A remark on the optimal transport between two probability measures sharing the same copula

Aurélien Alfonsi, Benjamin Jourdain

A remark on the optimal transport between two probability measures sharing the same copula

A. Alfonsi, B. Jourdain*

July 16, 2013

Abstract

We are interested in the Wasserstein distance between two probability measures on $\mathbb{R}^n$ sharing the same copula C. The image of the probability measure dC by the vectors of pseudo-inverses of marginal distributions is a natural generalization of the coupling known to be optimal in dimension $n = 1$. It turns out that for cost functions $c(x, y)$ equal to the p-th power of the L^q norm of $x - y$ in $\mathbb{R}^n$, this coupling is optimal only when $p = q$ i.e. when $c(x, y)$ may be decomposed as the sum of coordinate-wise costs.

Keywords: Optimal transport, Copula, Wasserstein distance, Inversion of the cumulative distribution function.

AMS Classification (2010): 60E05, 60E15.

1 Optimal transport between two probability measures sharing the same copula

Given two probability measures μ and ρ, the optimal transport theory aims at minimizing $\int c(x, y)\nu(dx, dy)$ over all couplings ν with first marginal $\nu \circ ((x, y) \mapsto x)^{-1} = \mu$ and second marginal $\nu \circ ((x, y) \mapsto y)^{-1} = \rho$ for a measurable non-negative cost function c. We use the notation $\nu <^{\mu}_{\rho}$ for such couplings. In the present note, we are interested in the particular case of the so-called Wasserstein distance between two probability measures μ and ρ on $\mathbb{R}^n$:

$$\mathcal{W}_{p,q}(\mu, \rho) = \inf_{\nu <^{\mu}_{\rho}} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|_q^p \nu(dx, dy) \right)^{1/p} \tag{1.1}$$

*Université Paris-Est, CERMICS, Projet MathFi ENPC-INRIA-UMLV, 6 et 8 avenue Blaise Pascal, 77455 Marne La Vallée, Cedex 2, France, e-mails : alfonsi@cermics.enpc.fr, jourdain@cermics.enpc.fr. This research benefited from the support of the "Chaire Risques Financiers", Fondation du Risque and the French National Research Agency (ANR) under the program ANR-12-BLAN Stab.

obtained for the choice $c(x,y)=\|x-y\|_q^p$. Here $\mathbb{R}^n$ is endowed with the norm $\|(x_1,\dots,x_n)\|_q = (\sum_{i=1}^n |x_i|^q)^{1/q}$ for $q\in[1,+\infty)$ whereas $p\in[1,+\infty)$ is the power of this norm in the cost function.

In dimension $n=1$, $\|x\|_q=|x|$ so that the Wasserstein distance does not depend on q and is simply denoted by $\mathcal{W}_p$. Moreover, the optimal transport is given by the inversion of the cumulative distribution functions : whatever $p\in[1,+\infty)$, the optimal coupling is the image of the Lebesgue measure on $(0,1)$ by $u\mapsto(F_\mu^{-1}(u),F_\rho^{-1}(u))$ where for $u\in(0,1)$, $F_\mu^{-1}(u)=\inf\{x\in\mathbb{R}:\mu((-\infty,x])\geq u\}$ and $F_\rho^{-1}(u)=\inf\{x\in\mathbb{R}:\rho((-\infty,x])\geq u\}$ (see for instance Theorem 3.1.2 in [3]). This implies that $\mathcal{W}_p^p(\mu,\rho)=\int_{(0,1)}|F_\mu^{-1}(u)-F_\rho^{-1}(u)|^p du$.

In higher dimensions, according to Sklar's theorem (see for instance Theorem 2.10.11 in Nelsen [1]),

$$\mu\left(\prod_{i=1}^n(-\infty,x_i]\right)=C\left(\mu_1((-\infty,x_1]),\dots,\mu_n((-\infty,x_n])\right)$$

where we denote by $\mu_i=\mu\circ((x_1,\dots,x_n)\mapsto x_i)^{-1}$ the i-th marginal of μ and C is a copula function i.e. $C(u_1,\dots,u_n)=m\left(\prod_{i=1}^n[0,u_i]\right)$ for some probability measure m on $[0,1]^n$ with all marginals equal to the Lebesgue measure on $[0,1]$. The copula function C is uniquely determined on the product of the ranges of the marginal cumulative distribution functions $x_i\mapsto\mu_i((-\infty,x_i])$. In particular, when the marginals μ_i do not weight points, the copula C is uniquely determined. Sklar's theorem shows that the dependence structure associated with μ is encoded in the copula function C. Last, we give the well-known Fréchet-Hoeffding bounds

$$\forall u_1,\dots,u_n\in[0,1],\ C_n^-(u_1,\dots,u_n)\leq C(u_1,\dots,u_n)\leq C_n^+(u_1,\dots,u_n)$$

that hold for any copula function C with $C_n^+(u_1,\dots,u_n)=\min(u_1,\dots,u_n)$ and $C_n^-(u_1,\dots,u_n)=(u_1+\dots+u_n-n+1)^+$ (see Nelsen [1], Theorem 2.10.12 or Rachev and Rüschendorf [3], section 3.6). We recall that the copula C_n^+ is the n-dimensional cumulative distribution function of the image of the Lebesgue measure on $[0,1]$ by $\mathbb{R}\ni x\mapsto(x,\dots,x)\in\mathbb{R}^n$. Also the copula C_2^- is the 2-dimensional cumulative distribution function of the image of the Lebesgue measure on $[0,1]$ by $\mathbb{R}\ni x\mapsto(x,1-x)\in\mathbb{R}^2$ and, for $n\geq3$, C_n^- is not a copula.

In dimension $n=1$, the unique copula function is $C(u)=u$ and therefore the optimal coupling between μ and ρ, which necessarily share this copula, is the image of the probability measure dC by $u\mapsto(F_\mu^{-1}(u),F_\rho^{-1}(u))$. It is therefore natural to wonder whether, when μ and ρ share the same copula C in higher dimensions, the optimal coupling is still the image of the probability measure dC by $(u_1,\dots,u_n)\mapsto(F_{\mu_1}^{-1}(u_1),\dots,F_{\mu_n}^{-1}(u_n),F_{\rho_1}^{-1}(u_1),\dots,F_{\rho_n}^{-1}(u_n))$. We denote by $\mu\diamond\rho$ this probability law on $\mathbb{R}^{2n}$. It turns out that the picture is more complicated than in dimension one because of the choice of the index q of the norm.

Proposition 1.1 *Let $n\geq2$, μ and ρ be two probability measures on $\mathbb{R}^n$ sharing the same copula C and $\mathcal{W}_{p,q}(\mu,\rho)=\inf_{\nu<^\mu_\rho}\left(\int_{\mathbb{R}^n\times\mathbb{R}^n}\|x-y\|_q^p\nu(dx,dy)\right)^{1/p}$.*

- *If $p=q$, then an optimal coupling between μ and ρ is given by $\nu=\mu\diamond\rho$ and*

$$\mathcal{W}_{p,p}^p(\mu,\rho)=\int_{[0,1]^n}\sum_{i=1}^n|F_{\mu_i}^{-1}(u_i)-F_{\rho_i}^{-1}(u_i)|^p dC(u_1,\dots,u_n)=\int_{[0,1]}\sum_{i=1}^n|F_{\mu_i}^{-1}(u)-F_{\rho_i}^{-1}(u)|^p du.$$

- *If $p \neq q$, the coupling $\mu \diamond \rho$ is in general no longer optimal. For $p < q$, if $C \neq C_n^+$, we can construct probability measures μ and ρ on $\mathbb{R}^n$ admitting C as their unique copula such that*
$$\left(\int_{\mathbb{R}^n \times \mathbb{R}^n} \|x-y\|_q^p \mu \diamond \rho(dx,dy)\right)^{1/p} > \mathcal{W}_{p,q}(\mu,\rho).$$
For $p > q$, the same conclusion holds if $n \geq 3$ or $n = 2$ and $C \neq C_2^-$.

Remark 1.2 *Let μ and ρ be two probability measures on $\mathbb{R}^n$ and $\nu <^\mu_\rho$. For $n = 1$, ν is said to be comonotonic if $\nu((-\infty,x],(-\infty,y]) = C_2^+(\mu((-\infty,x]),\rho((-\infty,y]))$. Puccetti and Scarsini [2] investigate several extensions of this notion for $n \geq 2$. In particular, they say that ν is π-comonotonic (resp. c-comonotonic) if μ and ρ have a common copula and $\nu = \mu \diamond \rho$ (resp. ν maximizes $\int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x,y\rangle \tilde{\nu}(dx,dy)$ over all the coupling measures $\tilde{\nu} <^\mu_\rho$). Looking at some connections between their different definitions of comonotonicity, they show in Lemma 4.4 that π-comonotonicity implies c-comonotonicity. Since*
$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \|x-y\|_2^2 \tilde{\nu}(dx,dy) = \int_{\mathbb{R}^n} \|x\|_2^2 \mu(dx) + \int_{\mathbb{R}^n} \|y\|_2^2 \rho(dy) - 2\int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x,y\rangle \tilde{\nu}(dx,dy),$$
this yields our result in the case $p = q = 2$.

2 Proof of Proposition 1.1

The optimality in the case $q = p$, follows by choosing $d_1 = \ldots = d_n = d_1' = \ldots = d_n' = d_1'' = \ldots = d_n'' = 1$, $c_i(y_i,z_i) = |y_i - z_i|^p$, $\alpha = dC$, and $\varphi_i = F_{\mu_i}^{-1}$, $\psi_i = F_{\rho_i}^{-1}$ in the following Lemma.

Lemma 2.1 *Let α be a probability measure on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \ldots \times \mathbb{R}^{d_n}$ with respective marginals $\alpha_1, \ldots, \alpha_n$ on $\mathbb{R}^{d_1}, \ldots, \mathbb{R}^{d_n}$ and $\varphi_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d_i'}$, $\psi_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d_i''}$ and $c_i : \mathbb{R}^{d_i'} \times \mathbb{R}^{d_i''} \to \mathbb{R}_+$ be measurable functions such that*
$$\forall i \in \{1,\ldots,n\}, \quad \inf_{\nu_i <^{\alpha_i \circ \varphi_i^{-1}}_{\alpha_i \circ \psi_i^{-1}}} \int_{\mathbb{R}^{d_i'} \times \mathbb{R}^{d_i''}} c_i(y_i,z_i)\nu_i(dy_i,dz_i) = \int_{\mathbb{R}^{d_i}} c_i(\varphi_i(x_i),\psi_i(x_i))\alpha_i(dx_i). \quad (2.1)$$

Then setting $\varphi : x = (x_1,\ldots,x_n) \ni \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_n} \mapsto (\varphi_1(x_1),\ldots,\varphi_n(x_n)) \in \mathbb{R}^{d_1'+\ldots+d_n'}$ and $\psi : x \ni \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_n} \mapsto (\psi_1(x_1),\ldots,\psi_n(x_n)) \in \mathbb{R}^{d_1''+\ldots+d_n''}$, one has
$$\begin{aligned}&\inf_{\nu <^{\alpha\circ\varphi^{-1}}_{\alpha\circ\psi^{-1}}} \int_{\mathbb{R}^{d_1'+\ldots+d_n'} \times \mathbb{R}^{d_1''+\ldots+d_n''}} \sum_{i=1}^n c_i(y_i,z_i)\nu(dy,dz) \\ &\quad = \int_{\mathbb{R}^{d_1+\ldots+d_n}} \sum_{i=1}^n c_i(\varphi_i(x_i),\psi_i(x_i))\alpha(dx) = \sum_{i=1}^n \int_{\mathbb{R}^{d_i}} c_i(\varphi_i(x_i),\psi_i(x_i))\alpha_i(dx_i).\end{aligned}$$

Proof of Lemma 2.1. We give two alternative proofs of the Lemma. The first one is based on basic arguments.

$$
\begin{aligned}
&\int_{\mathbb{R}^{d_1+\ldots+d_n}} \sum_{i=1}^n c_i(\varphi_i(x_i), \psi_i(x_i))\alpha(dx) \\
&\qquad \geq \inf_{\nu < ^{\alpha\circ\varphi^{-1}}_{\alpha\circ\psi^{-1}}} \int_{\mathbb{R}^{d'_1+\ldots+d'_n}\times\mathbb{R}^{d''_1+\ldots+d''_n}} \sum_{i=1}^n c_i(y_i, z_i)\nu(dy, dz) \\
&\qquad \geq \sum_{i=1}^n \inf_{\nu < ^{\alpha\circ\varphi^{-1}}_{\alpha\circ\psi^{-1}}} \int_{\mathbb{R}^{d'_1+\ldots+d'_n}\times\mathbb{R}^{d''_1+\ldots+d''_n}} c_i(y_i, z_i)\nu(dy, dz) \\
&\qquad \geq \sum_{i=1}^n \inf_{\nu_i < ^{\alpha_i\circ\varphi_i^{-1}}_{\alpha_i\circ\psi_i^{-1}}} \int_{\mathbb{R}^{d'_i}\times\mathbb{R}^{d''_i}} c_i(y_i, z_i)\nu_i(dy_i, dz_i) \\
&\qquad = \sum_{i=1}^n \int_{\mathbb{R}^{d_i}} c_i(\varphi_i(x_i), \psi_i(x_i))\alpha_i(dx_i) \\
&\qquad = \int_{\mathbb{R}^{d_1+\ldots+d_n}} \sum_{i=1}^n c_i(\varphi_i(x_i), \psi_i(x_i))\alpha(dx),
\end{aligned}
$$

where we used that

- the probability measure $\alpha \circ (\varphi^{-1}, \psi^{-1})$ on $\mathbb{R}^{d'_1+\ldots+d'_n} \times \mathbb{R}^{d''_1+\ldots+d''_n}$ has respective marginals $\alpha \circ \varphi^{-1}$ and $\alpha \circ \psi^{-1}$ on $\mathbb{R}^{d'_1+\ldots+d'_n}$ and $\mathbb{R}^{d''_1+\ldots+d''_n}$, for the first inequality,
- the infimum of a sum is greater than the sum of infima, for the second inequality,
- the respective marginals of $\alpha \circ \varphi^{-1}$ and $\alpha \circ \psi^{-1}$ on $\mathbb{R}^{d'_i}$ and $\mathbb{R}^{d''_i}$ are $\alpha_i \circ \varphi_i^{-1}$ and $\alpha_i \circ \psi_i^{-1}$, for the third one,
- and the hypotheses for the first equality.

The second proof is given to illustrate the theory of optimal transport. It requires to make the following additional assumption on the cost function c_i, for all $1 \leq i \leq n$:

there exist measurable functions $g_i : \mathbb{R}^{d'_i} \to \mathbb{R}_+$ and $h_i : \mathbb{R}^{d''_i} \to \mathbb{R}_+$ such that

$$
c_i(y_i, z_i) \leq g_i(y_i) + h_i(z_i), \quad \int_{\mathbb{R}^{d_i}} g_i(\varphi_i(x_i))\alpha_i(dx_i) < \infty, \quad \int_{\mathbb{R}^{d_i}} h_i(\psi_i(x_i))\alpha_i(dx_i) < \infty. \tag{2.2}
$$

Basically, this assumption ensures that $\int_{\mathbb{R}^{d'_i}\times\mathbb{R}^{d''_i}} c_i(y_i, z_i)\nu_i(dy_i, dz_i) < \infty$ for any coupling $\nu_i < ^{\alpha_i\circ\varphi_i^{-1}}_{\alpha_i\circ\psi_i^{-1}}$.

We now introduce some definitions that are needed, and refer to the Section 3.3 of Rachev and Rüschendorf [3] for a full introduction. Let $\bar{c} : \mathbb{R}^{d'} \times \mathbb{R}^{d''} \to \mathbb{R}$. A function $f : \mathbb{R}^{d'} \to \bar{\mathbb{R}}$ is *$\bar{c}$-convex* if there is a function $a : \mathbb{R}^{d''} \to \bar{\mathbb{R}}$ such that $f(y) = \sup_{z\in\mathbb{R}^{d''}} \bar{c}(y, z) - a(z)$. For $y \in \mathbb{R}^{d'}$, we define the $\bar{c}$-subgradient:

$$
\partial_{\bar{c}} f(y) = \{z \in \mathbb{R}^{d''} \ s.t. \ \forall \tilde{y} \in \mathrm{dom} f, \ f(\tilde{y}) - f(y) \geq \bar{c}(\tilde{y}, z) - \bar{c}(y, z)\},
$$

where $\mathrm{dom} f = \{y \in \mathbb{R}^{d'}, f(y) < \infty\}$.

We are now in position to prove the result again. Let $X = (X_1, \dots, X_d)$ be a random variable with probability measure α. We define $Y_i = \varphi_i(X_i)$, $Z_i = \psi_i(X_i)$ and $\bar{c}_i = -c_i$. From (2.1), we know that (Y_i, Z_i) is an optimal coupling that maximizes $\mathbb{E}[\bar{c}_i(Y_i, Z_i)]$. From Theorem 3.3.11 of Rachev and Rüschendorf [3], this implies the existence of a $\bar{c}_i$-convex function $f_i : \mathbb{R}^{d'_i} \to \bar{\mathbb{R}}$ such that $Z_i \in \partial_{\bar{c}_i} f_i(Y_i)$. By definition of the $\bar{c}_i$-convexity, there exists a function $a_i : \mathbb{R}^{d''_i} \to \bar{\mathbb{R}}$ such that $f_i(y_i) = \sup_{z_i \in \mathbb{R}^{d''_i}} \bar{c}(y_i, z_i) - a(z_i)$. We define for $y = (y_1, \dots, y_n) \in \mathbb{R}^{d'_1} \times \dots \times \mathbb{R}^{d'_n}$ and $z = (z_1, \dots, z_n) \in \mathbb{R}^{d''_1} \times \dots \times \mathbb{R}^{d''_n}$,

$$f(y) = \sum_{i=1}^{n} f_i(y_i), \ a(z) = \sum_{i=1}^{n} a_i(z_i) \text{ and } \bar{c}(y, z) = \sum_{i=1}^{n} \bar{c}_i(y_i, z_i).$$

The function f is $\bar{c}$-convex since we clearly have $f(y) = \sup_{z \in \mathbb{R}^{d''_1 + \dots + d''_n}} \bar{c}(y, z) - a(z)$. Then, we have the straightforward inclusion:

$$\begin{aligned} \partial_{\bar{c}} f(y) = \{&(z_1, \dots, z_n) \in \mathbb{R}^{d''_1} \times \dots \times \mathbb{R}^{d''_n}, \\ &s.t. \ \forall \tilde{y} \in \mathrm{dom} f, \ \sum_{i=1}^{n} f_i(\tilde{y}_i) - f_i(y_i) \geq \sum_{i=1}^{n} \bar{c}_i(\tilde{y}_i, z_i) - \sum_{i=1}^{n} \bar{c}_i(y_i, z_i)\} \\ \supset \ &\partial_{\bar{c}_1} f_1(y_1) \times \dots \times \partial_{\bar{c}_n} f_n(y_n). \end{aligned}$$

This gives immediately $Y \in \partial_{\bar{c}} f(X)$. Using again Theorem 3.3.11 of [3], we get that the coupling (Y, Z) with law $\mu \diamond \rho$ is optimal in the sense that it maximizes $\mathbb{E}[\bar{c}(Y, Z)]$. ■

We now prove that the coupling $\mu \diamond \rho$ is in general no longer optimal when $q \neq p$. We first deal with the dimension $n = 2$. Given two copulas C_2 and C'_2 on $[0, 1]^2$, let (U_1, U'_2) be distributed according dC'_2. Given (U_1, U'_2), let U_2 be distributed according to the conditional distribution of the second coordinate given that the first one is equal to U_1 under dC_2 and U'_1 be distributed according to the conditional distribution of the first coordinate given that the second one in equal to U'_2 still under dC_2. This way the random variables U_1, U_2, U'_1 and U'_2 are uniformly distributed on $[0, 1]$ and both the vectors (U_1, U_2) and (U'_1, U'_2) are distributed according to dC_2.

For $\varepsilon \in [0, 1]$, we consider

$$Y_\varepsilon = (U_1, \varepsilon U_2), \ Z_\varepsilon = (\varepsilon U_1, U_2) \text{ and } Z'_\varepsilon = (\varepsilon U'_1, U'_2).$$

We notice that the copula of Y_ε and Z_ε is C_2 since the copula is preserved by coordinatewise increasing functions (see Nelsen [1], Theorem 2.4.3). Also, Z_ε and Z'_ε obviously have the same law. We will show that for a suitable choice of C'_2 and $\varepsilon > 0$ small enough, we generally have

$$\mathbb{E}[\|Y_\varepsilon - Z_\varepsilon\|_q^p] > \mathbb{E}[\|Y_\varepsilon - Z'_\varepsilon\|_q^p].$$

Denoting by μ^ε the law of Y_ε and ρ^ε the common law of Z_ε and Z'_ε, this implies the desired conclusion since

$$\begin{aligned} \int_{\mathbb{R}^2 \times \mathbb{R}^2} \|x - y\|_q^p \mu^\varepsilon \diamond \rho^\varepsilon(dx, dy) &= \int_{[0,1]^2} \|(F^{-1}_{\mu_1^\varepsilon}(u_1), F^{-1}_{\mu_2^\varepsilon}(u_2)) - (F^{-1}_{\rho_1^\varepsilon}(u_1), F^{-1}_{\rho_2^\varepsilon}(u_2))\|_q^p dC_2(u_1, u_2) \\ &= \mathbb{E}[\|Y_\varepsilon - Z_\varepsilon\|_q^p] > \mathbb{E}[\|Y_\varepsilon - Z'_\varepsilon\|_q^p] \geq \mathcal{W}^p_{p,q}(\mu, \rho). \end{aligned}$$

The coupling between Y_ε and Z_ε gives the score

$$\mathbb{E}[((1-\varepsilon)^q U_1^q + (1-\varepsilon)^q U_2^q)^{p/q}] \underset{\varepsilon\to 0}{\to} \mathbb{E}[(U_1^q + U_2^q)^{p/q}],$$

while the one between Y_ε and Z'_ε gives:

$$\mathbb{E}[(|U_1 - \varepsilon U'_1|^q + |U'_2 - \varepsilon U_2|^q)^{p/q}] \underset{\varepsilon\to 0}{\to} \mathbb{E}[(U_1^q + (U'_2)^q)^{p/q}].$$

We now focus on the cost function $\bar{c}(u_1,u_2) = -(u_1^q + u_2^q)^{p/q}$ for $u_1, u_2 \in (0,1)$. We have

$$\partial_{u_1}\partial_{u_2}\bar{c}(u_1,u_2) = p(q-p)u_1^{q-1}u_2^{q-1}(u_1^q + u_2^q)^{\frac{p}{q}-2}.$$

When $q < p$ (resp. $q > p$) this is negative (resp. positive) for any $u_1, u_2 \in (0,1)$, i.e. $\bar{c}$ (resp. $-\bar{c}$) satisfies the so-called Monge condition. By Theorem 3.1.2 of Rachev and Rüschendorf [3], we get that $\mathbb{E}[\bar{c}(U_1, U'_2)]$ is maximal for $U'_2 = 1 - U_1$ (resp. $U'_2 = U_1$), i.e when $C'_2(u,v) = C_2^-(u,v)$ (resp. $C'_2(u,v) = C_2^+(u,v)$). Besides, since $\partial_{u_1}\partial_{u_2}\bar{c}(u_1,u_2)$ does not vanish, we have

$$\mathbb{E}[\bar{c}(U_1,U_2)] < \mathbb{E}[\bar{c}(U_1, 1-U_1)] \text{ (resp. } \mathbb{E}[\bar{c}(U_1,U_2)] < \mathbb{E}[\bar{c}(U_1,U_1)])$$

when $C_2 \neq C_2^-$ (resp. $C_2 \neq C_2^+$). Taking $U'_2 = 1 - U_1$ (resp. $U'_2 = U_1$), we have in this case

$$\mathbb{E}[((1-\varepsilon)^q U_1^q + (1-\varepsilon)^q U_2^q)^{p/q}] > \mathbb{E}[(|U_1 - \varepsilon U'_1|^q + |U'_2 - \varepsilon U_2|^q)^{p/q}]$$

for $\varepsilon > 0$ small enough. Notice that since both Y_0 and Z_0 have one constant coordinate, the range of the cumulative distribution function of this coordinate is $\{0,1\}$ and the probability measures μ_0 and ρ_0 share any two dimensional copula and in particular C_2^- (resp. C_2^+). That is why one has to choose $\varepsilon > 0$.

This two dimensional example can be easily extended to dimension $n \geq 3$. Let C_n now denote a n-dimensional copula, $C_2(u_1,u_2) = C_n(u_1,u_2,1,\dots,1)$ and C'_2 be another two-dimensional copula. We define (U_1, U_2, U'_1, U'_2) as above. Then, we choose $(U_3,\dots,U_n)$ (resp. $(U'_3,\dots,U'_n)$) distributed according to the conditional law of the $n-2$ last coordinates given that the two first are equal to (U_1,U_2) (resp. (U'_1,U'_2)) under dC_n. Last, we define

$$Y_\varepsilon = (U_1, \varepsilon U_2, \varepsilon U_3, \dots, \varepsilon U_n),\ Z_\varepsilon = (\varepsilon U_1, U_2, \varepsilon U_3, \dots, \varepsilon U_n) \text{ and } Z'_\varepsilon = (\varepsilon U'_1, U'_2, \varepsilon U'_3, \dots, \varepsilon U'_n).$$

We still have on the one hand that Y_ε and Z_ε share the same copula and on the other hand that Z_ε and Z'_ε have the same law. Moreover,

$$\mathbb{E}[\|Y_\varepsilon - Z_\varepsilon\|_q^p] \underset{\varepsilon\to 0}{\to} \mathbb{E}[(U_1^q + U_2^q)^{p/q}] \text{ and } \mathbb{E}[\|Y_\varepsilon - Z'_\varepsilon\|_q^p] \underset{\varepsilon\to 0}{\to} \mathbb{E}[(U_1^q + (U'_2)^q)^{p/q}].$$

Taking $\varepsilon > 0$ small enough, we get that $\mathbb{E}[\|Y_\varepsilon - Z_\varepsilon\|_q^p] > \mathbb{E}[\|Y_\varepsilon - Z'_\varepsilon\|_q^p] \geq \mathcal{W}_{p,q}^p(\mu,\rho)$ if, for some $u_1, u_2 \in [0,1]$, $C_n(u_1,u_2,1,\dots,1) > C_2^-(u_1,u_2)$ when $q < p$ (resp. $C_n(u_1,u_2,1,\dots,1) < C_2^+(u_1,u_2)$ when $q > p$). If there exist $i < j$ such that

$$C_n(1,\dots,1,u_i,1,\dots,1,u_j,1,\dots,1) > C_2^-(u_i,u_j) \text{ (resp. } < C_2^+(u_i,u_j)\text{)},$$

then one may repeat the above reasoning with the i-th and j-th coordinates replacing the first and second ones. Then the coupling $\nu = \mu \diamond \rho$ is not optimal for $\varepsilon > 0$ small enough. For $n \geq 3$, there is no copula such that $C_n(1,\dots,1,u_i,1,\dots,1,u_j,1,\dots,1) = C_2^-(u_i,u_j)$ for any $i < j$, $u_i, u_j \in [0,1]$, since this would imply that $U_2 = 1 - U_1 = U_3$ and $U_3 = 1 - U_2$. Also, the only one copula satisfying $C_n(1,\dots,1,u_i,1,\dots,1,u_j,1,\dots,1) = C_2^+(u_i,u_j)$ for any $i < j$, $u_i, u_j \in [0,1]$, is C_n^+ since the former condition implies $U_i = U_j$ for any $i < j$.

References

[1] Nelsen, R.B. An introduction to copulas, 2nd edition. Springer-Verlag, 2006.

[2] Puccetti, G. and Scarsini, M., Multivariate comonotonicity, Journal of Multivariate Analysis, p. 291–304, Vol. 101, No. 1, 2010.

[3] Rachev, S.T. and Rüschendorf, L. Mass Transportation problems. Springer-Verlag, 1998.