On lower bounds for the L^2-Wasserstein metric in a Hilbert space *

J.A. Cuesta-Albertos, C. Matrán-Bea and A. Tuero-Díaz
Universidad de Cantabria and Universidad de Valladolid

Abstract

We provide two families of lower bounds for the L^2-Wasserstein metric in separable Hilbert spaces which depend on the basis chosen for the space. Then we focus on one of these families and we provide a necessary and sufficient condition for the supremum in it to be attained. In the finite dimensional case, we identify the basis which provides the most accurate lower bound in the family.

Key words and phrases: L^2-Wasserstein metric, Lower bound, Gaussian distributions, Hilbert spaces.
A.M.S. 1980 subject classification: 60E05, 62H20.

1 Introduction.

Although the L^2-Wasserstein distance can be defined in more general settings, we limit the scope of this paper to separable real Hilbert spaces. In such spaces the L^2-Wasserstein metric is defined as follows:

Let $(\mathcal{H}; \langle\ ,\ \rangle)$ be a separable real Hilbert space and let μ and ν be Radon probability measures defined on its Borel σ-algebra such that $\int \|x\|^2 d\mu$ and $\int \|x\|^2 d\nu$ are finite. The L^2-Wasserstein distance between μ and ν is defined as the positive square root of

$$W^2(\mu,\nu) := \inf \{ \int \|x-y\|^2 dP(x,y) \}$$

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

where the infimum is taken over the family of probability measures on $\mathcal{H}^2$ with marginals μ and ν respectively.

Some results have appeared in which the value of $W^2(\mu,\nu)$ is computed if $\mathcal{H}$ is finite dimensional and μ and ν are Gaussian (see for instance [6, 8, 10, 11, 12, 13]). It is then well known that if μ and ν are Gaussian distributions, with parameters $(m_\mu,)$ and $(m_\nu,)$ respectively, then

$$W^2(\mu,\nu) = \|m_\mu - m_\nu\|^2 + trace\,(+ - 2)\,.$$

In [7] it is proved that this expression provides a universal lower bound for the Wasserstein distance in the sense that if μ and ν are two probability measures on a Euclidean space and m_μ and m_ν and and are their expectations and covariance matrices respectively, then we have that

$$W^2(\mu,\nu) \geq \|m_\mu - m_\nu\|^2 + trace\,(+ - 2)\,. \tag{1}$$

A sufficient condition for the equality to hold appears in [7]. It is also shown in [7] that the equality is valid in separable Hilbert spaces if μ and ν are Gaussian.

Later, in [14], an incomplete but very simple proof for (1) has been given. Moreover, from that proof it also follows that Gelbrich's sufficient condition is also necessary. These authors work in the finite dimensional case.

In this paper we analize the situation in general separable Hilbert spaces from a different point of view. The basic idea consists of relating the Wasserstein distance between the probabilities to the Wasserstein distances between the marginal probabilities on every complete orthonormal system (orthonormal basis). For every such a basis such a procedure provides not only a lower bound in terms of the variances of the components, but even a better bound based on the Wasserstein distances between the components (Proposition 2.4). This methodology allows to obtain Gelbrich's bound (1) as the best of those lower bounds based on the variances even in the infinite dimensional case (Theorem 2.7). The proof is based on the fact that equality holds for Gaussian distributions (a new simple proof for this fact in the finite dimensional case is included, Theorem 2.5).

The orthonormal basis which provides the best lower bound is obtained in the same theorem if the dimension is finite and we give a counterexample on the existence of such a basis in the general case. Moreover, a better natural bound, the best of those based on marginal distances, BMD, appears. It is interesting to remark that Gelbrich's bound is related to linear dependence, while BMD is related to the structure of dependence (see [6]) of the probability measures. The paper also includes an analysis of the interplay between both lower bounds.

We finalize by giving a necessary and sufficient condition to that (1) becomes an equality in the infinite dimensional case (Theorem 2.15).

On a different line let us consider the right hand side in (1) and suppose that $m_\mu = m_\nu = 0$. Then this right hand side coincides with the Wasserstein distance between two centered Gaussian distributions with covariance operators and respectively. So if Σ_1 and Σ_2 are two self-adjoint, compact operators, then Σ_1^2 and Σ_2^2 are the covariance operators of two Gaussian distributions and the expression

$$d(\Sigma_1, \Sigma_2) = trace^{1/2}\left(\Sigma_1^2 + \Sigma_2^2 - 2\left(\Sigma_1\Sigma_2^2\Sigma_1\right)^{1/2}\right) \tag{2}$$

defines a distance on the family of the self-adjoint, compact linear operators. Moreover in spite of the fact that this distance is related to a norm, d is not a norm. This is because if d arised from a norm, then

$$d(\Sigma_1, \Sigma_2) = d(\Sigma_1 - \Sigma_2, 0)$$

while, by definition,

$$d(\Sigma_1 - \Sigma_2, 0) = trace^{1/2}\left((\Sigma_1 - \Sigma_2)^2\right) = trace^{1/2}\left(\Sigma_1^2 + \Sigma_2^2 - 2(\Sigma_1\Sigma_2)\right).$$

This expression is equal to the right hand side of (2) if Σ_1 and Σ_2 commute, but they differ in general (take for instance $\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$).

On the other hand it is well known (see e.g. [2, pag. 274]) that the expression

$$d^*(\Sigma_1,\Sigma_2) = trace^{1/2}\left(\Sigma_1^2+\Sigma_2^2-2(\Sigma_1\Sigma_2)\right)$$

defines a norm (the so-called Hilbert-Schmidt norm; in fact that norm is induced by an inner product). It may be of interest to note that this expression coincides with the Wasserstein distance between a probability measure concentrated in the point 0 and the centered Gaussian distribution with covariance operator $(\Sigma_1-\Sigma_2)^2$.

Through the paper we will often assume without loss of generality that $m_\mu = m_\nu = 0$ because (see for instance [1]) if μ and ν are two probability measures on $\mathcal{H}$ and μ^* and ν^* are the result of centering them in mean, then

$$W^2(\mu,\nu) = W^2(\mu^*,\nu^*) + \|m_\mu - m_\nu\|^2.$$

2 The results.

We begin with the notation to be employed and with some well-known definitions and properties related to the covariance operators in a Hilbert space. We state them for sake of completeness.

Let $(\mathcal{H},\langle\,,\,\rangle)$ be a separable real Hilbert space and let μ be a Radon probability measure (i.e. $\mu(B) = \sup\{\mu(K) : K \subset B,\, K \text{ compact}\}$) defined on the Borel σ-algebra. Let us suppose that μ is of strong order two (i.e. $\int \|x\|^2 d\mu < \infty$). Under these conditions the mean value of μ, m_μ, and the covariance operator of μ, , can be defined through the relations:

$$\langle m_\mu, a\rangle = \int \langle x, a\rangle \mu(dx)\,;\; a\in\mathcal{H}$$

and

$$\langle a, b\rangle = \int \langle x-m_\mu, a\rangle\langle x-m_\mu, b\rangle \mu(dx)\,;\; a,b\in\mathcal{H}.$$

is positive (i.e. $\langle a,a\rangle \geq 0\,;\, a\in\mathcal{H}$), self-adjoint (i.e., $\langle a,b\rangle = \langle a,b\rangle\,;\, a,b\in\mathcal{H}$), compact (i.e. maps every bounded set in $\mathcal{H}$ onto a relatively compact set in $\mathcal{H}$) and has finite trace (for definitions, properties and related facts see e.g. [9, 16]).

On the other hand, if Σ is a compact positive self-adjoint linear operator, there exists an orthonormal basis for $\mathcal{H}$, $\{\}$, consisting of eigenvectors of Σ whose associated eigenvalues are nonnegative [2, p. 48]. Therefore, if we denote by λ_n the eigenvalue associated with the eigenvector , we can define the linear operator $\Sigma^{1/2}$ by

$$\Sigma^{1/2}\vec{e}_n = \lambda_n^{1/2}\vec{e}_n$$

and, if we fix in $\mathcal{H}$ the basis given by the eigenvectors $\{\}$ and we consider $\mathcal{H}$ as a space of sequences, we could define Σ^- in an obvious way from the relations

$$\Sigma^-\vec{e}_n = \left| \begin{array}{ll} \lambda_n^{-1}\vec{e}_n, & \text{if } \lambda_n \neq 0 \\ 0, & \text{if } \lambda_n = 0 \end{array} \right. .$$

The basic result in this section is proved in two steps. We first solve the problem in the finite-dimensional case and then we use this partial result to obtain the general one.

Therefore we set $\mathcal{H} = \Re^k$ and let μ and ν be two probability measures with finite second order moment and covariance matrices and . Let us consider the matrix

$$A(,) := \tag{3}$$

(which defines a linear, self-adjoint, positive semidefinite operator). According to Theorem 2.13 in [6], if X is a $\Re^k$-valued random vector, then $\|X - AX\|_2 = W(P_X, P_{AX})$, where $A := A(,)$.

Unless the contrary is stated, while we are in the case $\mathcal{H} = \Re^k$, we represent by $\{\vec{e}_n\}$ an orthonormal basis of $\Re^k$ of eigenvectors of the linear operator $A(,)$ such that the basis of $KerA(,)$ contains a basis of Ker (this is possible because $Ker \subset KerA(,)$). Therefore, we have that $\vec{e}_n \in Ker$ or $\vec{e}_n \in (Ker)^\perp$ for every n.

In that follows, given the linear operator Σ, Π_Σ denotes the projection onto the subspace $(Ker\Sigma)^\perp$. Note the basic property relating the operators Π_Σ and Σ^-:

$$\Pi_\Sigma\vec{x} = \Sigma^-\Sigma\vec{x} = \Sigma\Sigma^-\vec{x}, \; \forall\vec{x} \in \mathcal{H}.$$

Lemma 2.1 *Let $\{\lambda_n\}$ be the eigenvalues of $A := A(,)$ associated with the eigenvectors $\{\vec{e}_n\}$. Then:*

$$trace = \sum_n \lambda_n \langle \vec{e}_n, \vec{e}_n \rangle.$$

PROOF.- As stated we have two possibilities:

If $\vec{e}_n \in Ker = Ker$, then $\vec{e}_n \in Ker$ and

$$\langle \vec{e}_n, \vec{e}_n \rangle = \langle A\vec{e}_n, \vec{e}_n \rangle = 0.$$

If $\vec{e}_n \in (Ker)^{\perp}$ then

$$\begin{aligned} \langle A, \rangle &= \langle \Pi_{\Sigma_\mu}, \rangle \\ &= \langle , \Pi_{\Sigma_\mu} \rangle \ = \ \langle , \rangle \end{aligned}$$

and we have that

$$\begin{aligned} trace &= trace\,(A) = trace(A) \\ &= \textstyle\sum_i \langle A, \rangle = \sum_i \lambda_i \langle , \rangle. \end{aligned}$$

The key ideas in the paper arise in the (proof of the) following theorem.

Theorem 2.2 *With μ and ν as above we have that*

$$W^2(\mu, \nu) \geq \|m_\mu - m_\nu\|^2 + trace(+ - 2).$$

The equality holds if and only if $\nu^* \circ \Pi_{\Sigma_\mu}^{-1} = \mu^* \circ A^{-1}$, where μ^* and ν^* are the centered in mean measures and A:=$A(,)$.

PROOF.- Without loss of generality let us assume that $m_\mu = m_\nu = 0$.

Let X and Y be two random vectors such that $P_X = \mu$, $P_Y = \nu$ and $W(\mu, \nu) = \|X - Y\|_2$. If we denote $X_i := \langle X, \rangle$, we have that

$$W^2(\mu, \nu) = \|X - Y\|_2^2 = \sum_i E|X_i - Y_i|^2.$$

Let $\{\hat{X}_i\}$ and $\{\hat{Y}_i\}$ be real r.v.'s. such that

$$\hat{X}_i \stackrel{d}{=} X_i,\ \hat{Y}_i \stackrel{d}{=} Y_i, \text{ and } W(P_{X_i}; P_{Y_i}) = \|\hat{X}_i - \hat{Y}_i\|_2.$$

Then we have the following inequalities

$$
\begin{aligned}
W^2(\mu,\nu) &= \textstyle\sum_i E|X_i - Y_i|^2 \geq \sum_i W^2(P_{X_i}; P_{Y_i}) = \sum_i E|\hat{X}_i - \hat{Y}_i|^2 \\
&\geq \textstyle\sum_i \left| \|\hat{X}_i\|_2 - \|\hat{Y}_i\|_2 \right|^2 = \sum_i |\sigma(X_i) - \sigma(Y_i)|^2 \\
&= \textstyle\sum_i \left| \langle , \rangle^{1/2} - \langle , \rangle^{1/2} \right|^2 \\
&= trace\,(+) - 2 \textstyle\sum_i \langle , \rangle^{1/2} \langle , \rangle^{1/2},
\end{aligned}
\tag{4}
$$

where $\sigma(X_i)$ is the standard deviation of X_i, $i = 1, ..., k$.

Now note that $Ker \subset Ker$. Therefore, if $x \in \Re^k$, we have that $x \in (Ker)^{\perp}$ and we obtain that

$$\Sigma_{AX} = AA == \Pi_{\Sigma_\mu} \Pi_{\Sigma_\mu}$$

and then

$$
\begin{aligned}
\langle , \rangle &= \langle \Pi_{\Sigma_\mu}, \Pi_{\Sigma_\mu} \rangle \; = \; \langle \Pi_{\Sigma_\mu} \Pi_{\Sigma_\mu}, \rangle \\
&= \langle AA, \rangle \; = \; \lambda_i \langle A, \rangle \; = \; \lambda_i^2 \langle , \rangle
\end{aligned}
$$

if $\in (Ker)^{\perp}$.

Therefore we have proved that

$$W^2(\mu,\nu) \geq trace(+) - 2 \sum_i \lambda_i \langle , \rangle.$$

By Lemma 2.1 the inequality in the theorem is proved.

To end the proof note that the first inequality in (4) is an equality if and only if $W^2(P_{X_i}, P_{Y_i}) = E|X_i - Y_i|^2$, i.e. if and only if we can assume that $X_i = \hat{X}_i$ and $Y_i = \hat{Y}_i$ a.s. and the second inequality in (4) is an equality if and only if $\hat{X}_i$ and $\hat{Y}_i$ are linearly related a.s.. But this implies that, if $\in (Ker)^{\perp}$ then $\hat{Y}_i = \alpha_i \hat{X}_i$ a.s. for some constant α_i. In that case $\alpha_i = \lambda_i$ and the second statement is also proved.

Next we extend Theorem 2.2 to the infinite dimensional case. This extension is carried out in several steps. First we show that inequalities in (4) hold in the infinite dimensional case. We will use Theorem 2.2 in our proof. To do this, we fix an orthonormal basis $\{\}$ in $\mathcal{H}$ and we denote by Π_n the projection on the linear subspace spanned by the vectors $\vec{e}_1, ...,$ and $\mu_n = \mu \circ \Pi_n^{-1}$ and $\nu_n = \nu \circ \Pi_n^{-1}$.

We also employ the following lemma which was proved for general Banach spaces in [1, Lemma 8.3].

Lemma 2.3 *Let $\{\mu_i\}$ and μ be strong order two probability measures defined on $\mathcal{H}$. Then the following statements are equivalent*

1. *$W(\mu_i, \mu) \to 0$.*
2. *$\mu_i \to \mu$ weakly and $\int \|x\|^2 d\mu_i \to \int \|x\|^2 d\mu$.*

Proposition 2.4 *Let μ and ν be two Radon probability measures of strong order two defined on $\mathcal{H}$. Let X and Y be two random vectors with distribution μ and ν respectively. Let $\{\vec{e}_n\}$ be an orthonormal basis on $\mathcal{H}$ and let us define $X_i := \langle X, \vec{e}_i\rangle$ and $Y_i := \langle Y, \vec{e}_i\rangle$. Then we have the following inequalities*

$$\begin{aligned} W^2(\mu,\nu) &\geq \sum_i W^2(P_{X_i}, P_{Y_i}) && (5)\\ &\geq \|m_\mu - m_\nu\|^2 + \sum_i \sigma^2(X_i) + \sum_i \sigma^2(Y_i) - 2\sum_i \sigma(X_i)\sigma(Y_i) && (6) \end{aligned}$$

where $\sigma(X_i)$ and $\sigma(Y_i)$ denote the standard deviations of the real random variables X_i and Y_i respectively.

PROOF.- We can assume without loss of generality that $m_\mu = m_\nu = 0$.

Since $\|\Pi_n X - X\|_2 \to 0$, we can apply Lemma 2.3 to μ_n to get that $\{W(\mu_n, \mu)\}$ converges to zero, so that $\lim_n W(\mu_n, \nu_n) = W(\mu, \nu)$.

On the other hand, inequalities in (4) hold for every orthonormal basis, therefore, if we apply them to μ_n and ν_n, we have that

$$\begin{aligned} W^2(\mu,\nu) &= \lim_n W^2(\mu_n,\nu_n) \geq \lim_n \textstyle\sum_{i\leq n} W^2(P_{X_i}, P_{Y_i})\\ &\geq \lim_n \textstyle\sum_{i\leq n} |\sigma(X_i) - \sigma(Y_i)|^2, \end{aligned}$$

and the proof ends.

REMARKS

1. Inequality (5) is an equality if and only if μ and ν have the same structure of dependence on the basis $\{\vec{e}_n\}$ (see [6]).

2. Inequality (6) is an equality if and only if the marginal distributions on the basis $\{\vec{e}_n\}$ are affine positively linearly related.

3. The same kind of proof of the Theorem 2.2 allows to improve the lower bound given by (5) by considering marginal distributions of higher order. For example, let us assume that k is an even number and let P_j (resp. Q_j) be the two-dimensional marginal probability of μ (resp. ν) on the subspace generated by $\{\vec{e}_{2j-1}, \vec{e}_{2j}\}$ $j = 1, 2, ...$, then

$$W^2(\mu, \nu) \geq \sum_j W^2(P_j, Q_j) \geq \sum_i W^2(P_{X_i}, P_{Y_i}).$$

   (It is also possible consider bounds based on marginal distributions on orthogonal subspaces V_j of different dimensions with the only restriction $H = \bigoplus V_j$).

4. The best bounds which can be obtained from (5) and (6) are obviously

$$\begin{aligned} W^2(\mu, \nu) &\geq \sup \sum_i W^2(P_{X_i}, P_{Y_i}) \\ &\geq \sup \left(\|m_\mu - m_\nu\|^2 + \sum_i \sigma^2(X_i) + \sum_i \sigma^2(Y_i) - 2 \sum_i \sigma(X_i)\sigma(Y_i) \right) \end{aligned}$$

   where both suprema are taken over the different orthonormal basis of $\mathcal{H}$. Indirect considerations (related with the second bound) show that the more precise bound, BMD, cannot be generally attained in a linear way.

Note that the lower bounds that inequality (6) gives do not depend on the particular random vector or even on the probability distributions we choose. They only depend on and . Therefore, given the orthonormal basis $B := \{\vec{e}_n\}$, let us to denote by $Low(B, ,)$ the lower bound obtained in (6) when we choose B as a orthonormal basis of $\mathcal{H}$.

All these lower bounds do not coincide in general. For instance let us suppose that $\mathcal{H} = \Re^2$ and that

$$= \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then, by Proposition 2.4 we have that $W^2(\mu, \nu) \geq 4$; but if we consider the basis

$$B = \{(1/\sqrt{2}, 1/\sqrt{2}), (1/\sqrt{2}, -1/\sqrt{2})\}$$

then

$$= \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

and $Low\,(B, ,) = 0$.

This suggests the question of the choice of the best lower bound and the identification of the basis which produces it. Both questions will be solved in the finite dimensional setting by analyzing the Gaussian case. We are going to prove that if μ and ν are two Gaussian distributions and we choose the basis we handled in the proof of Theorem 2.2 then equality holds. Therefore the lower bound obtained through the orthonormal basis of eigenvectors of $A(,)$ has to be the best one.

Theorem 2.5 *Let and be two positive, symmetric matrices on $\Re^k$. Let B_0 be an orthonormal basis of eigenvalues of $A(,)$ which contains an orthonormal basis of Ker. Then*

$$Low(B, ,) \leq Low(B_0, ,).$$

Moreover, if μ and ν are two Gaussian distributions with covariance matrices and respectively, then

$$W^2(\mu, \nu) = Low(B_0, ,).$$

PROOF.- Let μ and ν be two Gaussian probability measures with and as covariance matrices respectively. Taking into account that $\Sigma_{AX} = \Pi_{\Sigma_\mu} \Pi_{\Sigma_\mu}$ we obtain that Theorem 2.2 gives an equality for μ and ν. However, from the reasoning in that theorem we also have that

$$\|m_\mu - m_\nu\|^2 + trace(+ - 2) = Low(B_0, ,)$$

what joined to Proposition 2.4 gives us the result.

To better explain the different roles played by Gelbrich's bound and BMD let us consider in $\Re^2$ the probability μ given by the standard bidimensional normal law $N_2(0, I_2)$ and ν a nonnormal distribution with uncorrelated standard normal, $N(0,1)$, marginal laws (in a given basis). Then Gelbrich's bound is zero while BMD is strictly greater because some one-dimensional marginal distribution of ν in some basis must be nonnormal. However, the last term in inequality (6) is the same for each orthonormal basis. Therefore it could be still argued that BMD will be attained in a basis where Gelbrich's bound is also attained. The following simple example shows a new situation where Gelbrich's bound is useless while BMD coincides with the Wasserstein distance. Moreover the BMD is attained for an orthonormal basis in which Gelbrich's bound is not attained.

Example 2.6 For a given orthonormal basis in $\Re^2$ let μ be the uniform distribution on the triangle surface determined by the points (0,0), (1,1) and (1,-1). Now consider a random vector $X = (X_1, X_2)$ with distribution μ and let ν be the distribution of the random vector $Y = (X_1^2, (X_2+1)^2)$. Clearly:

$$\begin{aligned} W^2\left(P_{X_1}, P_{X_1^2}\right) + W^2\left(P_{X_2}, P_{(X_2+1)^2}\right) &\leq W^2(\mu,\nu) \leq E\|X-Y\|^2 \\ &= E|X_1 - X_1^2|^2 + E|X_2 - (X_2+1)^2|^2. \end{aligned} \tag{7}$$

Moreover, taking into account that the mappings $\lambda \mapsto \lambda^2$ and $\lambda \mapsto (\lambda+1)^2$ defined on the interval $[-1,\infty)$ are increasing, we obtain as a consequence of Corollary 2.9 in [17] and Corollary 5.2, ii) in [3], that:

$$W^2\left(P_{X_1}, P_{X_1^2}\right) = E|X_1 - X_1^2|^2$$

$$W^2\left(P_{X_2}, P_{(X_2+1)^2}\right) = E|X_2 - (X_2+1)^2|^2.$$

It then follows that the inequalities in (7) are equalities and we obtain that

$$W^2(\mu,\nu) = W^2\left(P_{X_1}, P_{X_1^2}\right) + W^2\left(P_{X_2}, P_{(X_2+1)^2}\right) = 48/30.$$

On the other hand, simple computations show that:

$$\Sigma_\mu = \frac{1}{18}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}, \Sigma_\nu = \frac{1}{180}\begin{pmatrix} 15 & 5 \\ 5 & 1 \end{pmatrix}$$

and

$$\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2} = \frac{1}{3240}\begin{pmatrix} 15 & 5\sqrt{3} \\ 5\sqrt{3} & 381 \end{pmatrix}$$

so that the matrix $A(,)$ is not diagonal because in other case $\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}$ would be diagonal.

Finally, numerical computations show that $trace\left(\Sigma_\mu + \Sigma_\nu - 2(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\right)$ is approximately 0,189941 (computations realized with MATHEMATICA with twenty digits accuracy).

Next we show that Gelbrich's bound (1) also works in the infinite dimensional case. At the same time we obtain the relation between this bound and those bounds given in (5) and (6).

We employ the following additional notation. Given the strong order two Radon probability measure μ we denote by G_μ to the Gaussian probability measure with the same mean and covariance operator as μ (recall that $\mu_n = \mu \circ \Pi_n^{-1}$).

It is interesting to note that it could be possible that the optimal basis for $G_{\mu_{n+1}}$ and $G_{\nu_{n+1}}$ does not contain the corresponding optimal one for G_{μ_n} and G_{ν_n}. Therefore some caution is needed to carry out the extension we are looking for.

Theorem 2.7 *Let μ and ν be two Radon probability measures of strong order two defined on $\mathcal{H}$. Then we have that*

$$\begin{aligned} W^2(\mu,\nu) &\geq sup \sum_i W^2(P_{X_i}, P_{Y_i}) \\ &\geq \|m_\mu - m_\nu\|^2 + trace(+-2) \\ &= W^2(G_\mu, G_\nu) \\ &= sup \left(\|m_\mu - m_\nu\|^2 + \sum_i \sigma^2(X_i) + \sum_i \sigma^2(Y_i) - 2\sum_i \sigma(X_i)\sigma(Y_i)\right), \end{aligned}$$

where both supremums are taken on all possible orthonormal basis and X_i, Y_i, $i = 1, 2, ...$ are real random variables with the same distribution as the corresponding marginals of μ and ν respectively.

PROOF.- We are going to use the fact (proved in [7, Theorem 3.5]) that, even in the infinite dimensional case, if $m_\mu = m_\nu = 0$, then

$$W^2(G_\mu, G_\nu) = trace(+ - 2). \qquad (8)$$

On the other hand it is obvious that $G_{\mu_n} = G_\mu \circ \Pi_n^{-1}$. So, if we apply the same reasoning as in Proposition 2.4 we have that

$$\begin{aligned} W^2(\mu,\nu) \ &\geq \ \lim_n \textstyle\sum_{i\leq n} W^2(P_{X_i}, P_{Y_i}) \\ &\geq \ \lim_n W^2(G_{\mu_n}, G_{\nu_n}) \geq \ \lim_n \textstyle\sum_{i\leq n} |\sigma(X_i) - \sigma(Y_i)|^2. \end{aligned}$$

From here, by taking the supremum on all orthonormal basis on $\mathcal{H}$, we obtain that

$$\begin{aligned} W^2(\mu,\nu) \ &\geq \sup \ \textstyle\sum_i W^2(P_{X_i}, P_{Y_i}) \geq W^2(G_\mu, G_\nu) \\ &\geq \sup \ \left(\textstyle\sum_i \sigma^2(X_i) + \sum_i \sigma^2(Y_i) - 2\sum_i \sigma(X_i)\sigma(Y_i)\right). \end{aligned}$$

Now we only have to prove that the third inequality is, in fact, an equality. To do this, let us fix, once more, an orthonormal basis, $\{\}$, on $\mathcal{H}$. Let $\epsilon > 0$ and let k be a natural number such that $W^2(G_\mu, G_\nu) - \epsilon \leq W^2(G_{\mu_k}, G_{\nu_k})$. If we apply Theorem 2.5 to the distributions G_{μ_k} and G_{ν_k} we have that there exist k orthonormal vectors, $\vec{e}_1^{\,k}, ..., \vec{e}_k^{\,k}$, which span the same subspace as $\vec{e}_1, ..., \vec{e}_k$ such that if we denote $X_i^k =< X, \vec{e}_i^{\,k} >$ and $Y_i^k =< Y, \vec{e}_i^{\,k} >$, $i = 1, ..., k$, where X and Y are two random elements with distributions μ and ν, then

$$W^2(G_{\mu_k}, G_{\nu_k}) = \sum_{i\leq k} |\sigma(X_i^k) - \sigma(Y_i^k)|^2$$

and, if we consider the orthonormal basis of $\mathcal{H}$ given by $\{\vec{e}_1^{\,k}, ..., \vec{e}_k^{\,k}, \vec{e}_{k+1}, \vec{e}_{k+2}, ...\}$ we obtain that

$$\begin{aligned} W^2(G_\mu, G_\nu) - \epsilon \leq \ & W^2(G_{\mu_k}, G_{\nu_k}) = \textstyle\sum_{i\leq k} |\sigma(X_i^k) - \sigma(Y_i^k)|^2 \\ \leq \ & \textstyle\sum_{i=1}^\infty |\sigma(X_i^k) - \sigma(Y_i^k)|^2 \\ \leq \ & \sup \ \left(\textstyle\sum_i \sigma^2(X_i) + \sum_i \sigma^2(Y_i) - 2\sum_i \sigma(X_i)\sigma(Y_i)\right), \end{aligned}$$

where X_i^k and Y_i^k are defined in an obvious way when $i \geq k$, and the theorem is proved.

Now, we will study the conditions for the equality in Theorem 2.7 to hold in the infinite dimensional case. This presents some difficulties. For instance, consider the conditions to have equality in Theorem 2.2. The first question is that now the operator $A(,)$ is not well defined through the expression (3) because if we consider in $\mathcal{H}$ an orthonormal basis of eigenvectors of and X is a random element with Gaussian distribution μ then the marginal distributions of X are independent standard Gaussian ones which has sense if we consider $\mathcal{H}$ as an espace of sequences, while $()^{-}X$ does not belong to the Hilbert space $\mathcal{H}$ μ-a.s., so X is not well defined.

However we can get an almost surely defined operator with the desired properties through the following methodology.

Proposition 2.8 *Let and be covariance operators. Let X be a random element with covariance operator . Let $\{\}$ be an orthonormal basis consisting of eigenvectors of and let $\{\}$ be the corresponding eigenvalues. Let $L = \{i \in \mathcal{N} \ : \notin Ker\}$ and suppose that $\geq$ if $i, j \in L$ and $i < j$.*

Then there exist real random variables $\{Z_i; i \in L\}$ such that

$$\frac{\sum_{\substack{j\in L\\ 1\leq j\leq k}} \langle X, \rangle \langle\ ,\ \rangle}{\ } \xrightarrow[k\to\infty]{a.e.} Z_i,\ i \in L, \tag{9}$$

and

$$\sum_{i\in L} Z_i^2 < \infty \ a.e.$$

PROOF.- Note that the assumption $\geq$ if $i, j \in L$ and $i < j$ is right because $\sum < \infty$.

Let $i \in L$. For each $j \in L$ let us define

$$Y_j = \frac{\langle X, \rangle \langle\ ,\ \rangle}{\ }.$$

The sequence $\{Y_j\}_{j\in L}$ is orthogonal (i.e. $EY_j^2 < \infty$ for all $j \in L$ and $EY_kY_j = 0$ for all $k \neq j$), so to obtain the convergence in (9) it suffices to prove that $\sum_{j\in L} jE(Y_j^2) < \infty$ (see, e.g.,Corollary 2.2.1 in [15]).

But, we have:

$$\sum_{j\in L} \frac{1}{}E(Y_j^2) = \sum_{j\in L} \frac{1}{}E\frac{\langle X, \rangle^2\langle\ ,\ \rangle^2}{}$$
$$= \frac{1}{}\sum_{j\in L}\frac{1}{}\langle\ ,\ \rangle^2$$
$$\leq \frac{1}{}\sum_{j\in L}\frac{1}{}\|\|^2$$
$$= \frac{1}{}\sum_{j\in L}\frac{1}{}\langle\ ,\ \rangle$$
$$= \frac{1}{}\sum_{j\in L}\frac{1}{}\langle\left(^{1/2 1/2}\right)\ ,\ \rangle = \frac{1}{}\sum_{j\in L}\langle,\rangle < \infty$$

because has finite trace.

Now, the convergence of the series $\sum \lambda_i$ implies that, if $\epsilon > 0$ then there exists $N_0 \in \mathcal{N}$ such that, for every $m \geq N_0$ we have that

$$\epsilon > \lambda_{N_0} + \ldots + \lambda_m \geq (m - N_0)\lambda_m$$

and we have that $\lim_m m\lambda_m = 0$. Therefore we conclude that $\frac{1}{} >$ n if $n \geq n_0$ and $\sum_{j\in L} jE(Y_j^2) < \infty$.

To prove the convergence of the series $\sum_{i\in L} Z_i^2$ we will see that $\sum_{i\in L} E(Z_i^2) < \infty$. Clearly this condition would be satisfied if $E(Z_i^2) \leq$ $\langle,\rangle$; but,

$$E(Z_i^2) \leq \liminf_k E\left(\sum_{\substack{j\in L\\ 1\leq j\leq k}} Y_j\right)^2$$
$$= \liminf_k \sum_{\substack{j\in L\\ 1\leq j\leq k}} \frac{\langle\ ,\ \rangle^2}{}$$
$$= \frac{1}{}\sum_{j\in L}\langle\ ,\ \rangle^2$$
$$= \frac{1}{}\sum_{j\in L}\langle\ ,\ \rangle\langle\ ,\ \rangle$$

$$= \frac{1}{-}\langle\ ,\rangle$$
$$= \frac{1}{-}\langle\left(^{1/21/2}\right)\ ,\ \rangle \ = \ \langle,\rangle.$$

This completes the proof of this proposition.

In view of Proposition 2.8 there exists a set Ω_0 with $P_X(\Omega_0) = 1$ where we can define an $\mathcal{H}$-valued map $T\,(= T(,))$ by :

$$T(x) := \sum_{i\in L}\left(\sum_{j\in L}\frac{\langle x,\rangle\langle\ ,\ \rangle}{}\right). \tag{10}$$

Note that in the finite dimensional case T coincides with the operator in (3). On the other hand, in any situation it is easy to obtain that T is self-adjoint and linear in its domain.

We have to face another difficulty yet. To analize it let us consider the following proposition, whose proof appears in [5].

Proposition 2.9 *Let μ and ν be strong order two probability measures and assume that for the probability μ there exists an orthogonal basis, $\{\}$, such that, for each n and almost everywhere ω in the orthogonal subspace to , the conditional distribution function on the subspace generated by given ω is atomless.*

Then if $W(\mu,\nu) = \|X - Y_1\|_2 = \|X - Y_2\|_2$ we have that $Y_1 = Y_2$ a.s.

Now let B be the self-adjoint positive linear operator defined on the Hilbert space $L^2[0,1]$ by

$$B[f(x)] = xf(x),\ f \in L^2[0,1],\ x \in [0,1].$$

According to Theorem 2.13 in [6] we have that $W(P_X, P_{BX}) = \|X - BX\|_2$. Moreover, if μ is a non-singular Gaussian distribution then it verifies the conditions in Proposition 2.9 [4, Proposition 2.11]. Therefore, if X has a Gaussian distribution, by Proposition 2.9, B is the only map F which verifies that $W(P_X, P_{BX}) = \|X - F(X)\|_2$. But B has no eigenvector and therefore one of the key ideas of the proof of Theorem 2.2 (to consider the basis where the linear operator A is diagonal) is not valid now.

Moreover, in this case, for every basis we fix, if we denote $Y = BX$, we have that

$$trace(+-2) > \sum_i \sigma^2(X_i) + \sum_i \sigma^2(Y_i) - 2\sum_i \sigma(X_i)\sigma(Y_i),$$

because, if this equation were an equality for any basis, then by the reasoning in Theorem 2.2 we would have that $Y_i = \lambda_i X_i$ almost surely in this basis and then B would have eigenvalues

However, as announced, we can obtain the following result.

Theorem 2.10 *Let and be covariance operators. If X is a random element with covariance operator , mean vector $m_X = 0$ and T is the map given by (10), then the covariance operator of T(X) is* $\Pi\Pi$, *where* Π *denotes the projection on the subspace* $(Ker\Sigma_1)^{\perp}$, *and*

$$W^2(P_X, P_{T(X)}) = E\|X - T(X)\|^2 = trace(+\Pi\Pi - 2).$$

PROOF.- The covariance operator $\Sigma_{T(X)}$ associated with the probability measure $P_{T(X)}$ is linear and bounded, so to prove $\Sigma_{T(X)} = \Pi\Pi$ it suffices to show that $\langle \Sigma_{T(X)}, \rangle = \langle \Pi\Pi, \rangle$ for every i, j.

Taking into account that $Ker \subset Ker$ we have that, if $i, j \in L$:

$$\begin{aligned}
\langle \Sigma_{T(X)}, \rangle &= \int \langle x, \rangle \langle x, \rangle P_{T(X)}(dx) \;=\; \int \langle T(y), \rangle \langle T(y), \rangle P_X(dy) \\
&= \int \langle y, T() \rangle \langle y, T() \rangle P_X(dy) \;=\; \langle T, T \rangle \\
&= \sum_{k \in L} \frac{\langle , \vec{e}_k \rangle \langle , \vec{e}_k \rangle}{} \\
&= \frac{\langle , \rangle}{} \\
&= \langle , \rangle \;=\; \langle \Pi\Pi, \rangle.
\end{aligned}$$

The equality is evident if i or j does not belong to L because then

$$\langle \Sigma_{T(X)}, \rangle = \;0\; = \langle \Pi\Pi, \rangle.$$

Consider now

$$\begin{aligned}
E\|X - T(X)\|^2 &= \sum_i E(\langle X, \rangle - \langle T(X), \rangle)^2 \\
&= trace(+\Pi\Pi) - 2\sum_i E\langle X, \rangle \langle T(X), \rangle
\end{aligned}$$

and

$$\begin{aligned} E\langle X, \rangle\langle T(X), \rangle &= E\langle X, \rangle\langle X, T()\rangle = \langle , T()\rangle \\ &= \langle T(), \rangle = \langle , \rangle. \end{aligned}$$

Consequently:

$$E\|X - T(X)\|^2 = trace(+\Pi\Pi - 2)$$

and the theorem is proved.

REMARK

Let and be covariance operators. If X is a Gaussian random variable with covariance operator and mean vector $m_X = 0$ then T(X) is a Gaussian random variable with covariance operator $\Pi\Pi$, because if $v \in \mathcal{H}$, then the real-valued random variable

$$\langle T(X), v\rangle = \sum_{i\in L}\langle v, \rangle \left(\sum_{j\in L} \frac{\langle X, \rangle\langle\ ,\ \rangle}{\ } \right)$$

is the a.e. limit of normal random variables and, in consequence,

$$W^2(P_X, P_{T(X)}) = E\|X - T(X)\|^2 = trace(+\Pi\Pi - 2).$$

We will need the next result, which has independent interest.

Proposition 2.11 *Let μ be a strong order two probability measure which verifies the condition in the Proposition 2.9. Let Σ be a covariance operator. Consider the set $\mathcal{P}_\Sigma$ of all strong order two probability measures whose covariance operator is Σ and define*

$$W^2(\mu, \mathcal{P}_\Sigma) = inf_{\nu\in\mathcal{P}_\Sigma} W^2(\mu, \nu).$$

Then there exists a unique probability measure ν^* such that

$$W^2(\mu, \mathcal{P}_\Sigma) = W^2(\mu, \nu^*).$$

PROOF.- Let X be a random element such that $P_X = \mu$. The hyphoteses imply that $Ker\Sigma_\mu = \{0\}$. So if we consider the mapping $T := T(,)$ then (see Theorem 2.10) $\Sigma_{T(X)} = \Sigma$ and $T(X) \in \mathcal{P}_\Sigma$.

Theorems 2.7 and 2.10 imply that $W^2(\mu, P_{T(X)}) = W^2(\mu, \mathcal{P}_\Sigma)$ and Theorem 3.1 in [5] gives us that $P_{T(X)}$ is the only probability whith this property.

Lemma 2.12 *Let X be a random element. Let G be a Gaussian random element and let B be a Bernoulli random variable with parameter p in (0,1) such that X, G and B are mutually independent.*

Then the random element $Z := B \times X + (1 - B) \times G$ satisfies the condition in Proposition 2.9

PROOF.- Let $\{\}$ be an orthogonal basis in $\mathcal{H}$ of eigenvectors of the covariance operator of G. Let us denote by Z^i, X^i, G^i, $i \in N$, the marginal components in this basis of the r.e.'s Z, X and G respectively.

Let $x \in \Re$ and $n \in N$. We have to show that

$$P[Z^n = x/Z^i,\, i \neq n] = 0.$$

First note that

$$P[Z^n = x/Z^i,\, i \neq n] = E\left[E[I_{\{x\}}(Z^n)/B, X^n, Z^i,\, i \neq n]/Z^i,\, i \neq n\right].$$

Now take into account that, in the basis we are considering, the marginal r.v.'s of G are mutually independent. This and the independence assumption imply that the conditional distribution of Z^n given $(B, X^n, Z^i,\, i \neq n)$, is an one-dimensional Gaussian and then

$$P[Z^n = x/B, X^n, Z^i,\, i \neq n] = 0.$$

Next we show the necessary condition under the assumption that Σ_μ is non singular.

Proposition 2.13 *Let μ and ν be probability distributions of strong order two such that*

$$W^2(\mu, \nu) = \|m_\mu - m_\nu\|^2 + trace(+ - 2),$$

and assume that $Ker = \{0\}$.

Then there exists a transformation, T, defined μ-a.s. linear in its domain, such that if X is a random element with $P_X = \mu$ then $P_{TX} = \nu$.

PROOF.- As usual, we assume to simplify the notation, that $m_\mu = m_\nu = 0$.

Let us first assume, that μ satisfies the condition in the Proposition 2.9. As stated in that proposition, there exists a unique probability measure ν^* in $\mathcal{P}_{\Sigma_\nu}$ such that

$$W^2(\mu, \mathcal{P}_{\Sigma_\nu}) = W^2(\mu, \nu^*).$$

On the other hand, Gelbrich's bound only depends on the covariance operators we are considering, and therefore

$$W^2(\mu, \mathcal{P}_{\Sigma_\nu}) \geq trace(+ - 2).$$

So, we have that $\nu^* = \nu$. If we consider the "operator" T $(= T(,))$ defined in (10), then $\mu \circ T^{-1}$ belongs to $\mathcal{P}_{\Sigma_\nu}$ and, by Theorem 2.10 we have that

$$W^2(\mu, \mathcal{P}_{\Sigma_\nu}) = W^2(\mu, \mu \circ T^{-1}).$$

Therefore $\nu = \mu \circ T^{-1}$ and the result is proved if μ fulfils the condition in Proposition 2.9.

For the general case, let X and Y be two random elements, with distribution μ and ν respectively, such that

$$W^2(\mu, \nu) = E\|X - Y\|^2.$$

Let B be a Bernoulli random variable and let G_μ be a random element with Gaussian distribution with the same mean and covariance operator as μ. Let us assume that these random variables are mutually independent and independent of X and Y respectively.

Let $T := T(,)$ as defined in (10) and consider $G_\nu = T \circ G_\mu$. Thus G_ν is a random element with Gaussian distribution with the same mean and covariance operator as ν.

We define the random elements $U := B \times X + (1 - B) \times G_\mu$ and $V := B \times Y + (1 - B) \times G_\nu$. Then the mean vector and the covariance operator of U and V coincide with those of X and Y respectively and

$$W^2(P_U, P_V) \leq E\left[[U - V]^2\right] = E\left[I_{\{B=1\}}[X - Y]^2\right] + E\left[I_{\{B=0\}}[G_\mu - G_\nu]^2\right]$$
$$= trace(+ - 2)$$

and the inequality is an equality by the Theorem 2.7.

Lemma 2.12 implies that U satisfies the condition in Proposition 2.9. Then the first part in this theorem implies that that $V = TU$ a.s. and we have that in fact $Y = TX$ a.s. as we wanted to prove because T is linear.

As a consequence of the proof of the previous theorem we obtain the following corollary which gives the unicity of the representation of the pair in which the Wasserstein distance between two probability measures is reached when there exists a linear relation between them (compare with Proposition 2.9).

Corollary 2.14 *Let μ and ν be two strong order two probability measures such that $Ker = \{0\}$. Let us suppose that there exists a linear operator A such that $\nu = \mu \circ A^{-1}$. Then if $W(\mu, \nu) = \|X - Y\|_2$ we have that $Y = AX$ a.s.*

Theorem 2.15 *Under the hypotheses in Theorem 2.7 we have that*

$$W^2(\mu, \nu) \geq \|m_\mu - m_\nu\|^2 + trace\left[+ - 2\,()^{1/2}\right]. \tag{11}$$

Moreover, the following statements are equivalent

1. *Equality holds in (11).*
2. $\nu^* \circ \Pi_\mu^{-1} = \mu^* \circ T(,)^{-1}$.
3. $\mu^* \circ \Pi_\nu^{-1} = \nu^* \circ T(,)^{-1}$.

(Recall that μ^* and ν^* are the centered in mean probablity measures, and Π_{Σ_μ} and Π_{Σ_ν} denote the projections on the subspaces $(Ker)^\perp$ and $(Ker)^\perp$ respectively).

PROOF.- Let $\{\vec{e}_n\}$ be an orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of and let $L = \{i : \vec{e}_i \notin Ker\}$.

If X and Y are two random elements with distributions μ and ν respectively we have that

$$\|X - Y\|^2 = \sum_{i \notin L} E < Y, \vec{e}_i >^2 + \|\Pi_{\Sigma_\mu} X - \Pi_{\Sigma_\mu} Y\|^2 =$$

$$trace(\Pi_\mu^\perp \Pi_\mu^\perp) + \|\Pi_\mu X - \Pi_\mu Y\|^2$$

where $\Pi_\mu^\perp$ denotes for the projection on Ker.

Therefore the first term in the right-hand side is a constant which cannot be modified and the equivalence between statements 1 and 2 in the theorem is obtained trivially applying Theorem 2.10 and Proposition 2.13 to the subspace $(Ker)^\perp$.

The equivalence between 1 and 3 is proved similarly.

Acknowledgment

Authors want to thank an anonymous referee for helpful comments which have improved considerably the paper.

References

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

[2] CONWAY, J.B. (1985). *A Course in Functional Analysis.* Springer. New York.

[3] CUESTA-ALBERTOS, J.A., DOMINGUEZ-MENCHERO, J.S. and MATRAN-BEA, C. (1991). Some stochastics on monotone functions. Preprint.

[4] CUESTA-ALBERTOS, J.A. and MATRAN-BEA, C. (1991). Notes on the Wasserstein metric in Hilbert spaces. *Ann. Prob.* 17, 1264-1276.

[5] CUESTA-ALBERTOS, J.A., C. MATRAN-BEA, L and TUERO-DIAZ, A. (1993). Optimal maps for the L^2-Wasserstein distance. Preprint.

[6] CUESTA-ALBERTOS, J.A., RÜSCHENDORF, L and TUERO-DIAZ, A. (1993). Optimal coupling of multivariate distributions and stochastic processes. To appear in *J. Multivariate Anal.*.

[7] GELBRICH, M. (1990). On a formula for the L^2-Wasserstein metric between measures on Euclidean and Hilbert Spaces. *Math. Nachr.* 147, 185-203.

[8] KNOTT, M. and SMITH, C.S. (1984). On the optimal mapping of distributions. *J. Optim. Theory Appl.* 43, 39-49.

[9] LAHA, R.G. and ROHATGI, V.K. (1979). *Probability Theory.* Wiley. Chichester.

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

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

[12] RÜSCHENDORF, L. (1991). Fréchet bounds and their applications. In: *Advances in probability distributions with given marginals.* Eds. G. Dall'Aglio, S. Kotz and G. Salinetti. Pgs. 151-187.

[13] SCHWEIZER, B. (1991). Thirty years of copulas. In: *Advances in probability distributions with given marginals.* Eds. G. Dall'Aglio, S. Kotz and G. Salinetti, Pgs. 13-50.

[14] SMITH, C. and KNOTT, M. (1990). A note on the bound for the L^2 Wasserstein Metric. Unpublished paper.

[15] STOUT, W.F. (1974). *Almost Sure Convergence.* Academic Press. New York.

[16] VAKHANIA, N.N., TARIELADZE, V.I. and CHOBANYAN, S.A. (1987). *Probability Distributions on Banach Spaces.* Reidel. Dordrecht.

[17] WHITT,W.(1976). Bivariate distributions with given marginals.*Ann. Statist.* 4, 1280-1289.

J.A. Cuesta-Albertos
Departamento de Matemáticas, Estadística y Computación.
Facultad de Ciencias. Universidad de Cantabria.
39005 Santander (SPAIN).

C. Matrán-Bea
Departamento de Estadística e Investigación Operativa
Facultad de Ciencias. Universidad de Valladolid.
47002 Valladolid (SPAIN).

A. Tuero-Diaz
Departamento de Matemáticas, Estadística y Computación.
Facultad de Ciencias. Universidad de Cantabria.
39005 Santander (SPAIN).