *Wasserstein distance* between them is defined as the square root of

$$W^2(P,Q) := \inf\left\{\int \|X-Y\|^2 d\mu : P_X = P \text{ and } P_Y = Q\right\}.$$

In this paper we are interested in the following problem. Let $\mathcal{C}$ be the family of the uniform probabilities defined on compact convex sets of $I\!R^d$ with a non-empty interior. Obviously $\mathcal{C} \subset \mathcal{P}_2$. Given $P \in \mathcal{P}_2$, we are interested in finding conditions to assure that there exists $Q_P \in \mathcal{C}$ such that

$$W^2(P, Q_P) = \inf_{Q\in\mathcal{C}} W^2(P,Q). \tag{1}$$

We will denote the infimum of the expression in the term on the right by $W^2(P,\mathcal{C})$.

As has been pointed out, the interest of this problem from a statistical point of view arises from the fact that the shape of the support of Q_P could give important hints on the general shape of P. Moreover, $W^2(P,\mathcal{C})$ can be employed as an alternative parameter to measure the flatness (as opposed to the peakness) of P. These aspects have been analyzed in [?].

Some additional notation to be employed will be the following. Given $P \in \mathcal{P}_2$, $\sigma(P)$ will denote its standard deviation. If $x \in I\!R^d$ and $\mathcal{H}$ is a subspace, $x_{\mathcal{H}}$ will denote the projection of x on $\mathcal{H}$. Given S, a non-empty subset of $I\!R^d$, and $x \in \mathcal{H}$ we will denote by

$$S_{\mathcal{H}}(x) := \{z \in S : z_{\mathcal{H}} = x\}$$

the section of S on x. To avoid unnecessary complications in the notation, we will suppress the subindex $\mathcal{H}$ in $S_{\mathcal{H}}(x)$ if the subspace is clearly identified.

2 Preliminary results

In order to show that the infimum in (**??**) is reached, we need to assume some regularity condition on P. Notice that, for instance, there is no sense in trying to approximate a distribution which is concentrated in a point by a continuous distribution. Thus, from now on, we will assume that P is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, λ_d. This condition suffices to obtain the following result which relates our distance between probabilities with the mass transportation problem (see [**?**]).

Proposition 2.1 *Let $P, Q \in \mathcal{P}_2$. If P is absolutely continuous with respect to λ_d, then, there exists a measurable map $H_{P,Q} : \mathbb{R}^d \to \mathbb{R}^d$ such that if X is a r.v. with distribution P, then*

1. *The distribution of $H_{P,Q}(X)$ is Q.*
2. *$W^2(P,Q) = \int \|X - H_{P,Q}(X)\|^2 d\mu$.*

Moreover, $H_{P,Q}$ is unique up to P_X-a.s. equivalence.

Remark 2.2 In fact, as shown in [**?**], a slightly weaker condition than absolute continuity of P is enough to obtain conclusions in Proposition **??**. All the results in this paper would hold under this weaker condition. However, we prefer to assume absolute continuity in order to avoid unnecessary technical complications.

The map $H_{P,Q}$ obtained in the previous proposition is called an *optimal transportation plan* (o.t.p.) between P and Q.

In the one dimensional case, if F_P is the distribution function of P and F_P^{-1} its generalized inverse, then $H_{P,Q} = F_Q^{-1} \circ F_P$. Therefore, if we denote by Q_a the uniform distribution on $[-a, a]$, then $F_{Q_a}^{-1} = aF_{Q_1}^{-1}$ and $H_{P,Q_a} = aH_{P,Q_1}$. Thus, we have that

$$\begin{aligned} W^2(P, Q_a) &= \int |X - aH_{P,Q_1}(X)|^2 d\mu \\ &= E[X^2] - 2aE[XH_{P,Q_1}(X)] + a^2 E[H_{P,Q_1}(X)^2] \\ &= E[X^2] - 2aC(P) + \frac{a^2}{3}, \end{aligned} \tag{2}$$

where $C(P)$ denotes for the covariance between X and $H_{P,Q_1}(X)$ which, obviously, depends just on P. Moreover, $C(P) > 0$ unless there exists $t_0 \in \mathbb{R}$ such that $P[t_0] = 1$ and, if this does not happen, as a consequence of (**??**) we obtain that

$$\inf_{a>0} W^2(P, Q_a) > W^2(P, Q_0) = E[X^2]. \tag{3}$$

The interest of Wasserstein distance, from a probabilistic point of view, comes from its relation with the weak convergence of probability measures plus convergence of moments. We will employ this relation later as stated in the next proposition. A proof of it can be seen e.g. in [**?**].

Proposition 2.3 *Given $Q_n, n \in I\!N, Q \in \mathcal{P}_2$, the following statements are equivalent*

1. $\lim_n W^2(Q, Q_n) = 0$.
2. *The sequence $\{Q_n\}_n$ converges weakly to Q and*

$$\lim_n \int \|x\|^2 Q_n(dx) = \int \|x\|^2 Q(dx).$$

Now we will give some properties and definitions which will be used later. The next proposition has been taken from [**?**].

Proposition 2.4 *Let $P \in \mathcal{P}_2$ absolutely continuous with respect to λ_d. Let $Q \in \mathcal{P}_2$ such that its support is contained in the subspace $\mathcal{H}$. Then, there exists an o.t.p. $H_{P,Q}$ between P and Q such that if $x, y \in I\!R^d$ satisfy that $x_{\mathcal{H}} = y_{\mathcal{H}}$, then $H_{P,Q}(x) = H_{P,Q}(y)$.*

A very useful property of the Wasserstein distance is that it is reduced if the probabilities under consideration are translated in order to make both share the same mean (see [**?**]).

Proposition 2.5 *Given $P, Q \in \mathcal{P}_2$, if P^* and Q^* denote for their centered in mean translations, then*

$$W^2(P, Q) = W^2(P^*, Q^*) + \left\| \int xP(dx) - \int xQ(dx) \right\|^2.$$

To bring to an end the properties of the Wasserstein distance, we include two lower bounds. Here, given $P \in \mathcal{P}_2$ and once a basis in $I\!R^d$ is fixed, we denote by $\{P_1, P_2, ..., P_d\}$ the family of the one-dimensional marginal distributions of P.

The first lower bound in this proposition was obtained in [**?**]. The second one is a simple consequence of properties of the L_2-norm.

Proposition 2.6 *Given $P, Q \in \mathcal{P}_2$ centered in mean, then*

$$W^2(P, Q) \geq \sum_{i=1}^{d} W^2(P_i, Q_i) \geq \sum_{i=1}^{d} \left(\sigma(P_i) - \sigma(Q_i)\right)^2.$$

We will also need some relevant properties of convex sets. They have been taken from [?] and are now stated for sake of completeness.

Given $x \in \mathbb{R}^d$ and $r > 0$, $B(x,r)$ will denote for the closed ball with center at x and radius r. Let $C \subset \mathbb{R}^d$ and let us define $C^\delta := \cup_{x\in C} B(x,\delta)$ and, given the closed, bounded sets C_1 and C_2, we define

$$\Delta(C_1, C_2) := \delta_1 + \delta_2,$$

where $\delta_1 := \inf\{\delta > 0 : C_2 \subset C_1^\delta\}$ and $\delta_2 := \inf\{\delta > 0 : C_1 \subset C_2^\delta\}$. It turns out that Δ is a metric on the family of the bounded convex and closed sets. Two important properties of Δ are the following.

Proposition 2.7 *Let $K > 0$. The family of the closed and convex sets contained in $B(0,K)$ is sequentially compact with respect to Δ.*

Proposition 2.8 *The Lebesgue measure is continuous with respect to Δ. I.e., if $\{C_n\}_n$ is a sequence of bounded, convex, closed sets which converges to the closed and convex set C, then*

$$\lim_n \lambda_d(C_n) = \lambda_d(C).$$

We will also need some properties of the probabilities in $\mathcal{C}$. We have taken them from [?]. The following notation will be used in them. Let $a, b \geq 0$, $s > 0$ and $\theta \in (0,1)$. The *s-generalized mean $M_s(a,b;\theta)$* is defined by

$$M_s(a,b;\theta) := [(1-\theta)a^s + \theta b^s]^{1/s}.$$

The limit case $s = \infty$ is defined by continuity as

$$M_\infty(a,b;\theta) := \max(a,b).$$

Remark 2.9 The s-generalized means can be also defined for negative values of s but here we only handle positive ones.

Definition 2.10 *Let Q be a probability measure defined on $\mathbb{R}^d$ and $t \in \mathbb{R}^+ \cup \{\infty\}$. We say that Q is t-concave if for both measurable sets A_0 and A_1 in its support and $\theta \in (0,1)$,*

$$Q[(1-\theta)A_0 + \theta A_1] \geq M_t[Q(A_0), Q(A_1); \theta].$$

Proposition 2.11 *If the probability measure Q is t-concave, then its support is convex.*

Proposition 2.12 *Let Q be a t-concave probability measure. If the interior of the support of Q is non-empty, then $t \leq 1/d$ and if $\mathcal{H}$ is the affine hull of its support, then Q is absolutely continuous with respect to the Lebesgue measure on $\mathcal{H}$.*

It is possible to repeat the proof of Theorem 2.10 in [?] to obtain the following proposition.

Proposition 2.13 *Let $\{Q_n\}_n$ be a sequence of t-concave probability measures which converges weakly to the probability measure Q. Then, Q is also t-concave.*

t-concavity of a probability measure is related to certain concavity property of its density function. This fact is related to the next definition through the proposition which follows:

Definition 2.14 *Given a non-negative function f on an open convex set $C \subset I\!R^d$ it is said to be s-concave if for any x_0, x_1 in C and $\theta \in (0,1)$ we have*

$$f[(1-\theta)x_0 + \theta x_1] \geq M_s[f(x_0), f(x_1); \theta].$$

Remark 2.15 Notice that if $s \neq \infty$, then f is s-concave if and only if f^s is concave.

Proposition 2.16 *Let Q be a probability measure supported by a convex set in $I\!R^d$. Let h be the dimension of the affine hull of its support. Then Q is t-concave if and only if Q admits a s-concave density function where*

$$s = \begin{cases} \frac{t}{1-th}, & \textit{if } 0 < t < 1/h, \\ \infty, & \textit{if } t = \frac{1}{h}. \end{cases}$$

The next proposition is valid for any value of s even though, in this paper, we are interested solely in the case $s = \infty$.

Proposition 2.17 *Let us assume that Q is a probability distribution on $I\!R^d$ which admits an ∞-concave density function. Let $\mathcal{H}$ be a subspace with dimension h. Then, the marginal distribution of Q on $\mathcal{H}$ admits a $1/(d-h)$-concave density function.*

The next corollary is a simple consequence of the fact that the densities of probabilities in $\mathcal{C}$ are ∞-concave and of the application of the two previous propositions.

Corollary 2.18 *Let us assume that $Q \in \mathcal{C}$. Let $\mathcal{H}$ be a subspace with dimension $h < d$. Then, the marginal distribution of Q on $\mathcal{H}$ is $1/d$-concave and admits a $1/(d-h)$-concave density function.*

3 The Results

Let $P \in \mathcal{P}_2$ absolutely continuous with respect to λ_d. We will show that there exists a probability measure $Q \in \mathcal{C}$ satisfying (**??**). In order to prove this result we begin by analyzing the laws which are feasible as a weak limit of a sequence of distributions in $\mathcal{C}$. This is done in the following theorem.

Theorem 3.1 *Let $\{Q_n\}_n$ be a sequence of probabilities in $\mathcal{C}$ that converges weakly to the probability measure Q. Let us assume that their supports are uniformly bounded. Then the support of Q, C, is convex and there exists $x_0 \in I\!R^d$ such that $Q[x_0] = 1$, or*

1. *If the interior of C is empty and if A_C denotes for the affine hull of C, then, Q is absolutely continuous with respect to the Lebesgue measure on A_C and its density is $1/(d-h)$-concave where h is the dimension of A_C.*
2. *If the interior of C is non-empty then Q is uniform on C.*

PROOF.- Let us assume that there exists no $x_0 \in I\!R^d$ such that $Q[x_0] = 1$. We begin by assuming also that the interior of C is empty. If we consider the sequence of the marginal distributions of $\{Q_n\}_n$ on A_C, this sequence converges weakly to the corresponding marginal of Q which, obviously, is Q. Let h be the dimension of A_C.

By Corollary **??**, the marginals of the probabilities in the sequence are $1/d$-concave. From Proposition **??**, Q is also $1/d$ concave and, by Proposition **??**, the support of Q is convex. Therefore, since it generates A_C, the interior of C is not empty in this affine subspace and Proposition **??** implies that Q admits a $1/(d-h)$-concave density function with respect to the Lebesgue measure on A_C.

To end the proof, let us assume that the interior of C is not empty. Let C_n be the support of Q_n, $n \in I\!N$. The sequence $\{C_n\}_n$ is bounded and, from Proposition **??**, we have that there exists a closed and convex set, C_0, which is the Δ-limit of a subsequence $\{C_{n_k}\}_{n_k}$. Moreover, from the definition of the Δ-convergence, we have that the support of Q is contained in C_0. Thus the interior of C_0 is not empty.

Let x_0 be an element of the interior of C_0 and let $0 < \delta_1 < \delta_2$ be such that $B(x_0, \delta_2) \subset C_0$. Let $\gamma := (\delta_2 - \delta_1)/2$.

Let $n_0 \in I\!N$ and let us assume that $B(x_0, \delta_1) \not\subset C_{n_0}$. Thus, $C_{n_0}^\gamma$ does not contain $B(x_0, \delta_1)^\gamma = B(x_0, \delta_1 + \gamma) \subset B(x_0, \delta_2)$. But, by definition of Δ-convergence, from an index onward, $C_0 \subset C_{n_k}^\gamma$ and for those indices, it must be $B(x_0, \delta_1) \subset C_{n_k}$.

From the previous reasoning, we can conclude that if x_0 is a point in the interior of C_0 and $\delta > 0$ satisfies that $B(x_0, \delta) \subset C_0$, then, from an index (which depends, of course, on x_0 and δ) onward we have that $B(x_0, \delta) \subset C_{n_k}$ and then, for those indices, we have that

$$Q_{n_k}[B(x_0, \delta)] = \frac{\lambda_d[B(x_0, \delta)]}{\lambda_d(C_{n_k})} \to \frac{\lambda_d[B(x_0, \delta)]}{\lambda_d(C_0)},$$

where the last convergence comes from Proposition **??**. From here a standard reasoning allows us to conclude that C_0 is the support of Q_0 and that Q_0 is uniform on C_0.

•

Lemma 3.2 *Let $Q \in \mathcal{C}$ and let $v \in I\!R^d$. Then, the marginal of Q on the one-dimensional subspace generated by v, Q^v, is supported by a bounded interval $[a,b]$ and it satisfies*

$$\sigma(Q^v) \geq \gamma(b-a)$$

where γ is a strictly positive constant depending on d, but not on Q.

PROOF.- Without loss of generality we can assume that Q is centered in mean. Obviously, the support of Q^v is a bounded interval $[a,b]$. Let S be the support of Q. Given $x \in [a,b]$ there exists the density of Q^v on x, $f(x)$, and its value is

$$f(x) = \frac{\lambda_{d-1}(S(x))}{\lambda_d(S)}.$$

However, since S is a closed and convex set, its boundary is continuous, and then the map

$$x :\rightarrow L(x) := \lambda_{d-1}(S(x))$$

is continuous on $[a,b]$. Let M be the maximum of L on $[a,b]$ and let $c \in [a,b]$ be such that $L(c) = M$.

Since S is a closed set, there exists $y_a \in S(a)$. Let C_a be the bounded and convex cone with vertex on y_a and basis $S(c)$. Convexity of S implies that $C_a \subset S$. Therefore, if $x \in [a,c]$, then the section of C_a on x, $C_a(x)$, is contained on the section of S on x, $S(x)$, and we have that

$$f(x) \geq \frac{\lambda_{d-1}(C_a(x))}{\lambda_d(S)} = \frac{M}{\lambda_d(S)} \frac{(x-a)^{d-1}}{(c-a)^{d-1}} \geq \frac{(x-a)^{d-1}}{(b-a)^d},$$

where, in the last inequality, we have taken into account that $a \leq c \leq b$ and that $M(b-a)/\lambda_d(S) \geq 1$ because f is a density function.

Using the same argument with the bounded convex cone with vertex on $y_b \in S(b)$ and basis $S(c)$ we have that

$$\sigma^2(Q^v) = \int_a^b x^2 f(x)dx \geq \int_a^c x^2 \frac{(x-a)^{d-1}}{(b-a)^d} dx + \int_c^b x^2 \frac{(b-x)^{d-1}}{(b-a)^d} dx =: I(c).$$

Derivation on c allows us to conclude that the function $I(c)$, as defined above, has an absolute minimum on $c_0 = (a+b)/2$. Moreover, by computing the integrals in the definition of I we have that

$$I(c_0) = \frac{(a+b)^2}{d\,2^{d+1}} + \frac{(b-a)^2}{d(d+1)(d+2)2^d} \geq \frac{(b-a)^2}{d(d+1)(d+2)2^d},$$

which gives the result by taking $\gamma = (d(d+1)(d+2)2^d)^{-1/2}$.

•

In the next lemma we show that if $\{Q_n\}_n$ are centered in mean probabilities in $\mathcal{C}$, such that their Wasserstein distances to P approach the infimum in (**??**), then their support sets are uniformly bounded.

Lemma 3.3 *Let $P \in \mathcal{P}_2$ and $\{Q_n\}_n \subset \mathcal{C}$ centered in mean such that*

$$\lim_n W^2(P, Q_n) = W^2(P, \mathcal{C}). \tag{4}$$

Then, there exists $K > 0$ such that the supports of all probabilities in the sequence are contained in the closed ball $B(0, K)$.

PROOF.- Let $n \in I\!N$ and and $v \in I\!R^d$. Let us consider the one-dimensional subspace generated by v. Let P^v and Q_n^v, respectively, be the marginals of P and Q_n on this subspace. By Proposition **??** we have

$$W^2(P, Q_n) \geq (\sigma(P^v) - \sigma(Q_n^v))^2 \,. \tag{5}$$

Let $K > 0$. If the conclusion in the lemma is not satisfied, since all probabilities are centered in mean, there exists a subsequence of $\{Q_n\}$ such that every distribution in this subsequence admits, at least, a one-dimensional marginal whose support is not contained in an interval whose length is less than K. Thus, from (**??**), (**??**) and Lemma **??** we have that

$$W^2(P, \mathcal{C}) \geq \left(\gamma K - \sup_{v \in I\!R^d} \sigma(P^v) \right)^2 ,$$

which is impossible because $W^2(P, \mathcal{C})$ is finite and K is arbitrary.

•

Proposition 3.4 *Given $P \in \mathcal{P}_2$ centered in mean, there exists a probability measure Q_P such that*

$$W^2(P, Q_P) = W^2(P, \mathcal{C}). \tag{6}$$

Moreover, Q_P is concentrated in one point or it is $1/d$-concave.

PROOF.- Let $\{Q_n\}$ be a sequence in $\mathcal{C}$ such that

$$\lim_n W^2(P, Q_n) = W^2(P, \mathcal{C}), \tag{7}$$

by Proposition **??**, we can assume that the probabilities are centered. Because of Lemma **??**, the supports of the probabilities $\{Q_n\}$ are uniformly bounded. Thus the sequence is tight and there exists a subsequence which converges weakly to a probability measure Q which satisfies the conclusions in Theorem **??**. Therefore if Q is not concentrated in one point, by Proposition **??**, it is $1/d$-concave.

On the other hand, (**??**) and the boundness of the sequence $\{Q_n\}_n$ allows us to apply Proposition **??** to obtain that Q also satisfies (**??**).

•

From the previous proposition, the only remaining task is to show that the support of Q_P in this proposition has a non-empty interior. The proof is carried out in two steps. The first one is more simple and consists of excluding the possibility that Q_P is concentrated in one point. The second one excludes the possibility that Q_P is supported by a strict subspace of $\mathbb{R}^d$.

Proposition 3.5 *Let $P \in \mathcal{P}_2$ centered in mean and absolutely continuous with respect to λ_d. Let $Q = \delta_{\{0\}}$ be the Dirac's delta on 0. Then, there exists a $1/d$-concave distribution Q^* with one dimensional support such that*

$$W^2(P,Q) > W^2(P,Q^*).$$

PROOF.- Let X be a r.v. expressed in terms of some orthonormal basis $e_1, e_2, ..., e_d$ as $X = (X_1, X_2, ..., X_d)$ with distribution P. Let Q_a be the uniform distribution on $[-a, a]$ nearest to the law of X_1. By (**??**), $a > 0$ and

$$\begin{aligned} W^2(P,Q) &= W^2(P,\delta_{\{0\}}) = \sum_{j=1}^{d} W^2(P_{X_j},\delta_{\{0\}}) \\ &> W^2(P_{X_1},Q_a) + \sum_{j=2}^{d} EX_j^2 = W^2(P,Q^*), \end{aligned}$$

where Q^* is the uniform distribution on the convex set $[-a,a] \times \{0\} \times, ... \times \{0\}$, which is $1/d$-concave and the subspace generated by its support is one dimensional.

•

Proposition 3.6 *Let $P, Q \in \mathcal{P}_2$ centered in mean. Assume that P is absolutely continuous with respect to λ_d. Let $\mathcal{H}$ be the subspace generated by the support of Q. Let $1 \leq h < d$ be the dimension of $\mathcal{H}$ and assume that Q is $1/d$-concave. Then there exists $Q^* \in \mathcal{C}$ such that*

$$W^2(P,Q) > W^2(P,Q^*).$$

PROOF.- Let X be a r.v. defined as in Proposition **??**. Without loss of generality, we can assume that $\mathcal{H}$ is the subspace generated by the vectors $e_1, e_2, ..., e_h$ (therefore

$X_{\mathcal{H}} = (X_1, ..., X_h))$. Also let g be the density of the marginal distribution of Q on $\mathcal{H}$. This density does exist by Proposition **??** and is $1/(d-h)$-concave. Let S be the support of Q. Given $\alpha > 0$ and $y \in S$ we will denote $a(\alpha, y) = \alpha[g(y)]^{1/(d-h)}$.

Given $x \in \mathcal{H}$, let P_x^{h+1} be a regular conditional distribution of X_{h+1} given $X_{\mathcal{H}} = x$. By hypothesis, for $P_{(X_1,...,X_h)}$-a.e. $x \in \mathcal{H}$, P_x^{h+1} is absolutely continuous with respect to λ_1. Let us denote by H_x the map obtained by applying Proposition **??** to P_x^{h+1} and the uniform distribution on $[-1, 1]$, Q_1. I.e. $H_x = H_{P_x^{h+1}, Q_1}$.

Given $\alpha > 0$, let

$$C_\alpha := \{(y, z) \in S \times I\!R : |z| \leq a(\alpha, y)\}.$$

Notice that C_α is convex because the $1/(d-h)$-concavity of g implies that if $y_0, y_1 \in S$ and $\theta \in (0, 1)$, then

$$[g(\theta y_0 + (1-\theta)y_1)]^{1/(d-h)} \geq \theta[g(y_0)]^{1/(d-h)} + (1-\theta)[g(y_1)]^{1/(d-h)}.$$

Let Q^α be the distribution on C_α whose density function is

$$g^\alpha(y, z) := \frac{1}{2\alpha}[g(y)]^{(d-h-1)/(d-h)},\ (y, z) \in C_\alpha.$$

Obviously, the subspace generated by the convex supporting Q^α is $(h+1)$-dimensional and a simple computation shows that the density of Q^α is $1/(d-h-1)$-concave. If $Y^\alpha = (Y_1^\alpha, ..., Y_{h+1}^\alpha)$ is a r.v. with distribution Q^α, then the conditional distribution of Y_{h+1}^α given that $(Y_1^\alpha, ..., Y_h^\alpha) = y \in S$ is the uniform law $Q_{a(\alpha,y)}$ and the marginal distribution of $(Y_1^\alpha, ..., Y_h^\alpha)$ is Q.

On the other hand, there exists $H_{P,Q} : I\!R^d \to S$, an o.t.p. between P and Q which satisfies the conclusions in Proposition **??**. We will denote $H := H_{P,Q}$ in order to keep the notation as simple as possible.

Let us consider the r.v.

$$Y^\alpha := (H(X), a(\alpha, H(X))H_{X_{\mathcal{H}}}(X_{h+1}))$$

Taking into account that H satisfies Proposition **??**, we have that the distribution of Y^α is Q^α and

$$\begin{aligned}
\|X - H(X)\|^2 - \|X - Y^\alpha\|^2 &= \|(X_1, X_2, ..., X_h) - H(X)\|^2 + \sum_{j=h+1}^{d} X_j^2 \\
&\quad -\|(X_1, X_2, ..., X_h) - H(X)\|^2 \\
&\quad - [X_{h+1} - a(\alpha, H(X))H_{X_{\mathcal{H}}}(X_{h+1})]^2 - \sum_{j=h+2}^{d} X_j^2 \\
&= X_{h+1}^2 - [X_{h+1} - a(\alpha, H(X))H_{X_{\mathcal{H}}}(X_{h+1})]^2.
\end{aligned}$$

Given $x \in X_{\mathcal{H}}$, let us define

$$\sigma^2_{h+1}(x) := E\left[X^2_{h+1}/X_{\mathcal{H}} = x\right] \qquad \text{and} \qquad H^*(x) := H(x, 0, ..., 0).$$

As a consequence of the construction of Y^α we obtain that

$$\begin{aligned}
W^2(P,Q) - W^2(P,Q^\alpha) &\geq E\|X - H(X)\|^2 - E\|X - Y^\alpha\|^2 \\
&= \int \left(\sigma^2_{h+1}(x) - E\left[\left(X_{h+1} - a(\alpha, H(X))H_{X_{\mathcal{H}}}(X_{h+1})\right)^2 \Big/ X_{\mathcal{H}} = x\right]\right) P_{X_{\mathcal{H}}}(dx) \\
&= \int \sigma^2_{h+1}(x) P_{X_{\mathcal{H}}}(dx) - \int E\left[W^2\left(P^{h+1}_x, Q_{a(\alpha,H^*(x))}\right) \Big/ X_{\mathcal{H}} = x\right] P_{X_{\mathcal{H}}}(dx) \\
&= \int \left(2a(\alpha, H^*(x)C(P^{h+1}_x) - \frac{a^2(\alpha, H^*(x))}{3}\right) P_{X_{\mathcal{H}}}(dx) \qquad (8) \\
&= \alpha \int \left(2g[H^*(x)]^{1/(d-h)}C(P^{h+1}_x) - \frac{\alpha}{3} g[H^*(x)]^{2/(d-h)}\right) P_{X_{\mathcal{H}}}(dx) \qquad (9)
\end{aligned}$$

where (**??**) comes from the application of (**??**) to the involved distributions.

To end the proof, notice that $1/(d-h)$ concavity of g implies that g is bounded. Therefore, taking into account that $C(P^{h+1}_x) > 0$, $P_{X_{\mathcal{H}}}$-a.s. if we take small enough α, from (**??**) we have that

$$W^2(P,Q) - W^2(P,Q^\alpha) > 0,$$

and the proof ends if $h = d-1$ because $Q^\alpha \in \mathcal{C}$. If not, we only need to apply $(n-h-1)$ times the previous reasoning beginning with Q^α and, at the end we will have a distribution Q^* with the desired property.

•

Main result in this paper is an easy consequence of Propositions **??**, **??** and **??**.

Theorem 3.7 *Let $P \in \mathcal{P}_2$ absolutely continuous with respect to λ_d. Then there exists a probability $Q_P \in \mathcal{C}$ such that*

$$W^2(P, Q_P) = W^2(P, \mathcal{C}).$$

Acknowledgment. The authors wish to thank an anonymous referee who pointed out a mistake in the main result in a previous version of this paper and whose comments have contributed considerably to improving the paper.

References

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

[2] BARRIO, E. del, CUESTA-ALBERTOS, J.A. and MATRÁN, C. (2000). Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests (with discussion). *TEST* **9**, 1-96.

[3] BARRIO, E. del, CUESTA-ALBERTOS, J.A., MATRÁN, C. and RODRÍGUEZ-RODRÍGUEZ, J.M. (1999). Tests of goodness of fit based on the L_2-Wasserstein distance. *Ann. Statist.* **27**, 1230-1239.

[4] CSÖRGŐ, S. (2002). Weighted correlation tests for scale families. *Preprint.*

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

[6] CUESTA-ALBERTOS, J.A., MATRÁN, C. and RODRÍGUEZ-RODRÍGUEZ, J.M. (2002). Shape of a distribution through the Wasserstein distance. In *Distributions with given marginals and statistical modeling,* 51-62. Eds. C. M. Cuadras, J. Fortiana and J.A. Rodríguez-Lallena. Kluwer A. Pub.

[7] CUESTA-ALBERTOS, J.A., MATRÁN, C. and TUERO, A. (1996). On lower bounds for the L^2-Wasserstein metric in a Hilbert space. *J. Theoret. Probab.* **9**, 263-283.

[8] CUESTA-ALBERTOS, J.A., MATRÁN, C. and TUERO, A. (1997). Optimal transportation plans and convergence in distribution. *J. Multivariate Analysis.* **60**, 72-83.

[9] de WET, T. (2001). Goodness-of-fit tests for location and scale families based on a weighted L2-Wasserstein distance measure. *Preprint.*

[10] DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity and Applications.* Academic Press: New York.

[11] EGGLESTON, H. (1969) *Convexity.* Cambridge University Press: Cambridge.