Optimal coupling of multivariate distributions and stochastic processes

J.A. Cuesta-Albertos[1], L. Rüschendorf[2] and A. Tuero-Diaz.[1]

[1] Departamento de Matemáticas, Estadística y Computación.
Universidad de Cantabria. Spain.
[2] Institut für Mathematische Statistik.
University of Münster. Germany.

Abstract

Some explicit optimal coupling results are derived with respect to minimal metrics of l_p-type. In particular the optimality of radial transformations, positive transformations and monotone transformations of the components is established.

Key words and phrases: Optimal couplings, Wasserstein metrics, multivariate distributions, radial transformations, positive transformations, monotone transformations.

A.M.S. 1980 subject classification: 60E05, 62H20.

1 Introduction.

Let (U,$\mathcal{L}$) be a measure space and let $\sigma : U \times U \to \Re_+$ denote a product measurable function. The Kantorovich-functional induced by σ (which we also denote by σ) on the set $\mathcal{M}^1(U)$ of all probability distributions on $\mathcal{L}$ is defined by

$$\sigma(P,Q) = \inf\left\{\int \sigma d\mu; \mu \in \mathcal{M}(P,Q)\right\},$$

where $\mathcal{M}$(P,Q) is the set of all probability measures on $\mathcal{L} \otimes \mathcal{L}$ with marginals P and Q. Note that $\sigma(P,Q)$ is a measure of the proximity betwen P and Q. A probability $\mu^* \in \mathcal{M}(P,Q)$ is called an optimal coupling with respect to σ (o.c.(σ)) if

$$\sigma(P,Q) = \int \sigma d\mu^*.$$

Similarly a pair, (X,Y), of $U \times U$-valued random variables (r.v.'s) on a common probability space, (Ω,$\mathcal{A}$, λ), is called an o.c.(σ) between P and Q if $X \stackrel{\mathrm{d}}{=} P$, $Y \stackrel{\mathrm{d}}{=} Q$ ($\stackrel{\mathrm{d}}{=}$ means equality in distribution) and

$$\sigma(P,Q) = \int \sigma(X,Y)d\lambda.$$

The existence of o.c.(σ)'s in the case of tight measures P and Q and σ lower semicontinuos follows from the weak compactness of $\mathcal{M}$(P,Q) (cf. [10, 16]).

The most interesting case concerns the situation where (U,d) is a metric space with Borel σ-algebra, $\mathcal{L}$, and $\sigma = d^r$, $r > 0$. In this case $[\sigma(P,Q)]^{(1/r)\wedge 1} =: l_r(P,Q)$ is a probability metric, the minimal L_r-metric (cf. [13]). In this paper we shall consider in particular the determination of o.c.'s for the L_2-metric, which we denote "Wasserstein Distance", W, as do most articles on this subject in spite of the fact that the priority in the definition belongs to Kantorovich [9] (see [13]). This metric is defined as the square root of

$$W^2(P,Q) = \inf\left\{\int d^2(x,y)d\mu(x,y); \mu \in \mathcal{M}(P,Q)\right\}.$$

While for the real line, $U = \Re$, there are a lot of results on o.c.'s, there are for the euclidean spaces, $U = \Re^n$, or even more general spaces, only very few explicit results available (cf. [7, 13, 18, 21, 19, 22] and the references therein). In the case of a Hilbert space, $U = \mathcal{H}$, with inner product, $\langle\cdot,\cdot\rangle$, and P and Q with finite second moments the following result (cf.[11, 19]) is basic for the determination of the Wasserstein distance. Let $X \stackrel{\mathrm{d}}{=} P$, $Y \stackrel{\mathrm{d}}{=} Q$ and let P and Q have finite second moments. Then (X,Y) is an o.c. with respect to the Wasserstein distance if and only if

$$Y \in \partial f(X),\ \lambda\text{-a.s.} \tag{1.1}$$

for some closed (= lower semicontinuous) convex function f, where $\partial f(x)$ denotes the subgradient of f in x:

$$\partial f(x) = \{y : f(x') - f(x) \geq \langle y, x' - x\rangle,\ \forall x' \in \mathrm{dom} f\}.$$

In particular, in the case $U = \Re^n$, if ϕ is continuously differentiable, then $(X, \phi(X))$ is an o.c.(W) for all X in the domain of ϕ if and only if :

$$\begin{array}{l} \text{a) } \phi \text{ is monotone (i.e. } \langle x-y, \phi(x)-\phi(y)\rangle \geq 0,\ \forall x,y); \text{ and} \\ \text{b) } D\phi = \left(\frac{\partial \phi_i}{\partial x_j}\right) \text{ is symmetric } \left(\text{ i.e. } \frac{\partial \phi_i}{\partial x_j} = \frac{\partial \phi_j}{\partial x_i}\right) \end{array} \tag{1.2}$$

(cf. [20, page 162]).

That ϕ is $P^X \times P^X$-a.s. monotone if $(X, \phi(X))$ is an o.c.(W) was shown in [3] for every Hilbert space.

An interesting consequence of (1.1) is the fact that independently of the distribution of X certain functions (as specified in (1.2)) have the optimal coupling property w.r.t. W. In particular one obtains that for any symmetric, positive semidefinite matrix A, (X,AX) is an o.c.(W). This applies to the normal case in the following way: Given two multivariate normal distributions, $P = N(0, \Sigma_1)$ and $Q = N(0, \Sigma_2)$, with regular covariance matrices Σ_1 and Σ_2; if $X \stackrel{\mathrm{d}}{=} N(0, \Sigma_1)$ and $A = \Sigma_1^{-1/2}\left(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\right)^{1/2}\Sigma_1^{-1/2}$ then (X,AX) is an o.c.(W) between $N(0, \Sigma_1)$ and $N(0, \Sigma_2)$ (cf. [6, 11, 12, 19, 20, 22]). This fact was established without (1.2), only by some involved calculations, but in Section 2 we shall give a surprisingly simple proof of this result based on the fact that X and AX have the same dependence structure. Moreover, our proof will allow us to prove also some far reaching extensions of the o.c. property of positive semidefinite functions which also apply to stochastic processes. We furthemore consider monotone transformations of the components (in the case $U = \Re^T$). In this way we obtain optimal couplings between distributions with the same dependence structure.

In Section 3 we propose a new class of functions with a general strong optimal coupling property, the radial transformations. These transformations give optimal couplings not only with respect to W but also with respect to a wide class of Kantorovich functionals induced by $\sigma(X;Y) = \phi(\|X - Y\|)$. We prove that radial transformations always give o.c.'s and that if $\phi(x) = \delta(x)x$ gives an o.c., then, under some restrictions on δ, ϕ must be a radial transformation. We also obtain in this way an o.c. for spherically equivalent distributions.

For general references to the problem of o.c.'s we refer to the recent conference volume: *"Advances in Probability distributions with given marginals"* edited by G. Dall'Aglio, S.Kotz and G. Salinetti, to the book of Rachev [14] and the dissertation of Tuero [24].

2 Distributions with the same dependence structure.

Let X_i be real r.v.'s defined on the probability space $(\Omega, \mathcal{A}, \lambda)$ with distribution functions (d.f.'s) F_i and corresponding distributions P_i, $1 \leq i \leq n$, and let $\{V_i\}_{1\leq i \leq n}$ be i.i.d. r.v.'s uniformly distributed on (0,1) also defined on $(\Omega, \mathcal{A}, \lambda)$ and independent of $\{X_i\}_{1\leq i\leq n}$. Define for $x \in \Re$, $\alpha \in (0, 1)$

$$\mathcal{F}_i(x, \alpha) := \lambda(X_i < x) + \alpha\lambda(X_i = x) \text{ and } U_i := \mathcal{F}_i(X_i, V_i),\ 1 \leq i \leq n,$$

then (cf. [17])

$$U_i \stackrel{\mathrm{d}}{=} U(0, 1) \text{ and } X_i = F_i^{-1} \circ U_i,\ \lambda\text{-a.s.},\ 1 \leq i \leq n, \tag{2.1}$$

where U(0,1) is the uniform distribution on the interval (0,1) and $F_i^{-1}(u) = \inf\{y : F_i(y) \geq u\}$ is the generalized inverse of F_i.

Define, furthermore, that X_1, X_2 are similarly ordered ($X_1 \stackrel{\text{s.o.}}{\sim} X_2$) if

$$(X_1(w) - X_1(w'))\,(X_2(w) - X_2(w')) \geq 0,\ \lambda \otimes \lambda\text{-a.s.}$$

Next we include a proposition concerning the optimal couplings of real r.v.'s. In it we assume, without loss of generality, that the probability space is rich enough to allow us to define on it a r.v. with distribution U(0,1).

Proposition 2.1 *Let X_1 and X_2 be real, square integrable r.v.'s with d.f.'s F_1 and F_2. Then:*

a) Equivalent are:

a.i) (X_1, X_2) *is an o.c.(W).*

a.ii) $F_{(X_1,X_2)}(x,y) = \min\{F_1(x), F_2(y)\}$, $\forall x, y$.

a.iii) There exists a r.v. $U \stackrel{d}{=} U(0,1)$, *such that for some nondecreasing functions* ϕ_1, ϕ_2:

$$X_1 = \phi_1(U) \text{ and } X_2 = \phi_2(U),\ \lambda - a.s.$$

a.iv) $X_1 \stackrel{s.o.}{\sim} X_2$.

b) The functions ϕ_i *in a) are essentially unique,* $\phi_i = F_i^{-1}$ *a.s. with respect to Lebesgue measure.*

c) The pairs $\left(X_1, F_2^{-1} \circ \mathcal{F}_1(X_1, V_1)\right)$ *and* $\left(F_1^{-1} \circ \mathcal{F}_2(X_2, V_2), X_2\right)$ *are o.c.(W).*

d) If P_1 *is nonatomic and* (X_1, Y_1) *is an o.c.(W) between* P_1 *and* P_2 *then*

$$Y_1 = F_2^{-1} \circ F_1(X_1), \lambda\text{-}a.s.$$

e) If $Y_1 = \phi_1(X_1)$ *with* ϕ_1 *non-decreasing, then* (X_1, Y_1) *is an o.c.(W).*

PROOF.- The first equivalence in a) follows, e.g., from the Hoeffding representation

$$EX_1X_2 = \int \left(F_{(X_1,X_2)}(x,y) - F_1(x)F_2(y)\right) \mathrm{dx\ dy} + (EX_1)(EX_2) \tag{2.2}$$

together with the Fréchet bounds saying that the class of d.f.'s with marginals P_1 and P_2, $\mathcal{F}(P_1, P_2)$, has a largest element $\overline{F}$ and a smallest element $\underline{F}$ (the upper and lower Fréchet bounds)

$$\overline{F}(x,y) = \min\{F_1(x), F_2(y)\} \text{ and } \underline{F}(x,y) = (F_1(x) + F_2(y) - 1)^+.$$

It is evident that *iii)* implies *ii).* That *iv)* is obtained from *i)* was proved in [23]. Therefore, to show a), we only have to prove that *iv)* implies *iii).*

Let us consider the real r.v.'s $U_1 = \mathcal{F}_1(X_1, V_1)$ and $U_2 = \mathcal{F}_2(X_2, U_1)$ where V_1 is a r.v. with distribution U(0,1) and independent from X_1 and X_2.

First we are going to prove that the distribution of U_2 is absolutely continuous. To this end let $\omega, \omega' \in \Omega$ such that $U_2(\omega) = U_2(\omega')$ and let us denote $x_1 = X_1(\omega)$, $x_1' = X_1(\omega')$, $x_2 = X_2(\omega)$ and $x_2' = X_2(\omega')$.

Note that $U_1(\omega) = U_1(\omega')$ if $x_2 = x_2'$, and that $\lambda[X_2 < x_2] = \lambda[X_2 < x_2']$ and $\lambda[X_2 = x_2] = \lambda[X_2 = x_2'] = 0$ if $x_2 < x_2'$ and $U_1(\omega)$ and $U_1(\omega')$ belong to (0,1). So we have proved that

$$\lambda \otimes \lambda\{(\omega, \omega') : U_2(\omega) = U_2(\omega')\} \leq \lambda \otimes \lambda\{(\omega, \omega') : U_1(\omega) = U_1(\omega')\} +$$

$$\lambda \otimes \lambda\{(\omega, \omega') : X_2(\omega),\ X_2(\omega') \in A\} + 2\lambda\{\omega : U_1(\omega) \notin (0,1)\}$$

where A is the P_2-probability zero set defined as the union of all nontrivial, non-empty and closed intervals with P_2-probability zero.

The second summand in the last term is zero by definition of A. The other ones are also zero because the distribution of U_1 is U(0,1). Therefore we have proved that the distribution of U_2 is continuous.

On the other hand, let $\omega, \omega' \in \Omega$ and let us use the same notation as above. If $x_2 < x_2'$ then, by construction, $U_2(\omega) \leq U_2(\omega')$ and, by hypothesis, $x_1 \leq x_1'$, λ-a.s. If $x_2 = x_2'$ and $x_1 < x_1'$ then, by construction, $U_2(\omega) \leq U_2(\omega')$. Therefore we have shown that $X_1 \overset{\text{s.o.}}{\sim} U_2$. This and the fact that the distribution of U_2 is absolutely continuous, permit us to apply the reasoning developed in Theorem 1 in [23] to obtain the existence of an increasing function, ϕ_1, such that $X_1 = \phi_1(U_2)$, λ-a.s.

Finally, the same argument which conducts to (2.1) gives us that $X_2 = F_2^{-1}(U_2)$, λ-a.s. and the proof ends because U_2 can be written as an increasing function of a r.v. with distribution U(0,1).

b) If $X_1 = \Phi_1(U)$, Φ_1 nondecreasing, then

$$F_1(x) = \lambda(X_1 \leq x) = \lambda(\Phi_1(U) \leq x) = \lambda(U \leq \Phi_1^{-1}(x)) = \Phi_1^{-1}(x)$$

and, therefore, $\Phi_1 = F_1^{-1}$. Cf. also Lemma 2.3 in [2].

c) and e) follow from a) since

$$X_1 \overset{\text{s.o.}}{\sim} F_2^{-1} \circ \mathcal{F}_1(X_1, V_1),\ X_2 \overset{\text{s.o.}}{\sim} F_1^{-1} \circ \mathcal{F}_2(X_2, V_2)$$

and $X_1 \overset{\text{s.o.}}{\sim} \phi_1(X_1)$ if ϕ_1 is nondecreasing.

Alternatively, define $\alpha(s) := F_2^{-1} \circ F_1(s)$ and $F(t) := \int_0^t \alpha(s)ds$. Then α is monotonically nondecreasing and so F is convex. Furthermore, $\partial F(s) = [\alpha(s-), \alpha(s+)]$ (c.f. [15, Th. 24.2]) .

Since $P(X_1 < x_1) \leq F_1(x_1, v_1) \leq F_1(x_1)$ for all $x_1 \in \Re^1, v_1 \in (0,1)$, we obtain

$$\alpha(s-) \leq F_2^{-1} \circ F_1(s, v_1) \leq \alpha(s+)$$

Therefore,

$$F_2^{-1} \circ F_1(X_1, V_1) \in \partial F(X_1) \text{ a.s.}$$

and, by (1.1), $\left(X_1, F_2^{-1} \circ F_1(X_1, V_1)\right)$ is an optimal coupling. e) can be proved similarly.

d) follows from b) and the fact that if F_1 is continuous, then $F_1 \circ F_1^{-1}(u) = u$. ■

Remark 1

1. We remark that $a.ii)$, $a.iii)$ and $a.iv)$ in Proposition 2.1 are equivalent and imply $a.i)$ without any assumption on the square integrability of X_1 and X_2.

   For the proof note that the statement "$a.ii)$ implies $a.i)$" follows from Theorem 1 in [1] and that the proof that $a.iv)$ (resp. $a.iii)$) implies $a.iii)$ (resp. $a.ii)$) that we have proposed in Proposition 2.1 does not use any integrability argument. So we have only to prove that $a.ii)$ implies $a.iv)$.

   To prove this let us assume that $a.ii)$ is verified but $a.iv)$ does not hold. In this case the reasoning in Theorem 1 in [23] permits us to conclude that there exist two real numbers j and k such that

   $$\lambda \otimes \lambda \{(\omega, \omega') : X(\omega) \leq j,\ X(\omega') > j,\ Y(\omega') \leq k \text{ and } Y(\omega) > k\} > 0.$$

   Therefore

   $$\lambda\left(\{X \leq j\} \cap \{Y > k\}\right) > 0 \text{ and } \lambda\left(\{Y \leq k\} \cap \{X > j\}\right) > 0$$

   and we have that $\lambda\left(\{X \leq j\} \cap \{Y \leq k\}\right) < \inf\left(\lambda\{X \leq j\}, \lambda\{Y \leq k\}\right)$ which contradicts $a.ii)$.

2. The arguments in [23] to prove that $a.i)$ implies $a.ii)$ hold under the weaker assumption that $W(F_1, F_2)$ is finite. Then we see that, under this assumption, statements $a)$ to $e)$ in Proposition 2.1 are equivalent.

3. In $a)$ we have proved the existence and the uniqueness of the joint optimal d.f. (namely $\overline{F}$) for the o.c.(W). It is well-known that $\overline{F}$ is optimal for σ a quasi-monotone, right-continuous function, e.g. $\sigma(x, y) = \phi(x - y)$, ϕ convex (cf. [5]). The uniqueness property holds also for all l_p-distance, $p > 1$ (see [5]) and for strictly convex ϕ but not for the l_1-distance. This can be seen from the formula (analogous to (2.2))

   $$E \mid X_1 - X_2 \mid = \int \Big(F_1(u) + F_2(u) - 2F_{(X_1,X_2)}(u, u)\Big) \,\mathrm{du}.$$

   Therefore, (X_1, X_2) is an o.c.(l_1) if and only if

   $$F_{(X_1,X_2)}(u, u) = \min\{F_1(u), F_2(u)\},\ \forall u.$$

   For this and related results cf. the recent survey article [5].

In the multivariate case there is not a similar complete result as Proposition 2.1. For instance, in the real case, according to b), each probability measure P on $\Re$ has associated a function Φ_P in such a way that an o.c.(W) between P and Q is given by $(\Phi_P(U), \Phi_Q(U))$ where U $\stackrel{\mathrm{d}}{=}$U(0,1). This kind of universal representation does not exist in $\Re^n$ as the following example (proposed in [4]) shows.

Example 2.2 Let us consider the points m_i in $\Re^2$:

$$m_0 = (0,0),\ m_1 = (1,0),\ m_2 = (\sqrt{2}/2, \sqrt{2}/2),\ m_3 = (-1/2, \sqrt{3}/2).$$

and the probabilities P_i, i=1,2,3, which give probability 1/2 to each one of m_0, and m_i, i=1,2,3, respectively.

If we suppose that there exist X_i, i=1,2,3, defined on some probability space, such that (X_i, X_j) is an o.c.(W) between P_i and P_j, $\forall i, j$, then it is easy to show that (X_1, X_2, X_3) has the following incompatible two-dimensional marginals:

$$\lambda\{(X_1, X_2) = (m_0, m_0)\} = \lambda\{(X_1, X_2) = (m_1, m_2)\} = \frac{1}{2},$$

$$\lambda\{(X_2, X_3) = (m_0, m_0)\} = \lambda\{(X_2, X_3) = (m_2, m_3)\} = \frac{1}{2},$$

$$\lambda\{(X_1, X_3) = (m_0, m_3)\} = \lambda\{(X_1, X_3) = (m_1, m_0)\} = \frac{1}{2}.$$

The following proposition shows the difference with respect to the generalization of the Fréchet bounds. Let $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$ be r.v.'s in $\Re^n$ with d.f.'s F and G and distributions P and Q. Let $\mathcal{F}(P, Q)$ be the class of d.f.'s on $\Re^{2n}$ with marginal d.f.'s F, G. Let P_i and Q_i denote the marginals of P and Q and F_i and G_i the marginals of F and G, i=1,...,n.

Proposition 2.3

a) $\max\{H(x,y); H \in \mathcal{F}(F,G)\} = \min\{F(x), G(y)\} =: \overline{F}(x,y),\ \forall x, y \in \Re^n.$

b) $\min\{H(x,y); H \in \mathcal{F}(F,G)\} = (F(x) + G(y) - 1)^+ =: \underline{F}(x,y),\ \forall x, y \in \Re^n.$

c) If $\overline{F}$ is a d.f. and $(X,Y) \overset{d}{=} \overline{F}$, then $X_i \overset{s.o.}{\sim} Y_j$ for all i,j. If P_1 is nonatomic, then $Y_j = \phi_j(X_1)$ for some nondecreasing functions ϕ_j. If P_1 and Q_i, for some i, are nonatomic, then additionally $X_j = \psi_j(X_1)$ for some nondecreasing functions ψ_j.

PROOF.- a) and b) are consequences of Theorem 6 in [16] on the sharpness of Fréchet bounds.

Let us see c). If $\overline{F}$ is a d.f., then (with an obvious notation) $\overline{F}_{ij}(x_i, y_j) = \min\{F_i(x_i), G_j(y_j)\}$ and, therefore, by Proposition 2.1,a) for $(X,Y) \overset{\mathrm{d}}{=} \overline{F}$ one obtains $X_i \overset{\mathrm{s.o.}}{\sim} Y_j$, $\forall i, j$. If P_1 is nonatomic, by Proposition 2.1,d), $Y_j = G_j^{-1} \circ F_1(X_1) = \phi_j(X_1)$. If Q_i also is nonatomic, then from $X_j \overset{\mathrm{s.o.}}{\sim} Y_i$, $X_j = \tilde{\phi}_j(Y_i)$ with $\tilde{\phi}_j$ nondecreasing and, therefore, $X_j = \tilde{\phi}_j[\phi_i(X_1)] = \psi_j(X_1)$ and ψ_j is nondecreasing.

■

Remark 2

1. Similarly, if $\underline{F}(x,y)$ is a d.f., then for any $(X,Y)\stackrel{\rm d}{=}\underline{F}$, we have that $X_i \stackrel{\rm o.o.}{\sim} Y_j$ ($\stackrel{\rm o.o.}{\sim}$ means oppositely ordered). If P_1 is nonatomic, then $Y_j = \phi_j(X_1)$, ϕ_j decreasing. If additionally Q_i is nonatomic, then also $X_j = \psi_j(X_1)$, ψ_j increasing. This shows that only in very exceptional cases there exist a smallest or a largest d.f. in $\mathcal{F}$(F,G) for $n \geq 2$.

2. If P_1 is continuous and $Q = \varepsilon_{\{a\}}$, the one point measure in the point $a \in \Re^n$, if $X\stackrel{\rm d}{=}P$ and $Y\stackrel{\rm d}{=}Q$, then $F_{(X,Y)}(x,y) = \min\{F(x), G(y)\} = \overline{F}(x,y)$ holds. Obviously, $X_j = \psi_j(X_1)$ with ψ_j increasing does not hold generally so that the condition of a nonatomic marginal Q_i can not be omitted in c) of Proposition 2.3. This is related to the fact that $\stackrel{\rm s.o.}{\sim}$ is not transitive (generally), i.e. $X_1 \stackrel{\rm s.o.}{\sim} Y_1$ and $Y_1 \stackrel{\rm s.o.}{\sim} Z_1$ does not imply $X_1 \stackrel{\rm s.o.}{\sim} Z_1$.

In spite of the previous considerations, a simple positive optimal coupling result holds for translations.

Lemma 2.4 *Let $X\stackrel{d}{=}P$ and for some $a \in \Re^n$, let $X+a\stackrel{d}{=}Q$. Let $\|\cdot\|$ denote the euclidean norm on $\Re^n$ and let $\phi : \Re_+ \to \Re_+$ be a convex, increasing function. Then $(X, X+a)$ is an o.c.(σ) between P and Q for $\sigma = \phi \circ \|\ \|$, i.e. translations are o.c.(σ)*

PROOF.- Let (X,Y) be a pair of r.v.'s with marginal distributions P and Q respectively. Then by Jensen's inequality and the increasing character of ϕ we have that

$$E\left(\phi\|X-Y\|\right) \geq \phi\left(E\|X-Y\|\right) \geq \phi\left(\|E[X-Y]\|\right) =$$

$$= \phi\left(\|a\|\right) = E\phi\left(\|X-(X+a)\|\right).$$

So (X,X+a) is an o.c.(σ).

∎

Note that the proof of the preceeding lemma remains valid for every normed linear space. But it does not hold necessarily in any metric space or even if ϕ is not convex as the following example shows.

Example 2.5 Let P $\stackrel{\rm d}{=}$U(0,1) and take $a = \frac{1}{2}$ and

$$\sigma(x,y) = \begin{cases} 1, & \text{if } x \neq y \\ 0, & \text{if } x = y. \end{cases}$$

A transformation which behaves better than the translation is the following one:

$$T(x) = \begin{cases} x, & \text{if } \frac{1}{2} \leq x \leq 1 \\ x+1, & \text{if } 0 \leq x \leq \frac{1}{2}. \end{cases}$$

On the other hand translations are special cases of monotonic, component to component transformations. Optimality of these transformations is a consequence of the next result.

Proposition 2.6 *Let* $\sigma(x,y) = \sum_{i=1}^n \sigma_i(x_i,y_i)$*, where* σ_i *are quasimonotone and right-continuous. Let* $P,Q \in \mathcal{M}^1(\Re^n,\mathcal{L}^n)$ *with marginal distributions* P_i*,* Q_i*,* $1 \le i \le n$*, and let* $X \stackrel{d}{=} P$*,* $Y \stackrel{d}{=} Q$*, then*

a) $\sigma(P,Q) \ge \sum_{i=1}^n \sigma_i(P_i,Q_i)$*.*

b) If $X_i \stackrel{s.o.}{\sim} Y_i$*,* $1 \le i \le n$*, then* (X,Y) *is an o.c.*(σ) *and* $E\sigma(X,Y) = \sum_{i=1}^n \sigma_i(P_i,Q_i)$*.*

c) If $\sigma_i(x_i,y_i) = \phi_i(x_i - y_i)$*,* ϕ_i *strictly convex, then the converse in b) holds also, i.e. given the r.v.'s* X *and* Y*, then*

$$(X,Y) \text{ is an o.c.}(\sigma) \text{ and } \sigma(P,Q) = \sum_{i=1}^n \sigma_i(P_i,Q_i) \Longleftrightarrow X_i \stackrel{s.o.}{\sim} Y_i,\ 1 \le i \le n.$$

PROOF.- With respect to a), it is evident that

$$\begin{aligned}
\sigma(P,Q) &= \inf\left\{\int \sum_{i=1}^n \sigma_i(x_i,y_i)d\mu(x,y)\ ;\ \mu \in \mathcal{M}(P,Q)\right\} \\
&\ge \sum_{i=1}^n \inf\left\{\int \sigma_i(x_i,y_i)d\mu(x,y)\ ;\ \mu \in \mathcal{M}(P,Q)\right\} \\
&= \sum_{i=1}^n \inf\left\{\int \sigma_i(x_i,y_i)d\mu_i(x_i,y_i)\ ;\ \mu_i \in \mathcal{M}(P_i,Q_i)\right\} \\
&= \sum_{i=1}^n \sigma_i(P_i,Q_i).
\end{aligned}$$

On the other hand, equality holds in a) if and only if there exists an o.c.(σ), (X,Y), between P and Q such that $E\sigma_i(X_i,Y_i) = \sigma_i(P_i,Q_i)$, $1 \le i \le n$.

A sufficient condition is $X_i \stackrel{\text{s.o.}}{\sim} Y_i$, $1 \le i \le n$, which is also necessary for ϕ_i strictly convex functions (cf. Proposition 2.1 and Remark 1).

∎

From the previous Proposition and Proposition 2.1 we obtain:

Corollary 2.7 *Let* $\sigma(x,y) = \sum_{i=1}^n \phi_i(x_i - y_i)$*,* ϕ_i *strictly convex. Let* $P,Q \in \mathcal{M}^1(\Re^n,\mathcal{L}^n)$ *with marginal distributions* P_i*,* Q_i*, where* P_i *are nonatomic, and marginal d.f.'s* F_i *and* G_i*,* $1 \le i \le n$*, and let* $X \stackrel{d}{=} P$*,* $Y \stackrel{d}{=} Q$*; then equivalent are:*

1. (X,Y) *is an o.c.*(σ) *and* $\sigma(P,Q) = \sum_{i=1}^n \sigma_i(P_i,Q_i)$*.*
2. $Y_i = G_i^{-1} \circ F_i(X_i)$*,* λ*-a.s.,* $1 \le i \le n$*.*

An inmediate consequence of Proposition 2.6 is:

Corollary 2.8 *If the function $\overline{F}$ defined in Proposition 2.3.a) is a d.f., then $\overline{F}$ gives an o.c.(σ) for σ as in Proposition 2.6.*

As is well-known, for any n-dimensional d.f., F, there exists a d.f. $C \in \mathcal{M}(U(0,1),...,U(0,1))$, such that

$$F(x_1,...,x_n) = C\left(F_1(x_1),...,F_n(x_n)\right). \tag{2.3}$$

We call any C as in (2.3) a dependence function of F (copula). If F_i are continuous, $1 \leq i \leq n$, then C is uniquely determined (cf., for instance, the review article [22]).

A natural construction of C is given by

$$C^*(u_1,...,u_n) = \lambda(U_1 \leq u_1,...,U_n \leq u_n) \tag{2.4}$$

where $(X_1,...,X_n) \stackrel{\mathrm{d}}{=} F$, $U_i = \mathcal{F}_i(X_i,V_i)$, $\{V_i\}$ are i.i.d., U(0,1)-distributed r.v.'s (cf. [17]). Note that C^* depends on the choice of $\{V_i\}$).

If F_i are continuous then C^* coincides with the structure function

$$D(u_1,...,u_n) = \lambda\left\{F_1(X_1) \leq u_1,...,F_n(X_n) \leq u_n\right\}.$$

We say that P and Q have the same dependence structure, if there exists a common dependence function of P and Q; or equivalently, if

$$C_X^* = C_Y^* \text{ for some X } \stackrel{\mathrm{d}}{=}\text{P, Y } \stackrel{\mathrm{d}}{=}\text{Q (with the same choice of } \{V_i\}).$$

As the main result of this section we now can state:

Theorem 2.9 *Let $P,Q \in \mathcal{M}^1(\Re^n,\mathcal{L}^n)$ and $X \stackrel{d}{=} P$, then:*

a) Equivalent are:

a.i) $W(P,Q) = \sum_{i=1}^n W(P_i,Q_i)$.

a.ii) P and Q have the same dependence structure.

a.iii) $(X,\overline{Y})$ is an o.c.(W) between P and Q where $\overline{Y}_i = G_i^{-1} \circ \mathcal{F}_i(X_i,V_i)$, $\{V_i\}$ i.i.d., independent of X, $V_i \stackrel{d}{=} U(0,1)$.

a.iv) There exists $Y \stackrel{d}{=} Q$, such that $X_i \stackrel{s.o.}{\sim} Y_i$, $1 \leq i \leq n$.

b) The same characterization as in a) also holds for

$$\sigma(x,y) = \sum_{i=1}^n \phi_i(x_i - y_i)$$

with ϕ_i strictly convex.

PROOF.- By Proposition 2.6,c) $W(P,Q)=\sum_{i=1}^{n} W(P_i,Q_i)$ is equivalent to iv).

It is obvious that iii) implies iv).

Let us see that ii) implies iii). In effect, if P and Q have the same dependence structure and F$_i$, G$_i$ are the marginal d.f.'s of P and Q respectively, then for some $X \overset{\mathrm{d}}{=} P$, $Y \overset{\mathrm{d}}{=} Q$, $U_i=\mathcal{F}_i(X_i,V_i)$, $\tilde{U}_i=\mathcal{G}_i(Y_i,V_i)$, we have

$$U=(U_1,...,U_n)\overset{\mathrm{d}}{=}\tilde{U}=(\tilde{U}_1,...,\tilde{U}_n)$$

and, therefore,

$$Q\overset{\mathrm{d}}{=}Y=\left(G_1^{-1}(\tilde{U}_1),...,G_n^{-1}(\tilde{U}_n)\right) \overset{\mathrm{d}}{=} \left(G_1^{-1}(U_1),...,G_n^{-1}(U_n)\right)=\overline{Y}$$

In other words $\overline{Y}\overset{\mathrm{d}}{=}Q$ and, by c) in Proposition 2.1 and Remark 1 we have obtained iii).

To end let us show that iv) implies ii). If $X\overset{\mathrm{d}}{=}P$, $Y\overset{\mathrm{d}}{=}Q$, $X_i \overset{\mathrm{s.o.}}{\sim} Y_i$ then by Proposition 2.1,a) and b), there exist $U_1,...,U_n$ such that $X_i=F_i^{-1}(U_i)$, $Y_i=G_i^{-1}(U_i)$. Let $C(x_1,..,x_n)=\lambda(U_1\le x_1,...,U_n\le x_n)$, then

$$F(x_1,...,x_n)=C\left(F_1(x_1),...,F_n(x_n)\right)$$

and

$$G(x_1,...,x_n)=C\left(G_1(x_1),...,G_n(x_n)\right)$$

where $F=F_P$ and $G=F_Q$, and P and Q have the same dependence structure.

■

From Theorem 2.9 we obtain in particular that affine transformations like $T(x_1,...,x_n)=(a_1+\lambda_1x_1,\dots,a_n+\lambda_nx_n)$, $a_i,\lambda_i\in\Re$, $\lambda_i>0$ give an o.c.(W). This gives in particular an o.c.(W) between uniform distributions on ellipses.

We note that the preceeding ideas can be extended to cover some infinite dimensional cases. For instance, let us suppose that $U=\Re^T$. Then, if P and Q are two probability measures on $\mathcal{L}^T$, we say that P and Q have the same dependence structure if there exist two U-valued stochastic processes $X=(X_t,t\in T)$ and $Y=(Y_t,t\in T)$, such that

a) $X\overset{\mathrm{d}}{=}P$ and $Y\overset{\mathrm{d}}{=}Q$, and

b) $X_t \overset{\mathrm{s.o.}}{\sim} Y_t$, for every t ∈T.

The equivalence between *a.ii)* and *a.iv)* in the Theorem 2.9 guarantees that this definition generalizes the previous one.

Now let $x=(x_t,t\in T)$, $y=(y_t,t\in T)\in\Re^T$ and $\sigma_t:\Re\times\Re\longrightarrow\Re_+$ be quasimonotone, right-continuous functions. Let us define

$$\sigma(x,y)=\begin{cases}\sum_{t\in T}\sigma_t(x_t,y_t), & \text{if T is denumerable}\\ \int_T\sigma_t(x_t,y_t)dt, & \text{if T is an interval.}\end{cases}$$

All results which we have stated in the finite-dimensional case, remain valid in the new situation with the only additional assumption that there exist X $\stackrel{d}{=}$P and Y $\stackrel{d}{=}$Q such that $E\sigma(X,Y)<\infty$.

This include some interesting cases. For instance, suppose that $\{h_t,\ t\in[0,1]\}$ is a family of one-dimensional increasing functions. Then the stochastic processes $\{X_t,\ t\in[0,1]\}$ and $\{h_t(X_t),\ t\in[0,1]\}$ are an o.c.(σ). This gives an o.c.(σ) between a stochastic process and its modification by some coordinate-weight strategies.

On the other side, if $U=\Re^n$, A is a positive semidefinite matrix and X is a r.v., then the pair (X,AX) is an o.c.(W) with respect to the euclidean norm (cf. Section 1). This result, which was proved before by other involved procedures, is a simple consequence of Theorem 2.9 and of the fact that X and AX have the same dependence structure if we choose the right basis on U. More generally, based on the discussion above we have the following result:

Theorem 2.10 *Let us suppose that $(\mathcal{H},\langle\cdot,\cdot\rangle)$ is a real, separable Hilbert space, let X be an $\mathcal{H}$-valued r.v. and let A be a self-adjoint and positive operator with purely discrete spectrum. Then, (X,AX) is an o.c.(W).*

PROOF.- Let $\{e_n\}_{n\in\mathcal{N}}$ be an orthonormal basis for $\mathcal{H}$ consisting of eigenvectors of A. Then

$$X=\sum_{n=1}^{\infty}\langle X,e_n\rangle\, e_n=\sum_{n=1}^{\infty}X_n e_n,$$

$$AX=\sum_{n=1}^{\infty}\delta_n\langle X,e_n\rangle\, e_n=\sum_{n=1}^{\infty}\delta_n X_n e_n$$

where $\delta_n\geq 0$ are the eigenvalues of A. Since $E\|X-AX\|^2=\sum_{n=1}^{\infty}E(X_n-\delta_n X_n)^2$ and $X_n=\langle X,e_n\rangle\stackrel{\text{s.o.}}{\sim}\delta_n\langle X,e_n\rangle=\langle AX,e_n\rangle$, the above remark applies if we identify X with (X_n).

X and AX (considered in the basis $\{e_n\}$) have the same dependence structure and (X,AX) is an o.c.(W).

■

This result has some new consequences. For instance, if T is a denumerable set, P_{X_t} is nonatomic, t $\in$T, Y $\stackrel{d}{=}$AX and (X,Y) is an o.c.(W), then, by Corollary 2.7, $Y=AX$, λ-a.s., i.e., we have the uniqueness of optimal couplings.

In particular if P and Q are Gaussian measures on $\Re^n$ with mean vector zero and regular variance-covariance matrices Σ_1 and Σ_2, and X $\stackrel{d}{=}$P, then BX $\stackrel{d}{=}$Q, where

$$B=\Sigma_1^{-1/2}\left(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\right)^{1/2}\Sigma_1^{-1/2},$$

(see the introduction) and therefore the o.c.(W) between two Gaussian distributions is unique.

As a new application, let $(T, \mathcal{A}, \mu)$ be a measure space, let $\mathcal{H} = L^2(T, \mathcal{A}, \mu)$ and let A be a positive semidefinite, selfadjoint L^2-operator given by a L^2-kernel (which we denote by $\tilde{A}$)

$$AX(s) := \int \tilde{A}(s,t) X_t d\mu(t), \; s \in [a,b];$$

we have the following corollary:

Corollary 2.11 *If X is a continuous, Gaussian L^2-process with mean zero and covariance function R_1 and A is a positive semidefinite, selfadjoint L^2-operator given by the kernel $\tilde{A}$, then (X,AX) is an o.c.(W) w.r.t. $\|x\| = \left(\int x_s^2 d\mu(s)\right)^{1/2}$. AX is a Gaussian process with mean zero and covariance function*

$$R_2(s,t) = \int\int \tilde{A}(s,u) R_1(u,v) \tilde{A}(v,t) d\mu(u) d\mu(v).$$

Finally:

Corollary 2.12 *If P is orthogonally invariant, if A is a $n \times n$ matrix and $AX \stackrel{d}{=} Q$ for $X \stackrel{d}{=} P$ then, the pair (X, A_+X) is an o.c.(W) of P and Q, where $A = A_+O$ is the polar decomposition of A.*

PROOF.- Since A_+ is positive semidefinite and $AX \stackrel{\mathrm{d}}{=} A_+X$, the result follows from Theorem 2.10.

■

Remark 3 If Q is an orthogonal matrix, then typically (X,QX) is not an o.c.(W). If, e.g., X is uniformly distributed in $\Re^2$ on $[-1,1] \times \{0\}$ and T_α is the rotation by the angle α, then for $\alpha \in [-\pi/2, \pi/2]$, $(X, T_\alpha X)$ is an optimal coupling. For $\alpha \in [-\pi/2, 3\pi/2]$ the pair $\left(X, T_{\alpha - \pi/2}\right)$ is an optimal coupling (but not $(X, T_\alpha X)$). For $\alpha \in \{\pi/2, -\pi/2\}$ we do not have uniqueness of an optimal coupling.

Moreover, in [24] it is proved that, in $\Re^n$, if P is uniform on a set A with finite Lebesgue measure and such that it contains a ball then $(X, T_\alpha(X))$ is not an optimal coupling for every $\alpha \neq 0$.

Based on the more involved criterion (1.1) we can omit the assumption of discrete spectrum of A in Theorem 2.10.

Theorem 2.13 *Let $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ be a real, separable, Hilbert space, let X be an $\mathcal{H}$-valued r.v. and let A be a selfadjoint, positive operator. Then (X,AX) is an o.c.(W).*

PROOF.- The function $f(x) = \frac{1}{2}\langle x, Ax \rangle$ is closed convex. By (1.1) we have to show that $AX \in \partial f(X)$ a.s., i.e. $\forall z \in \mathcal{H}$ holds: $f(z) \geq \langle z - x, Ax \rangle$ (c.f. [8]).

Equivalent to this inequality is

$$\langle z, Az\rangle + 2\langle x, Ax\rangle - 2\langle z, Ax\rangle \geq 0, \, \forall z \in \mathcal{H}$$

and this inequality is a consequence of

$$0 \leq \langle z - x, A(z - x)\rangle = \langle z, Az\rangle + \langle x, Ax\rangle - 2\langle x, Az\rangle.$$

■

3 Optimality of radial transformations.

Let $U = \Re^n$ with euclidean norm $\|\cdot\| = \|\cdot\|_2$ and let $\phi : \Re^n \to \Re^n$ be a transformation of the form

$$\phi(x) = \begin{cases} \frac{\alpha(\|x\|)}{\|x\|}x, & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

with α monotonically nondecreasing; then ϕ is called a radial transformation.

If ϕ is a radial transformation and X $\stackrel{\text{d}}{=}$P and ϕ (X) $\stackrel{\text{d}}{=}$Q, then we say that P and Q are of the same radial type. If $Y \stackrel{\text{d}}{=} \phi(X)$ then $\|\phi(X)\| = \alpha(\|X\|)$ and α is determined by the equation

$$F_{\|Y\|}(y) = \lambda\{\alpha(\|X\|) \leq y\} = F_{\|X\|}\left(\alpha^{-1}(y)\right) \tag{3.1}$$

i.e. α maps the p-quantile of $\|X\|$ on the p-quantile of $\|Y\|$, p $\in$(0,1).

Proposition 3.1 *Let $X \stackrel{d}{=} P$ and $\phi(X) \stackrel{d}{=} Q$ for a radial transformation ϕ, then $(X, \phi(X))$ is an o.c.(W).*

PROOF.- Define $F(t) := \int_0^t \alpha(s)$ ds, and $f(x) := F(\|x\|)$. Then f(x) is convex because it is of the form $\varphi[h(x)]$ with h convex, and φ convex and nondecreasing.

If we show that for every x, x'

$$f(x') - f(x) \geq \langle \phi(x), x' - x\rangle$$

then by (1.1), $(X, \phi(X))$ is an o.c.(W) for P,Q. The increasing character of α ensures that:

$$f(x') - f(x) \geq (\|x'\| - \|x\|)\, \alpha(\|x\|)$$

On the other hand

$$\langle \phi(x), x' - x\rangle = \frac{\alpha(\|x\|)}{\|x\|} \langle x, x' - x\rangle$$

$$= \frac{\alpha(\|x\|)}{\|x\|} \left(\|x\|\|x'\|\cos\left(\text{ang}(x, x')\right) - \|x\|^2\right) \leq (\|x'\| - \|x\|)\, \alpha(\|x\|).$$

Alternatively, since $\partial F(t) = [\alpha(t-), \alpha(t+)]$ and $\partial\|x\| = \nabla(\|x\|) = \frac{x}{\|x\|}$ for $x \neq 0$, while $\partial\|0\|$ is the unit ball, we obtain for $x \neq 0$, $\phi(x) = \frac{\alpha(\|x\|)}{\|x\|} x \in \partial f(x)$ (c.f. [8]). For $x = 0$, $\phi(x) = 0$, this is obvious.

■

The result of Proposition 3.1 can be generalized to much more general situations and it turns out that radial transformations have very general optimal couplings properties.

Theorem 3.2 *Let $(U, \|\cdot\|)$ be any normed space with Borel σ-algebra $\mathcal{L}$. Define for $\varphi : \Re_+ \to \Re_+$ convex, increasing, the distance σ_φ by $\sigma_\varphi(x,y) := \varphi(\|x-y\|)$. If $X \stackrel{d}{=} P$, $Y \stackrel{d}{=} Q$ satisfy*

$$Y = \frac{\|Y\|}{\|X\|} X, \ \lambda\text{-a.s.} \tag{3.2}$$

and

$$\|X\| \stackrel{s.o.}{\sim} \|Y\|; \tag{3.3}$$

then (X,Y) is an o.c.(σ_φ).

PROOF.- Since φ is nondecreasing, $\varphi(\|x-y\|) \geq \varphi(|\|x\| - \|y\||)$ with equality if $y = \frac{\|y\|}{\|x\|} x$.

Since φ is convex it follows from (3.3), Proposition 2.1 and Remark 1, that $(\|X\|, \|Y\|)$ is an o.c.($\tilde{\varphi}$) where $\tilde{\varphi}(a,b) = \varphi(|a-b|)$ and, therefore, (X,Y) is an o.c.(σ_φ).

■

Corollary 3.3 *If $Y = \phi(X)$ λ-a.s. with ϕ a radial transformation on a normed space $(U, \|\cdot\|)$, then $(X, \phi(X))$ is an o.c.(σ_φ).*

Corollary 3.4 *If $U = \mathcal{H}$ is a Hilbert space, φ is strictly convex and $X \stackrel{d}{=} P$, $Y \stackrel{d}{=} Q$, then*

a)

$$E\sigma_\varphi(X,Y) \geq \tilde{\varphi}\left(P^{\|\,\|}, Q^{\|\,\|}\right), \tag{3.4}$$

and we have equality if and only if (3.2) and (3.3) hold.

b) If ϕ is a radial transformation, $Y \stackrel{d}{=} \phi(X)$ and $P^{\|\,\|}$ is nonatomic, then (X,Y) is an o.c.(σ_φ), if and only if

$$Y = \phi(X), \ \lambda\text{-a.s.}$$

PROOF.- Note that if (X,Y) is an o.c.(σ_φ), then

$$\sigma_\varphi(P,Q) = E\left[\varphi\|X-Y\|\right] \geq E\varphi\left[|\|X\| - \|Y\||\right] \geq \tilde{\varphi}\left(P^{\|\,\|}, Q^{\|\,\|}\right) \tag{3.5}$$

and (3.4) is proved.

Moreover, φ strictly convex implies that φ is strictly increasing. So the first inequality in (3.5) is an equality if and only if (3.2) holds and the second inequality is another equality if and only if (3.3) is verified (Proposition 2.1, Remark 1). So a) is proved.

With respect to b), as the pair (X,ϕ(X)) verifies (3.2) and (3.3), then (X,ϕ(X)) is an o.c.(σ_φ), the pair ($\|X\|, \|Y\|$) is an o.c.($\tilde{\phi}$) and (3.5) is an equality in this case. Therefore, if (X,Y) is an o.c.(σ_φ), then $Y = \frac{\|Y\|}{\|X\|}X$ because φ is strictly increasing.

But the o.c.($\tilde{\varphi}$) is unique (Proposition 2.1 and Remark 1). So $\|Y\| = \|\phi(X)\| = \alpha\left(\|X\|\right)$ $\lambda-$a.s., and the proof ends.

∎

Example 3.5 Here we propose three examples of applications of Theorem 3.2.

1. Spherically invariant distributions.

$P \in \mathcal{M}^1(\Re^n, \mathcal{L})$ is called spherically invariant if P is invariant with respect to orthogonal transformations on $\Re^n$. It is well-known that P is spherically invariant if and only if the conditional distributions

$$P\left(\cdot \mid \|x\| = t\right) = U_t$$

are the uniform distribution on $\{x : \|x\| = t\}$, ($\| \; \|$ denotes the euclidean norm on $\Re^n$). Typical examples of spherically invariant distributions are those which admit a density with respect to the Lebesgue measure which only depends on x through $\|x\|$.

Since

$$P = \int U_t dP^{\| \, \|}(t),\; Q = \int U_t dQ^{\| \, \|}(t)$$

if P and Q are spherically invariant distributions, there exist r.v.'s X $\stackrel{\mathrm{d}}{=}$P and Y $\stackrel{\mathrm{d}}{=}$Q with $\|X\| \stackrel{\text{s.o.}}{\sim} \|Y\|$ and $Y = \frac{\|Y\|}{\|X\|}X$ and, therefore, (X,Y) is an o.c.(σ_φ). A direct argument is as follows. If X $\stackrel{\mathrm{d}}{=}$P, then with $U := \frac{X}{\|X\|}$ (assuming $X \neq 0$, λ-a.s.) we have $X = U \cdot \|X\|$ and U, $\|X\|$ are independent. Define $Y := Z \cdot U$, where $Z \stackrel{\mathrm{d}}{=} Q^{\| \, \|}$, $\|X\| \stackrel{\text{s.o.}}{\sim} Z$ and Z, U are independent. Then Y $\stackrel{\mathrm{d}}{=}$Q and $Y = \frac{\|Y\|}{\|X\|}X$.

Moreover if φ is strictly convex and $P^{\| \, \|}$ is non-atomic, then the o.c.(σ_φ) is unique and is given by the radial transformation $Y = \phi(X)$, where

$$\phi(x) = \frac{\alpha\left(\|x\|\right)}{\|x\|}x,\; \alpha(t) := F^{-1}_{\|Y\|} \circ F_{\|X\|}(t). \tag{3.6}$$

In this way we obtain, for instance, an o.c.(σ_φ) between the uniform distribution on the unit ball $B(0,1) = \{x \in \Re^n : \|x\| \leq 1\}$, U(B(0,1)) and the normal distribution with covariance matrix $I_n = I$, $N(0,I) = \otimes_{i=1}^n N(0,1)$, by a radial transformation with α determined from (3.1) which is independent of φ. Note that $F_{\|Y\|^2}$ is a χ^2_n-distribution and that

$$F_{\|X\|}(t) = \frac{\text{Volume}\,(B(0;t))}{\text{Volume}\,(B(0,1))} = t^n,\ 0<t<1.$$

So we obtain that the

$$W^2(U(B(0,1)), N(0;I)) = E\|X-\phi(X)\|^2 =$$

$$E\,|\|X\| - \alpha(\|X\|)|^2 = \int_0^1 |t-\alpha(t)|^2 n t^{n-1} dt,$$

where, from (3.6), we have that

$$\alpha(t) = F^{-1}_{\chi^2_n}[F^2_{\|X\|}(t)] = F^{-1}_{\chi^2_n}(t^{2n}),\ 0<t<1.$$

2. Spherically equivalent distributions.

Given $P, Q \in \mathcal{M}^1(U,\mathcal{L})$ where $\mathcal{L}$ is the Borel σ-algebra in the normed space (U,$\|\cdot\|$) we denote by p_t and q_t, $t\in(0,1)$ a t-quantile of the distributions $P^{\|\,\|}$ and $Q^{\|\,\|}$ respectively. Then we call P and Q spherically equivalent, if

$$P\left(A \mid \|x\| = p_t\right) = Q\left(\frac{q_t}{p_t}A \mid \|x\| = q_t\right),\ \forall t$$

where A is any Borel set.

In particular, two spherically invariant distributions are spherically equivalent.

First we show that Theorem 3.2 applies to spherically equivalent distributions.

Proposition 3.6 *Let P and Q be spherically equivalent distributions. Then, there exist* $X\stackrel{d}{=}P$ *and* $Y\stackrel{d}{=}Q$ *such that* $\|X\| \stackrel{s.o.}{\sim} \|Y\|$ *and* $Y = \frac{\|Y\|}{\|X\|}X$.

PROOF.- Let $X\stackrel{\mathrm{d}}{=}P$ and let us denote by F_1 and F_2 the d.f.'s of $P^{\|\,\|}$ and $Q^{\|\,\|}$ respectively.

Let $U\stackrel{\mathrm{d}}{=}U(0,1)$ be independent from X and define:

$$\alpha(\|X\|, U) := F_2^{-1}\left(\mathcal{F}_1(\|X\|, U)\right) \text{ and } Y := \frac{\alpha(\|X\|, U)}{\|X\|}X$$

where F_2^{-1} and $\mathcal{F}$ are defined as in Section 2.

Then $\|Y\|\stackrel{\mathrm{d}}{=}Q^{\|\,\|}$ and $\|X\| \stackrel{\mathrm{s.o.}}{\sim} \|Y\|$.

Let $t^* \in (0,1)$. If we show that

$$\lambda\left\{Y \in A \mid \|Y\| = q_{t^*}\right\} = Q(A \mid \|x\| = q_{t^*}) \tag{3.7}$$

the proposition will be proved. To this end let t_0^*, t_1^* be such that $[t_0^*, t_1^*] = \{t : q_t = q_{t^*}\}$. If this interval is degenerated (i.e. $t_0^* = t_1^* = t^*$) and if we denote $[t_0, t_1] = \{t : p_t = p_{t^*}\}$, then

$$\{\|Y\| = q_{t^*}\} = \left\{\|X\| = p_{t^*}\,;\, U = \frac{t^* - t_0}{t_1 - t_0}\right\}$$

and for every $A \in \mathcal{L}$:

$$\begin{aligned}\lambda\{Y \in A \mid \|Y\| = q_{t^*}\} &= \lambda\left\{\frac{\alpha(\|X\|, U)}{\|X\|}X \in A \,\middle|\, \|X\| = p_{t^*}\,;\, U = \frac{t^* - t_0}{t_1 - t_0}\right\} \\ &= \lambda\left\{\frac{q_{t^*}}{p_{t^*}}X \in A \,\middle|\, \|X\| = p_{t^*}\right\} = Q(A \mid \|x\| = q_{t^*})\end{aligned}$$

where the second equality holds because X and U are independent.

Then, let us suppose that $t_0^* < t_1^*$ and let $[t_0, t_1] = \{t : p_t \in [p_{t_0^*}, p_{t_1^*}]\}$. It is evident that $t_0 \le t_0^*$, $t_1 \ge t_1^*$, $p_{t_0} = p_{t_0^*}, p_{t_1} = p_{t_1^*}$ and that there exist $a, b \in (0, 1)$ such that

$$(1 - a)\lambda\{\|X\| = p_{t_0}\} + \lambda\{\|X\| \in (p_{t_0}, p_{t_1})\} + b\lambda\{\|X\| = p_{t_1}\} = t_1^* - t_0^*.$$

Moreover, if we fix a and b through the relation

$$\lambda\{\|X\| < p_{t_1}\} + b\lambda\{\|X\| = p_{t_1}\} = t_1^*$$

we obtain that

$\{\|Y\| = q_{t^*}\} =$

$$\{\|X\| = p_{t_0}\,;\, U \ge 1 - a\} \cup \{\|X\| \in (p_{t_0}, p_{t_1})\} \cup \{\|X\| = p_{t_1}\,;\, U \le b\}$$

Therefore, if $A \in \mathcal{L}$:

$$\begin{aligned}\lambda\{Y \in A \mid \|Y\| = q_{t^*}\} &= \frac{1}{t_1^* - t_0^*}\lambda\{Y \in A\,;\, \|Y\| = q_{t^*}\} \\ &= \frac{1}{t_1^* - t_0^*}\left[\lambda\left\{\frac{q_{t^*}}{p_{t_0}}X \in A\,;\, \|X\| = p_{t_0}\,;\, U \ge 1 - a\right\}\right. \\ &\quad +\lambda\left\{\frac{q_{t^*}}{\|X\|}X \in A\,;\, \|X\| \in (p_{t_0^*}, p_{t_1^*})\right\} \qquad (3.8) \\ &\quad \left.+\lambda\left\{\frac{q_{t^*}}{p_{t_1}}X \in A\,;\, \|X\| = p_{t_1}\,;\, U \le b\right\}\right]\end{aligned}$$

where we have used that $p_{t_0} = p_{t_0^*}$ and $p_{t_1} = p_{t_1^*}$. Now we consider each term separately

$$\lambda\left\{\frac{q_{t^*}}{\|X\|}X \in A\,;\, \|X\| \in (p_{t_0^*}, p_{t_1^*})\right\} = \int_{(p_{t_0^*}, p_{t_1^*})} \lambda\left\{\frac{q_t}{p_t}X \in A \,\middle|\, \|X\| = p_t\right\} dP^{\|X\|}(p_t)$$

$$= \int_{(p_{t_0^*}, p_{t_1^*})} Q(A \mid \|x\| = q_t)P^{\|X\|}d(p_t) = Q(A \mid \|x\| = q_t)\lambda\{\|X\| \in (p_{t_0}, p_{t_1})\}$$

The first term gives:

$$\lambda\left\{\frac{q_{t^*}}{p_{t_0}}X \in A\,;\ \|X\| = p_{t_0}\,;\ U \geq 1-a\right\}$$
$$= (1-a)\lambda\left\{\frac{q_{t_0^*}}{p_{t_0^*}}X \in A \,\middle|\, \|X\| = p_{t_0^*}\right\}\lambda\left\{\|X\| = p_{t_0}\right\}$$
$$= (1-a)Q(A \mid \|x\| = q_{t^*})\lambda\left\{\|X\| = p_{t_0}\right\}$$

A similar reasoning with the last term and (3.8) gives us (3.7) and the proof ends.

■

Therefore, if P and Q are spherically equivalent, then Theorem 3.2 yields optimal couplings uniformly for all distances σ_φ. If $P^{\|\,\|}$ is continuous, an optimal coupling is obtained by a radial transformation as in (3.6). Moreover, according with (3.1), if ϕ is a radial transformation, X and $\phi(X)$ are spherically equivalent.

The following example shows that the spherically equivalent distributions are not the only ones to which Theorem 3.2 is applicable.

Example 3.5.1 Suppose that $U = \Re^2$ is supplied with the euclidean norm and that P and Q are uniform distributions on a circumference and on an ellipse respectively. Then, an o.c(W) between P and Q can be written as $(X, A(X)X)$, with $A(X) \in \Re_+$, but P and Q are not spherically equivalent in spite of the fact that $\|X\| \overset{\text{s.o.}}{\sim} \|A(X)X\|$.

In the n-dimensional case, if $\|\,\|$ is the euclidean norm, there exists a very simple characterization of spherically equivalence.

Proposition 3.7 *Let $X \overset{d}{=} P$ and $Y \overset{d}{=} Q$ and denote $\psi(X) = (\|X\|, \phi_1, ..., \phi_{n-1})$ and $\psi(Y) = (\|Y\|, \psi_1, ..., \psi_{n-1})$ the transformations of X and Y to polar coordinates respectively. Then P and Q are spherically equivalent if and only if $\psi(X)$ and $\psi(Y)$ have the same dependence structure and*

$$\phi_i \overset{d}{=} \psi_i,\ 1 \leq i \leq n-1. \tag{3.9}$$

This Proposition can be used to construct examples of spherically equivalent distributions. It also gives the following corollary. Note that we obtain as consequence of the proposition that if P and Q are spherically equivalent and $X \overset{\text{d}}{=} P$ and $Y \overset{\text{d}}{=} Q$ then there exists increasing functions, $T, f_1, ..., f_{n-1} : \Re \to \Re$ such that

$$\psi(Y) \overset{\text{d}}{=} (T\left(\|X\|\right), f_1(\phi_1), ..., f_{n-1}(\phi_{n-1})).$$

By (3.9), we can assume that $f_i = Id$, $i = 1, ..., n-1$. So, if we consider polar coordinates in $\Re^n$, and we assume that the distribution of $\|X\|$ is continuous, we have the following characterization of the class of the distributions which are spherically equivalent to a given one.

Corollary 3.8 *Let* $X \stackrel{d}{=} P$ *and let* $(\|X\|, \phi_1, ..., \phi_{n-1})$ *be the representation of* X *in polar coordinates. If* $P^{\|\,\|}$ *is continuous, we have that* P *and* Q *are spherically equivalent if and only if there exists an increasing function* $T : \Re \to \Re$ *such that* $(T(\|X\|), \phi_1, ..., \phi_{n-1}) \stackrel{d}{=} Q$*.*

It is not difficult to find examples of spherically equivalent distributions which are not related through radial transformations and viceversa; but in the next proposition we prove that there exists a strong link between both notions.

Proposition 3.9 *Let* P *and* Q *be two probability distributions and let* $X \stackrel{d}{=} P$*. Then:*

1. *If* P *and* Q *are spherically equivalent and the distribution of* $\|X\|$ *is continuous,then there exists a radial transformation* ϕ *such that* $\phi(X) \stackrel{d}{=} Q$*.*
2. *If* ϕ *is a radial transformation such that* $\phi(X) \stackrel{d}{=} Q$ *and* α *is strictly increasing, then* P *and* Q *are spherically equivalent distributions.*

PROOF.- Item 1 is a consequence of the Proposition 3.6 because in this case the function $\alpha(\|X\|, U)$ in the proof of this Proposition does not depend on U.

With respect to 2 note that $\|Y\| = \alpha(\|X\|)$ and that for q $\in \mathrm{Im}(\alpha)$ there exists an unique p such that $q = \alpha(p)$. So we have that

$$\lambda\left\{Y \in \frac{q}{p}A \,|\|Y\| = q\right\} = \lambda\left\{\frac{\alpha(\|X\|)}{\|X\|}X \in \frac{q}{p}A \,|\alpha(\|X\|) = q\right\} =$$

$$\lambda\left\{\frac{X}{\|X\|} \in \frac{1}{p}A \,|\alpha(\|X\|) = \alpha(p)\right\} = \lambda\left\{X \in A|\|X\| = p\right\}.$$

∎

As an application consider the p-norms on $\Re^n$

$$\|x\|_p = \begin{cases} \left(\sum |\, x_i \,|^p\right)^{1/p}, & \text{if } 1 \leq p < \infty \\ \max\{|\, x_i \,|\}, & \text{if } p = \infty. \end{cases}$$

Let P be the uniform distribution on the p-ball of radius 1 in $\Re^n_+$,

$$B_p(0,1) = \{x \in \Re^n_+ : \|x\|_p \leq 1\}$$

for $0 < p < \infty$, let $Q_{\lambda,p} := \otimes_{i=1}^n W_{\lambda,p}$, where $W_{\lambda,p}$ is a Weibull-distribution with density

$$f_{\lambda,p}(x) = \frac{\lambda^{1/p} p}{\Gamma(1/p)} e^{-\lambda x^p},\ x \geq 0,$$

and for $p = \infty$, let $Q_{\lambda,\infty} := \otimes_{i=1}^n U(0,\lambda)$ denote the uniform distribution on $(0,\lambda)^n$. Then an o.c.(σ_φ), $\sigma_\varphi(x,y) = \varphi(\|x-y\|_p)$, between P and $Q_{\lambda,p}$ is obtained by a radial transformation.

3. Norm weighted processes.

Let $X = \{X_t\}_{t\in[0,1]}$ be a stochastic process in $L^p([0,1])$, respectively $D([0,1])$, with norm $\|\cdot\|_p$, resp. $\|\cdot\|_\infty$, and define $Y = \{Y_t\}_{t\in[0,1]}$ by $Y_t = \|X\|_p^r X_t$, $r > -1$. So each path of X is weighted by its norm. From Theorem 3.2, (X,Y) is an optimal coupling between P_X and P_Y with respect to $\sigma_\varphi(x,y) = \varphi(\|x-y\|_p)$, $0 < p \leq \infty$. In particular, in this way one obtain optimal couplings between spherically invariant processes and more generally spherically equivalent processes.

Next we analyze some possible generalizations of the preceeding results.

The first conclusion is negative but, in some way, surprising. In a Hilbert space, if $\phi(x) = \frac{\alpha(A(x))}{B(x)}x$, the only possibility to assure that $(X,\phi(X))$ is an o.c. is to take $A(x) = f(\|x\|)$ if we restrict the selection of A to a suitable class. The key is the following proposition:

Proposition 3.10 *Let $(\mathcal{H};\langle\cdot,\cdot\rangle)$ be a Hilbert space and let $A : \mathcal{H} \to \Re_+$ be a continuous function such that there exist two points x and y verifying that*

$$A(y) < A(x) \text{ and } \langle x, y-x\rangle > 0.$$

If $B : \mathcal{H} \to \Re_+$ is continuous, then there exist an increasing map $\alpha : \Re_+ \to \Re_+$ such that for every r.v., X, with support equal to $\mathcal{H}$, $\left(X, \frac{\alpha(A(X))}{B(X)}X\right)$ is not an o.c.(W).

PROOF.- Trivially

$$\left\langle \frac{\alpha(A(y))}{B(y)}y - \frac{\alpha(A(x))}{B(x)}x, y-x \right\rangle = \frac{\alpha(A(y))}{B(y)}\langle y, y-x\rangle - \frac{\alpha(A(x))}{B(x)}\langle x, y-x\rangle.$$

The second term is negative. So, for a suitable increasing function, α, the sum is also negative. The continuity of A and B and the asumption on the support of X, implies that the map $x \to \frac{\alpha(A(x))}{B(x)}x$ is not λ-a.s. increasing, and then we have the result by Theorem 2.3 in [3].

■

This result includes a lot of cases, as the next proposition and corollaries show.

Proposition 3.11 *Let $A : \mathcal{H} \to \Re_+$ be a continuous function such that*

1. *The restriction of A to the set $\{\delta v : \delta \geq 0\}$ is strictly increasing for every $v \in \mathcal{H}$, and*
2. *If $A(y) < A(x)$ then $\langle x, y-x\rangle \leq 0$.*

Then $A(x) = f(\|x\|)$ for some strictly increasing function f.

PROOF.- In this proof we are going to denote by $\mathcal{L}[a,b,..]$ (resp. $\mathcal{L}[a,b,..]^\perp$) the linear subspace generated by the vectors {a,b,...} (resp. the ortogonal subspace to the vectors {a,b,..,}).

Note that if we show that for each $x, y \in \mathcal{H}$ with $A(x) = A(y)$ holds $\|x\| = \|y\|$ then we have as consequence $A(x) = f(\|x\|)$ with f strictly increasing by 1.

Therefore, let $x, y \in \mathcal{H}$ such that $A(x) = A(y)$ and $x \neq y$. 1. implies that $\theta = \text{ang}(x,y) \neq 0$. Let $n \in \mathcal{N}$, and define $x_n^0 = x$ and x_n^i be the vector in $\mathcal{L}[x,y] \cap \{x_n^{i-1} + \mathcal{L}[x_n^{i-1}]^\perp\}$ such that $\text{ang}(x_n^i, x_n^{i-1}) = \theta/n$, $1 \leq i \leq n$ and $x_n^n = \alpha y$; $\alpha > 0$. Therefore, $\langle x_n^{i-1}, x_n^i - x_n^{i-1}\rangle = 0$. This, 2. and the continuity of A imply that $A(x_n^i) \geq A(x_n^{i-1})$. So, $A(y) \leq A(x_n^n)$, and 1. implies that

$$\|y\| \leq \|x_n^n\|. \tag{3.10}$$

On the other hand, it is evident that $\|x_n^i\| = \frac{\|x_n^{i-1}\|}{\cos\theta/n}$; therefore, $\|x_n^n\| = \frac{\|x\|}{(\cos\theta/n)^n}$, which jointly with (3.10) gives that

$$\|y\| \leq \lim_n \frac{\|x\|}{(\cos\theta/n)^n} = \|x\|$$

and the proof ends.

■

Corollary 3.12 *Let $\phi : \mathcal{H} \to \mathcal{H}$ be such that $\phi(x) = \frac{\alpha(A(x))}{B(x)} x$ where α, A and B verify the conditions in Proposition 3.10 and 3.11. If ϕ verifies that $(X, \phi(X))$ is an o.c.($\| \ \|$) for every r.v. X, then ϕ is a radial transformation with respect to the norm $\| \ \|$.*

With this corollary we obtain that the condition on the support of X in Proposition 3.10 can not be deleted. For instance, consider again the example 3.5.1. There an o.c.(W) between P and Q could be written as $(X, A(X)X)$ but $A(X) \neq \alpha(\|X\|)$.

Corollary 3.13 *If $\| \ \|$ is a norm on $\Re^n$ different from the euclidean one, there exists $\alpha : \Re^+ \to \Re^+$, increasing, and a r.v., X, such that $\left(X, \frac{\alpha(\|X\|)}{\|X\|} X\right)$ is not an o.c.(W) with respect to the euclidean norm.*

In particular, if A is a symmetric, positive definite matrix different from the identity and $\|x\|_A = x^t A x$, then, for some r.v., X, the pair $\left(X, \frac{\alpha(\|X\|_A)}{\|X\|_A} X\right)$ is not an o.c.(W) with respect to the euclidean norm, but, according to Corollary 3.3, it is an o.c.(W) with respect to the norm $\| \ \|_A$.

The next result is also negative. It could be suspected that the composition of o.c.(W)'s gives a new o.c.(W). If this were true then, $\left(X, \frac{\alpha(\|X\|)}{\|X\|} AX\right)$ should be an o.c.(W) because in Proposition 3.1 and Theorem 2.10 we have proved that, if $\alpha : \Re \to \Re$ is increasing and A is a symmetric, positive semidefinite $n \times n$ matrix, then $\left(X, \frac{\alpha(\|X\|)}{\|X\|}X\right)$ and $\left(\frac{\alpha(\|X\|)}{\|X\|}X, \frac{\alpha(\|X\|)}{\|X\|}AX\right)$ are o.c.(W). But, we have the following proposition.

Proposition 3.14 *Let A be a symmetric, positive definite, $n\times n$ matrix. There exists an increasing map $\alpha : \Re_+ \to \Re_+$ such that if X is a r.v. with support containing $\Re^n_+$, then $\left(X, \frac{\alpha(\|X\|)}{\|X\|}AX\right)$ is not an o.c.(W).*

PROOF.- Let $H(x) = \frac{\alpha(\|x\|)}{\|x\|}Ax$ and $\{e_i\}_{1\leq i\leq n}$ be an orthonormal basis of eigenvectors. Suposse $\delta_1 \geq \delta_2$ where δ_1, δ_2 are eigenvalues of A. If $x = e_1 + e_2$, $y = \frac{1}{4}e_1 + \frac{6}{4}e_2$ then $\|x\|^2 = 2 < \|y\|^2 = \frac{37}{16}$. Also, $\langle y - x, Ax\rangle = -\frac{3}{4}\delta_1 + \frac{2}{4}\delta_2 < 0$. So, if we fixe $\alpha(\|x\|)$ and take $\alpha(\|y\|)$ big enough then $H(y) - H(x)$ is close to Ax, and H is not increasing. Therefore, (X,H(X)) is not an o.c. (see Theorem 2.3 in [3]).

■

In the following propositions we construct some transformations similar to the radial ones which have the optimal coupling property for W-coupling with respect to the euclidean distance.

Proposition 3.15 *For any symmetric, positive semidefinite matrix A and any r.v. X on $\Re^n$,*

$$\left(X, \frac{\alpha(\|X\|_A)}{\|X\|_A}AX\right) \tag{3.11}$$

is an o.c.(W) with respect to the euclidean norm, where $\|x\|_A = (x^tAx)^{1/2}$.

PROOF.- Define $f(x) := F(\|x\|_A)$, $F(t) = \int_0^t \alpha(s)\mathrm{d}s$. Then f is convex and, similarly to the proof of Proposition 3.1, $\frac{\alpha(\|x\|_A)}{\|x\|_A}Ax \in \partial f(x)$ which implies (3.11) by (1.1).

■

Proposition 3.16 *For $1 \leq p < \infty$, let*

$$\phi(x) = \frac{\alpha(\|x\|_p)}{\|x\|_p}\left(x_1 \mid x_1 \mid^{p-2}, \ldots, x_n \mid x_n \mid^{p-2}\right),\ x \in \Re^n$$

with α increasing. Then, for any r.v. X the pair $(X, \phi(X))$ is an o.c.(W) with respect to the euclidean norm.

PROOF.- With $F(t) = \int_0^t \alpha(s)\mathrm{d}s$ define $f(x) := F(\|x\|_p)$. Then f is convex and $\phi(x) \in \partial f(x)$ which implies Proposition 3.16.

■

Remark 4

a) For $0 < p < 1$, $\|x\|_p = \sum | x_i |^p$ and, therefore, for nondecreasing α

$$\phi(x) = \alpha\left(\|x\|_p\right)\Big(| x_1 |^{p-1} \operatorname{sign}(x_1), \ldots, | x_n |^{p-1} \operatorname{sign}(x_n)\Big), \; x \in \Re^n$$

has the o.c. property w.r.t. W and the euclidean distance.

b) For $p = \infty$, $\|x\|_\infty = \max | x_i |$, if we consider $f(x) = F\left(\|x\|_\infty\right)$ we obtain that

$$\phi(x) = \alpha\left(\|x\|_\infty\right) \operatorname{sign}(x_j) e,$$

for any e in the convex hull of e_j, over all j with $| x_j | = \|x\|_\infty$, has the o.c. property w.r.t. W and the euclidean distance.

Proposition 3.17 *The transformation* $\phi(x) = \left(\frac{e^{x_i}}{\sum e^{x_j}}\right)$ *from* $\Re^n_+$ *to the unit simplex* $\{x \in \Re^n_+ : \sum_{i=1}^n x_i = 1\}$ *has the o.c. property w.r.t. W and the euclidean distance.*

PROOF.- Take the convex function $f(x) = \ln\left(\sum e^{x_i}\right)$, then $\phi(x) = \nabla f(x)$ which implies optimality.

■

Note that for differentiable transformations of the form $\psi(x) = \left(\frac{h(x_i)}{\sum h(x_j)}\right)$ in the unit simplex, the transformation ϕ of Proposition 3.17 is the only optimal one. This follows from the symmetry condition $\frac{\partial \psi_i}{\partial x_j} = \frac{\partial \psi_j}{\partial x_i}$ (see (1.2)) which is satisfied only for ϕ.

Acknowledgements. An essential part of this paper was established during a visit of L. Rüschendorf to the Universidad de Cantabria. This visit and the authors J. Cuesta-Albertos and A. Tuero-Diaz have been partially supported by the grants number PB87-0905-C02-00 and 02 and PB91-0306-C02-01 of the Spanish DGICYT.

The authors want to thank to C. Matrán-Bea for helpful comments and suggestions related to this work and to a referee for some hints which have made the paper more readable; in particular items 1 and 2 in Remark 1 have been suggested by him.

References

[1] CAMBANIS, S., SIMONS, G. and STOUT, W. (1976). Inequalities for $\mathcal{E}$k(X,Y) when the marginals are fixed. *Z. Wahrsch. Verw. Geb.* 36, 285-294.

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

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

[4] CUESTA, J.A. and MATRAN, C. (1991). Skorohod representation theorem and Wasserstein metrics. Preprint.

[5] DALL'AGLIO, G. (1991). Fréchet classes: The beginnings. In: *Advances in probability distributions with given marginals.* ed.G. Dall'Aglio, S. Kotz and G. Salinetti. pg. 1-12.

[6] DOWSON, D.C. and LANDAU, B.V. (1982). The Fréchet distance between multivariate normal distributions. *J. Multivariate. Anal.* 12, 450-455.

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

[8] IOFFE, A.D. and TICHOMIROV, V.M. (1979). Theorie der Extremalaufgaben. *VEB* .

[9] KANTOROVICH, L.V. (1942). On the transfer of masses. Dokc. Akad. Nauk SSSR 37, 7-8.

[10] KELLERER, H. (1984). Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete* 67, 399-432.

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

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

[13] RACHEV, S.T. (1984). The Monge-Kantorovich mass transference problem and its stochastic applications. *Theory Probab. Appl.* 29, 647-676.

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

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

[16] RÜSCHENDORF, L. (1981). Sharpness of Fréchet bounds. *Z. Wahrsch. Verw. Gebiete* 57, 293-302.

[17] RÜSCHENDORF, L. (1981). Stochastically ordered distributions and monotonicity of the OC-fuction of sequential probability ratio tests. *Math. Operationsf., Series Statistics* 12, 327-338.

[18] RÜSCHENDORF, L. (1985). The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Gebiete* 70, 117-129.

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

[20] 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.

[21] RÜSCHENDORF, L. (1991). Bounds for distributions with multivariate marginals. *Stochastic Order and Decisions under Risk.* Ed. K. Mosler and M. Scarsini. I.M.S. Lecture Notes 19, 285-310.

[22] 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.

[23] TANAKA, H. (1973). An inequality of probability distributions and its application to Kac's one-dimensional model of a maxwellian gas. *Z. Wahrsch. Verw. Gebiete* 27, 47-52.

[24] TUERO, A. (1991). Aplicaciones crecientes. Relaciones con las métricas de Wasserstein. *Dissertation*. Universidad de Santander.