



**HAL**  
open science

# Kendall quantile ordering on $\mathbb{R}^2$ , probabilistic transport maps and their empirical counterpart

Philippe Berthet, Jean-Claude Fort

► **To cite this version:**

Philippe Berthet, Jean-Claude Fort. Kendall quantile ordering on  $\mathbb{R}^2$ , probabilistic transport maps and their empirical counterpart. 2023. hal-03987676v1

**HAL Id: hal-03987676**

**<https://hal.science/hal-03987676v1>**

Preprint submitted on 14 Feb 2023 (v1), last revised 6 Nov 2023 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# KENDALL QUANTILE ORDERING ON $\mathbb{R}^2$ , PROBABILISTIC TRANSPORT MAPS AND THEIR EMPIRICAL COUNTERPART

---

A PREPRINT

Philippe Berthet\*

Jean Claude Fort†

February 14, 2023

## ABSTRACT

We build a universal *r.v.* generator from the intrinsic geometry induced by what we define to be the Kendall quantile ordering of probability distribution functions on  $\mathbb{R}^2$ , having its own statistical interest. Using this generator we define a closed form transport map  $\tau_{FG}$  between any smooth distributions  $F$  and  $G$ . This  $\tau_{FG}$  is optimal when reduced to the Kendall quantile curves, for a large class of coordinate-wise convex costs. It coincides with the optimal transport if  $F$  and  $G$  share the same copula. The empirical counterpart  $\tau_{n,m}$  of  $\tau_{FG}$  is a non parametric transport plan that is easy to compute even for large samples. We illustrate the probabilistic geometry of  $\tau_{FG}$  by simulations of  $\tau_{n,m}$  that exhibit good performance with respect to the  $L^2$  Wasserstein distance, and point out some statistical applications.

**Keywords** Transport map · Generator · Explicit coupling · Bivariate ranks · Generalized quantile transform · Empirical distribution · Wasserstein distance.

**AMS Subject Classification:** 62G30 ; 60E10 ; 62H20 ; 60F15

## 1 Introduction

### 1.1 Overview

In this paper we define explicit transport maps between smooth distributions  $F$  and  $G$  on  $\mathbb{R}^2$  in such a way that their empirical counterparts are easily computed. Their performance can then be studied with respect to smooth coordinate-wise cost functions, including Wasserstein type costs. In the sequel, the cost  $c$  together with the distributions  $F$  and  $G$  are defined in some fixed orthogonal basis  $b_0$  of  $\mathbb{R}^2$  and are  $\mathcal{C}^2$ .

Our key tool is a new probability distribution generator  $\mathcal{G}_F$ . We call it universal since  $\mathcal{G}_F(Z)$  has distribution  $F$  if the random variable  $Z$  is uniformly distributed in the unit square – without orientation. This universal generator relies on what we introduce to be the Kendall ordering geometry derived from quantile-sets and quantile-curves shared by all smooth bivariate distributions – with an orientation. We shall write Kendall  $Q$ -sets,  $Q$ -curves, and  $Q$ -geometry for short. These geometrical features have a statistical interest by their own to define various bivariate descriptive quantities.

We first define the  $Q$ -geometry when the support of  $F$  is a rectangle, then  $\mathbb{R}^2$ . Given two such  $F$  and  $G$ , the isomorphism

$$\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1}$$

is designed to preserve the  $Q$ -curves indexing and to optimally transport each  $Q$ -curve. Moreover the degenerated coupling  $(x, \tau_{FG}(x))$  defines a closed form transport plan from  $F$  to  $G$ , thus having its own probabilistic interest. It is also a possible generalization of the univariate quantile transform in a bivariate setting. We insist that we first define

---

\*Institut de Mathématiques de Toulouse UMR 5219 ; Université Paul Sabatier, France. [philippe.berthet@math.univ-toulouse.fr](mailto:philippe.berthet@math.univ-toulouse.fr)

†MAP5 UMR 8145; Université Paris Cité, France. [jean-claude.fort@parisdescartes.fr](mailto:jean-claude.fort@parisdescartes.fr)

Kendall quantiles and ranks to generate distributions by propagating mass in some self-ordered way, then we deduce a class of transport synchronizing the propagation.

Due to the probabilistic nature of the  $Q$ -geometry, the empirical version  $\mathcal{G}_{F,n}$  of the generator  $\mathcal{G}_F$  is straightforward. Moreover, the ensuing empirical counterparts  $\tau_{n,m} = \mathcal{G}_{G,m} \circ \mathcal{G}_{F,n}^{-1}$ ,  $\tau_{n,G} = \mathcal{G}_G \circ \mathcal{G}_{F,n}^{-1}$  and  $\tau_{F,m} = \mathcal{G}_{G,m} \circ \mathcal{G}_F^{-1}$  of  $\tau_{FG}$  and their empirical transportation costs are coded easily and turn out to compute rather fast on personal computers. It is noteworthy that  $\tau_{n,m}$ ,  $\tau_{F,m}$  and  $\tau_{n,G}$  are almost everywhere continuous transport maps, in contrast with the usual discrete transport plan of one sample to the other.

To illustrate our approach we shall numerically study  $\tau_{FG}$  by using  $\tau_{n,m}$  and the popular  $L^2$  Wasserstein distance. This singular cost being independent of the orthogonal basis we optimize  $\tau_{FG}$  and  $\tau_{n,m}$  over the class of rotations of  $b_0$ . This provides convincing transports between multimodal distributions, such as Gaussian mixtures. Moreover, if  $F$  and  $G$  share the same copula in one of the rotated basis then the optimized  $\tau_{FG}$  coincides with the optimal transport. Likewise the optimal transport between two radial distributions is caught by working with polar coordinates.

Our main contribution is therefore a closed form transport plan  $\tau_{FG}$  based on the Kendall  $Q$ -geometry of distribution functions. This prospective work opens access to explicit limit theorems for well defined and easily computed non parametric empirical transport maps  $\tau_{nm}$ ,  $\tau_{Fm}$ ,  $\tau_{nG}$  and their costs – this will be considered elsewhere.

The paper is organized as follows. In the remainder of Section 1 we introduce more precisely the setting and informally present the new objects. In Section 2 we define and study the universal generator and the associated  $Q$ -geometry. In Section 3 we give some properties of the obtained transport maps related to some partial optimality. At Section 4 we derive properties of the Kendall quantiles and ranks. Section 5 is devoted to the definition and computation of an empirical transport map, and Section 6 provides some numerical examples. In Section 7 we conclude with further results to prove and possible new developments.

## 1.2 Optimal transport versus explicit coupling

Let  $c_1$  and  $c_2$  be  $\mathcal{C}^1$ , strictly convex functions on  $\mathbb{R}$ ,  $\mathcal{C}^2$  and positive on  $\mathbb{R}^*$ , such that  $c_1(0) = c_2(0) = 0$ . The cost functions we consider are defined as

$$c(x, y) = c_1(x_1 - y_1) + c_2(x_2 - y_2), \quad (1)$$

for  $x = (x_1, x_2)$ ,  $y = (y_1, y_2)$ . This includes all Wasserstein costs  $W_p$ , for  $c_1(w) = c_2(w) = |w|^p$  and  $p > 1$ .

Coordinate-wise costs do not depend on the origin. Only the choice of the orthogonal basis  $b_0$  of  $\mathbb{R}^2$  matters in the definition of  $c(x, y)$ , except if  $c$  is the squared euclidean norm  $W_2$ . More importantly, the distribution function  $F$  and consequently the associated  $Q$ -curves to be transported do not depend on the cost but strongly depend on quadrants of  $b_0$ .

In this framework the optimal transportation problem is to minimize, under suitable moment assumptions, among the couplings  $(X, Y)$  on  $\mathbb{R}^2 \times \mathbb{R}^2$  with marginals  $F$  and  $G$  on  $\mathbb{R}^2$  the mean transportation cost

$$W_c(F, G) = \min_{X \sim F, Y \sim G} \mathbb{E}(c(X, Y)). \quad (2)$$

Several questions arise naturally, that motivated our coupling.

*Optimal transport map.* First, which theoretical conditions ensure uniqueness and existence of an optimal coupling, called a transport map when it is degenerated. This question is adressed in details in the now classical work of [4], [10], [1], [11], [6] among many others. However very few explicit transport maps can be derived by using the theoretical tools developed to study uniqueness and existence. Whence our attempt to propose a simple, closed form transport map that always exists.

*Optimal cost.* In case of existence and uniqueness, thus for sufficiently smooth  $F$  and  $G$ , the second question is how to evaluate the minimal cost and, more deeply, how to approximate the optimal transport map. Usually the theoretical, methodological and numerical approaches developed in the wide literature on optimal transport focus on controlling error bounds for the cost evaluation rather than for the transport map approximation. Most often, optimality is not addressed in terms of distribution, coupling and transport map. Recent numerical methods proposed to find the optimal transportation cost use various gradient descent algorithms based on the dual formulation in terms of convex optimization on the very large space of conjugate functions. We refer for instance to [8], [9], [2], [3] aiming to approximate the solution of the dual Kantorovitch problem using relaxation techniques, the Sinkhorn algorithm or the Lagrange tessellation. Sometimes a relaxed version of the transport map is accessible, with an almost minimal cost. An other direction relies on PDE, SDE and limits of sequences of solutions associated to transport plans. Our proposal of a universal coupling is complementary. It is also meaningful since it involves some – arbitrary but explicit – probabilistic and statistical interpretation.

*Estimation.* A third question concerns the additional random approximation induced when  $F$  and  $G$  are only observed through two samples. This is the case in a statistical setting, or when  $F$  or  $G$  have no tractable densities. The plug-in solution is a discrete transport plan sending  $n$  sample points on  $m$  sample points as the original Monge problem and Kantorovitch answer, splitting local mass if  $n \neq m$ . Unfortunately, computing the optimal solution is numerically far too demanding for massive data and the resulting transport is defined on a finite set only. From a statistical viewpoint this may be misleading by overfitting, not providing a clear interpretation of the mass transfer and not estimating the key properties of  $F$  with respect to  $G$ . In contrast, we define everywhere a piece-wise continuous transport map carrying a bivariate quantile interpretation and the key  $Q$ -geometry can be learned non parametrically, then explicitly transported to any other one.

### 1.3 Universal generator and Kendall ordering

The random variable (*r.v.*)  $Z$  of our universal generator has uniform distribution  $U$  on the unit square  $\mathbb{U} = (0, 1)^2$ , with support  $\bar{\mathbb{U}} = [0, 1]^2$ . It is designed to generate *r.v.*'s in the plane that we identify to  $\text{vect}\{e_1, e_2\}$  with  $b_0 = (e_1, e_2)$  an orthonormal basis of  $\mathbb{R}^2$ . Due to the cost (1), the choice of  $b_0$  matters whereas the origin  $(0, 0)$  does not matter. Let denote the scalar product  $\langle x, y \rangle = x_1y_1 + x_2y_2$  and write  $x_i = \langle x, e_i \rangle$  for  $i = 1, 2$ .

*The oriented class  $\mathbb{F}$ .* Consider two *r.v.*  $X$  and  $Y$  having rectangular supports  $\mathcal{R}_X = (x_1^-, x_1^+) \times (x_2^-, x_2^+)$  and  $\mathcal{R}_Y = (y_1^-, y_1^+) \times (y_2^-, y_2^+)$  in  $b_0$ , with closure denoted  $\bar{\mathcal{R}}_X$  and  $\bar{\mathcal{R}}_Y$ . Let assume that on  $\mathcal{R}_X$  and  $\mathcal{R}_Y$  they have  $\mathcal{C}^2$  distribution functions and positive  $\mathcal{C}^1$  densities denoted  $(F, f)$  and  $(G, g)$  respectively. Write  $\mathbb{F}$  the  $b_0$ -dependent family of such smooth distributions with parallel bounded rectangular support and  $\bar{\mathbb{F}}$  its extension allowing the unbounded rectangles, including  $\mathbb{R}^2$ . Notice the difference between the universal generator coordinate space  $\mathbb{U}$  and the  $b_0$ -oriented square  $\mathcal{R}_X$  in the plane if  $F$  is uniform. The  $Q$ -geometry of  $F \in \mathbb{F}$  is defined by the quantile curves from which the generator  $\mathcal{G}_F$  is defined, and is intrinsically linked to the orientation  $b_0$  of  $\mathbb{F}$ .

*Quantiles and generator.* The main open problem to be addressed is the bivariate generalization of the univariate quantile transform of  $X$  into  $Y$  based on  $F$  and  $G$  only, whatever their parallel rectangular supports – bounded or unbounded. A convenient notion of bivariate quantile is then needed. To this aim we define in Section 2 a one to one generator map  $\mathcal{G}_F$  such that  $\mathcal{G}_F(Z)$  has distribution  $F$  on  $\mathcal{R}_X$  and its inverse map  $\mathcal{G}_F^{-1}$  is such that  $\mathcal{G}_F^{-1}(X)$  has distribution  $U$  on  $\mathbb{U}$ . It makes sense to call  $\mathcal{G}_F(z) = x$  the bivariate quantile of  $X$  of order  $z = (z_1, z_2) \in \mathbb{U}$  and to call  $\mathcal{G}_F^{-1}(x) = z$  the bivariate rank of  $x = (x_1, x_2) \in \mathcal{R}_X$ .

In the univariate case this corresponds to the quantile function  $\mathcal{G}_F = F^{-1}$  and the rank function  $\mathcal{G}_F^{-1} = F$  with the rank segment  $\mathbb{U} = (0, 1)$  endowed with its natural continuous increasing strict order. The unidimensional quantile transform  $F^{-1} \circ G$  simply consists in coupling through the rank, and turns out to be the optimal transport for any convex cost.

In the same spirit  $\mathcal{G}_F$  is such that it preserves the following strict order on the bivariate rank square  $\mathbb{U}$ . In general, the induced bidimensional quantile transform  $\mathcal{G}_G \circ \mathcal{G}_F^{-1}$  is a transport enjoying only partial optimality.

*Kendall ordering.* The universal generator coordinate  $z = (z_1, z_2) \in \mathbb{U}$  is endowed with the strict order

$$z < z^+ \text{ if either } z_1 < z_1^+ \text{ or } z_1 = z_1^+ \text{ and } z_2 < z_2^+. \quad (3)$$

Thus  $z \leq z'$  if  $z < z'$  or  $z = z'$ . The generator  $\mathcal{G}_F$  is based on the Kendall ordering we characterize as follows: if  $x = \mathcal{G}_F(z)$  then

$$\mathbb{P}(F(X) \leq F(x)) = z_1, \quad \mathbb{P}(X_2 \geq x_2 \mid F(X) = F(x)) = z_2,$$

which, in the quadrant oriented  $Q$ -geometry driving  $\mathcal{G}_F$ , is equivalent to

$$\mathbb{P}(F(X) \leq F(x)) = z_1, \quad \mathbb{P}(X_1 \leq x_1 \mid F(X) = F(x)) = z_2.$$

In other words, for all  $z \in \mathbb{U}$  and  $F \in \mathbb{F}$  we have

$$z_1 = \mathbb{P}(\mathcal{G}_F^{-1}(X) \leq z), \quad z_2 = \mathbb{P}(X_2 \geq \langle \mathcal{G}_F(z), e_2 \rangle \mid \mathcal{G}_F^{-1}(X) \in \{z_1\} \times (0, 1)) \quad (4)$$

and we also have  $z_2 = \mathbb{P}(X_1 \leq \langle \mathcal{G}_F(z), e_1 \rangle \mid \mathcal{G}_F^{-1}(X) \in \{z_1\} \times (0, 1))$ . As a matter of fact,  $\mathcal{G}_F^{-1}$  induces the following geometrical and stochastic orders on  $\mathbb{F}$ . We say that  $x \leq x^+$  in the Kendall  $Q$ -geometry of  $F$  if  $\mathcal{G}_F^{-1}(x) \leq \mathcal{G}_F^{-1}(x^+)$  in the sense (3) of the universal generator coordinates, which means that either  $F(x) < F(x^+)$  or

$$F(x) = F(x^+), \quad \mathbb{P}(X_2 \geq x_2 \mid F(X) = F(x)) \leq \mathbb{P}(X_2 \geq x_2^+ \mid F(X) = F(x)).$$

Likewise we write  $X(\omega) \leq Y(\omega')$  if  $\mathcal{G}_F^{-1}(X(\omega)) \leq \mathcal{G}_G^{-1}(Y(\omega'))$  even if  $X, Y$  have different initial probability spaces but parallel rectangular supports. We only compare the random ranks  $\mathcal{G}_F^{-1}(X)$  and  $\mathcal{G}_G^{-1}(Y)$ , and our coupling on the same probability space makes these ranks equal.

*Q-transform transport map.* Denote  $P^X$  and  $P^Y$  the probability distributions of  $X$  and  $Y$ . A smooth transport map from  $F$  to  $G$  is a one to one continuously differentiable map  $\tau : \mathcal{R}_X \rightarrow \mathcal{R}_Y$  such that  $P^Y = P^X \circ \tau^{-1}$ . Write  $\mathcal{T}_1(F, G)$  the set of continuously differentiable transport maps from  $F$  to  $G$ . As announced above, the uniquely determined  $Q$ -transform

$$\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1} \quad (5)$$

works on  $\mathbb{R}^2$  as the quantile-rank transform  $G^{-1} \circ F$  on  $\mathbb{R}$  since  $\tau_{FG}(X)$  is a version of  $Y$ . Our main result is the existence and uniqueness of a universal generator preserving Kendall quantiles and ranks.

**Theorem 1.** *There exists a unique map  $\mathcal{G} : F \in \mathbb{F} \rightarrow \mathcal{G}(F) = \mathcal{G}_F$  such that for  $F \in \mathbb{F}$ ,  $\mathcal{G}_F(Z)$  has distribution  $F$ ,  $\mathcal{G}_F$  satisfies (4) and, for  $(F, G) \in \mathbb{F} \times \mathbb{F}$ , the  $Q$ -transform  $\tau_{FG}$  of (5) satisfies  $\tau_{FG} \in \mathcal{T}_1(F, G)$ .*

The map  $\mathcal{G}$  is the universal generator explicitly constructed from the  $Q$ -geometry developed at Section 2. It has nice probabilistic (Section 3), statistical (Sections 4, 7) and numerical (Section 5, 6) properties.

## 2 Universal generator and $Q$ -transport map

In this section we introduce and study the generator  $\mathcal{G}_F$  of  $F \in \mathbb{F}$  and the  $Q$ -transform (5), in order to first derive Theorem 1 for bounded rectangles.

### 2.1 Generator equation along $Q$ -curves

Given  $F \in \mathbb{F}$  let denote  $F_1$  and  $F_2$  the cumulative distribution functions (*c.d.f.*) on  $\mathbb{R}$  of the marginal *r.v.*  $X_1 = \langle X, e_1 \rangle$  and  $X_2 = \langle X, e_2 \rangle$ . For  $\alpha \in [0, 1]$  consider the  $\alpha$ -level set of  $F$ ,

$$Q_F(\alpha) = \{x \in \overline{\mathcal{R}}_X : F(x) = \alpha\} \quad (6)$$

that we call the  $\alpha$ -th  $Q$ -curve of  $F$ . Observe that for  $\alpha \in (0, 1)$ ,  $Q_F(\alpha)$  is a curve joining  $(F_1^{-1}(\alpha), x_2^+)$  to  $(x_1^+, F_2^{-1}(\alpha))$ . Moreover  $Q_F(1) = \{(x_1^+, x_2^+)\}$  is the upper-right corner and  $Q_F(0) = \{(x_1, x_2) : x_1 = x_1^- \text{ or } x_2 = x_2^-\}$  is the lower left half-perimeter. Define  $\mathbb{Q}_F(0) = Q_F(0)$ ,  $\mathbb{Q}_F(1) = \mathcal{R}_X$  and, for  $\alpha \in [0, 1]$ , the  $\alpha$ -th  $Q$ -set of  $F$

$$\mathbb{Q}_F(\alpha) = \{x \in \mathcal{R}_X : F(x) \leq \alpha\} = \bigcup_{a=0}^{\alpha} Q_F(a). \quad (7)$$

**Definition 2.** *Let the Kendall distribution function of  $F$  on  $[0, 1]$  be*

$$K_F(\alpha) = F(\mathbb{Q}_F(\alpha)). \quad (8)$$

In other words  $K_F$  is the *c.d.f.* of the *r.v.*  $F(X)$ . We first notice that on  $\mathbb{F}$ ,  $K_F$  only depends on the copula function of  $F$ . Remind that  $F \in \mathbb{F}$  has positive density on the open rectangle  $\mathcal{R}_X$ .

**Proposition 3.** *If  $(F, G) \in \mathbb{F} \times \mathbb{F}$  have same copula then  $K_F = K_G$ .*

*Proof.* By definition, the copulas  $C_F$  of  $F$  and  $C_G$  of  $G$  satisfy, for  $x = (x_1, x_2) \in \mathcal{R}_X$  and  $y = (y_1, y_2) \in \mathcal{R}_Y$ ,  $F(x_1, x_2) = C_F(F_1(x_1), F_2(x_2))$  and  $G(y_1, y_2) = C_G(G_1(y_1), G_2(y_2))$ . For  $i = 1, 2$  the marginals  $F_i$  and  $G_i$  are continuous, strictly increasing on their respective supporting interval and  $F_i \circ F_i^{-1} = G_i \circ G_i^{-1}$  is the identity on  $(0, 1)$ . Thus, for  $\alpha \in (0, 1)$ ,

$$\begin{aligned} K_F(\alpha) &= \mathbb{P}(F(X) \leq \alpha) \\ &= \mathbb{P}(C_F(F(X_1), F(X_2)) \leq \alpha) \\ &= \mathbb{P}(C_G(G_1(G_1^{-1} \circ F_1(X_1)), G_2(G_2^{-1} \circ F_2(X_2))) \leq \alpha) \\ &= \mathbb{P}(G(Y) \leq \alpha) \\ &= K_G(\alpha) \end{aligned}$$

where  $Y = (G_1^{-1} \circ F_1(X_1), G_2^{-1} \circ F_2(X_2))$  has distribution  $G$  since

$$\begin{aligned} &\mathbb{P}((G_1^{-1} \circ F_1(X_1), G_2^{-1} \circ F_2(X_2)) \leq (y_1, y_2)) \\ &= \mathbb{P}((F_1(X_1), F_2(X_2)) \leq (G_1(y_1), G_2(y_2))) \\ &= C_F(G_1(y_1), G_2(y_2)) \\ &= G(y_1, y_2). \end{aligned}$$

□

We shall use the fact that for  $F \in \mathbb{F}$ ,  $K_F$  is  $\mathcal{C}^2$ .

**Proposition 4.** *If  $F$  is  $\mathcal{C}^2$  and has continuous positive density  $f$  on the open rectangle  $\mathcal{R}_X$  then  $K_F$  has continuous positive density on  $(0, 1)$ , the collection of  $Q$ -curves  $\alpha \rightarrow Q_F(\alpha)$  determine  $F$  and the collection of  $Q$ -sets  $\alpha \rightarrow \mathbb{Q}_F(\alpha)$  determine  $F$ . If moreover  $f$  is  $\mathcal{C}^1$  on  $\mathcal{R}_X$  then  $K_F$  is  $\mathcal{C}^2$ .*

Before proving Proposition 4 we need to describe more precisely (6) as a parametric curve  $Q_F(\alpha, t)$  with  $t \geq 0$ . Let  $\nabla F$  denote the gradient of  $F$  and  $\bar{\nabla} F$  its right-oriented orthogonal vector tangent to the smooth curve  $Q_\alpha(F)$ . Thus  $\nabla F$  and  $\bar{\nabla} F$  satisfy  $\langle \nabla F, e_1 \rangle \geq 0$ ,  $\langle \nabla F, e_2 \rangle \geq 0$ ,  $\langle \bar{\nabla} F, e_1 \rangle \geq 0$  and  $\langle \bar{\nabla} F, e_2 \rangle \leq 0$ .

**Definition 5.** *For any  $\alpha \in (0, 1)$  define  $t \in \mathbb{R}^+ \rightarrow Q_F(\alpha, t) \in Q_F(\alpha)$  to be the solution of the ordinary differential equation*

$$\frac{dQ_F(\alpha, t)}{dt} = \bar{\nabla} F(Q_F(\alpha, t)), \quad Q_F(\alpha, 0) = (F_1^{-1}(\alpha), x_2^+). \quad (9)$$

Write  $t_F : x \in \mathcal{R}_X \rightarrow t_F(x)$  the unique solution (in  $t$ ) of  $Q_F(F(x), t) = x$ . The total time along  $Q_F(\alpha)$  is

$$T_F(\alpha) = t_F((x_1^+, F_2^{-1}(\alpha))) = \min \{t : Q_F(\alpha, t) = (x_1^+, F_2^{-1}(\alpha))\}. \quad (10)$$

Define the mass-time density on  $[0, T_F(\alpha)]$  to be

$$f_\alpha(t) = \frac{f(Q_F(\alpha, t))}{k_F(\alpha)}, \quad (11)$$

where

$$k_F(\alpha) = \int_0^{T_F(\alpha)} f(Q_F(\alpha, t)) dt. \quad (12)$$

*Proof of Proposition 4.* It holds  $\|\bar{\nabla} F\|_2 = \|\nabla F\|_2 > 0$  on  $\mathcal{R}_X$ , as  $f > 0$ . Write  $f_1$  the density of  $F_1$ , and observe that  $\|\bar{\nabla} F(Q_F(\alpha, 0))\|_2 > 0$  for  $\alpha \in (0, 1)$ , since  $f_1(F_1^{-1}(\alpha)) > 0$ . The existence of  $k_F$  follows from a change of variable onto the parametrization of Definition 5. Recall that  $F$  is  $\mathcal{C}^2$  and  $\bar{\nabla} F$  is  $\mathcal{C}^1$  thus  $Q_F$  is  $\mathcal{C}^1$ . Since  $(\nabla F/\|\nabla F\|_2, \bar{\nabla} F/\|\nabla F\|_2)$  is an orthonormal basis and  $F(Q_F(\alpha, t)) = \alpha$  we get, by (9),

$$\begin{aligned} \frac{dQ_F(\alpha, t)}{d\alpha} \cdot \nabla F(Q_F(\alpha, t)) &= 1, \\ \left| \det \left( \frac{dQ_F(\alpha, t)}{dt}, \frac{dQ_F(\alpha, t)}{d\alpha} \right) \right| &= 1. \end{aligned}$$

Therefore, by putting  $x = Q_F(\alpha, t)$ ,

$$K_F(\alpha) = \int_{x \in \mathbb{Q}_F(\alpha)} f(x) dx = \int_0^\alpha \int_0^{T_F(a)} f(Q_F(a, t)) dt da = \int_0^\alpha k_F(a) da.$$

Hence  $k_F$  is a positive density of  $K_F$  on  $(0, 1)$ . Assuming that  $f$  is  $\mathcal{C}^1$ , the alternative representation (15) below shows that  $k_F$  is  $\mathcal{C}^1$  and, by (12),  $T_F(\alpha)$  is differentiable. Next, for  $\alpha' > \alpha$  it holds  $Q_F(\alpha) \cap Q_F(\alpha') = \emptyset$  and  $\mathbb{Q}_F(\alpha) \subset \mathbb{Q}_F(\alpha')$  thus the  $Q$ -curves determine  $F(x) = \min \{q : x \in Q_F(q)\}$  and the  $Q$ -sets determine  $F(x) = \min \{q : x \in \mathbb{Q}_F(q)\}$  for  $x \in \mathcal{R}_X$ .  $\square$

## 2.2 The generator property

In order to define the generator we need to change coordinate in the solution  $Q_F(\alpha, t)$  of (9).

**Definition 6.** *For  $\alpha \in (0, 1)$  let the mass-time c.d.f. along the  $\alpha$ -th  $Q$ -curve of  $F$  be*

$$F_\alpha(t) = \int_0^t f_\alpha(s) ds \in (0, 1), \quad \text{for } t \in [0, T_F(\alpha)]. \quad (13)$$

Write  $F_\alpha^{-1} : [0, 1] \rightarrow [0, T_F(\alpha)]$  its inverse function and set

$$x_F(\alpha, u) = Q_F(\alpha, t) \text{ such that } F_\alpha(t) = u, \quad \text{for } u \in [0, 1]. \quad (14)$$

Thus  $u \rightarrow x_F(\alpha, u) = Q_F(\alpha, F_\alpha^{-1}(u))$  is a parametrization of  $Q_F(\alpha)$  by  $[0, 1]$  through the  $Q$ -curve conditional probability. Likewise,  $x_F(\alpha, u)$  is a parametrization of  $\mathcal{R}_X$  by  $\mathbb{U} = (0, 1)^2$  through the  $Q$ -geometry.

**Remark 7.** Combined with (3), (14) induces a strict order on  $\mathcal{R}_X$  that characterizes  $F$ . Namely  $x_F(\alpha, u) < x_F(\alpha^+, u^+)$  if  $(\alpha, u) < (\alpha^+, u^+)$  in the sense of (3), which is equivalent to (4). The idea behind is that the mass has been ordered to prepare its generation then its transportation.

At this stage,  $(\alpha, u) \in (0, 1)^2$  stands as a generator coordinate system and  $x_F(\alpha, u)$  from (14) as a generator mapping process for  $F$ . Starting from two independent uniform coordinates  $Z$ ,  $x_F$  can be used as follows to perfectly simulate an  $X$  with distribution  $F$ .

**Theorem 8.** Let  $F \in \mathbb{F}$  and  $Z = (Z_1, Z_2)$  be a uniform r.v. on  $\mathbb{U}$ . Then the r.v.  $x_F(K_F^{-1}(Z_1), Z_2)$  has distribution  $F$ .

*Proof.* First we show that the one to one mapping  $x_F(\alpha, u)$  is also  $\mathcal{C}^1$  from  $\mathbb{U}$  to  $\mathcal{R}_X$ , and thus is a  $\mathcal{C}^1$  diffeomorphism. The smoothness with respect to  $u$  is clear, what remains to prove is the smoothness of  $x_F$  with respect to  $\alpha$ . From  $x_F(\alpha, u) = Q_F(\alpha, F_\alpha^{-1}(u))$  we only need to check the continuous differentiability of  $k_F(\alpha)$ . Recall that  $F$  is  $\mathcal{C}^2$ . Let parametrize  $Q_F(\alpha)$  with  $x_1 \in [F_1^{-1}(\alpha), x_1^+]$ . There exists  $r_\alpha$  such that  $x_2 = r_\alpha(x_1)$  when  $x \in Q_F(\alpha)$ . As  $F(x_1, r_\alpha(x_1)) = \alpha$  it holds

$$\frac{\partial r_\alpha(x_1)}{\partial \alpha} = -\frac{1}{\frac{\partial F(x_1, r_\alpha(x_1))}{\partial x_2}}.$$

Now, the fact that  $F$  is  $\mathcal{C}^2$  readily implies that  $\partial r_\alpha(x_1)/\partial \alpha$  is  $\mathcal{C}^1$  with respect to  $(\alpha, t)$ , then integrating over  $Q_F(\alpha)$  of (7) yields

$$K_F(\alpha) = \alpha + \int_{F_1^{-1}(\alpha)}^{x_1^+} dx_1 \int_0^{r_\alpha(x_1)} f(x_1, x_2) dx_2$$

then deriving with respect to  $\alpha$  gives

$$\begin{aligned} k_F(\alpha) &= 1 - \frac{1}{f_1(F_1^{-1}(\alpha))} \int_{x_2^-}^{x_2^+} f(F_1^{-1}(\alpha), x_2) dx_2 \\ &\quad + \int_{F_1^{-1}(\alpha)}^{x_1^+} \frac{\partial r_\alpha(x_1)}{\partial \alpha} f(x_1, r_\alpha(x_1)) dx_1. \end{aligned} \quad (15)$$

From (15) we get, as  $f$  and  $\partial r_\alpha(x_1)/\partial \alpha$  are  $\mathcal{C}^1$  and  $\partial f/\partial x_1$  is continuous, that  $k_F(\alpha)$  is  $\mathcal{C}^1$ . Next we prove that the function defined on  $\mathbb{U}$  by  $H(z_1, z_2) = x_F(K_F^{-1}(z_1), z_2)$  satisfies  $|\det \nabla H(z_1, z_2)| = \frac{1}{f(H(z_1, z_2))}$ . Since  $\partial Q_F(\alpha, s)/\partial t = \bar{\nabla} F(Q_F(\alpha, t))$  and  $\partial F_\alpha(t)/\partial t = f_\alpha(t)$ , we have

$$\frac{\partial H(z_1, z_2)}{\partial z_2} = \bar{\nabla} F(x_F(K_F^{-1}(z_1), z_2)) \frac{k_F(K_F^{-1}(z_1))}{f(x_F(K_F^{-1}(z_1), z_2))} \quad (16)$$

then it holds  $F(H(z_1, z_2)) = K_F^{-1}(z_1)$  and thus

$$\frac{\partial H(z_1, z_2)}{\partial z_1} \nabla F(H(z_1, z_2)) = \frac{1}{k_F(K_F^{-1}(z_1))}. \quad (17)$$

The proof is complete,  $(\nabla F/\|\nabla F\|_2, \bar{\nabla} F/\|\nabla F\|_2)$  being orthonormal.  $\square$

### 2.3 Universal generator map

We are now ready to define the class of generators of r.v. with distribution in  $\mathbb{F}$  mentioned in Theorem 1. Remind (8), (13) and (14).

**Definition 9.** The generator  $\mathcal{G}_F : \mathbb{U} \rightarrow \mathcal{R}_X$  is defined to be

$$\mathcal{G}_F(z) = Q_F\left(K_F^{-1}(z_1), F_{K_F^{-1}(z_1)}^{-1}(z_2)\right) = x_F(K_F^{-1}(z_1), z_2), \quad z \in \mathbb{U}.$$

The universal generator map is  $\mathcal{G} : F \in \mathbb{F} \rightarrow \mathcal{G}(F) = \mathcal{G}_F$ .

By Theorem 8 for  $Z \sim U$ , the r.v.  $\mathcal{G}_F(Z)$  has distribution  $F$ . By the proof of Theorem 8, and Proposition 4,  $\mathcal{G}_F$  is a  $\mathcal{C}^1$  diffeomorphism from  $\mathbb{U}$  to  $\mathcal{R}_X$ . Lets give a simple example that is not at all the simplest since  $K_F$  is not trivially invertible.

**Example 10.** Let us build  $\mathcal{G}_F$  explicitly for  $F = U$ , to show that  $\mathcal{G}_U$  is not at all the identity if  $\mathbb{U} = \mathcal{R}_Z$ . We easily get  $Q_U(\alpha) = \{x : x_1 x_2 = \alpha\}$ ,  $K_U(\alpha) = \alpha - \alpha \log \alpha$ ,  $k_F(\alpha) = T_U(\alpha) = -\log \alpha$ ,  $Q_U(\alpha, t) = (\alpha e^t, e^{-t})$ , and  $U_\alpha(t) = -t/\log \alpha$ ,  $0 \leq t \leq T_U(\alpha)$ . Then, by Definition 9,

$$\mathcal{G}_U(Z) = \left( K_U^{-1}(Z_1)^{(1-Z_2)}, K_U^{-1}(Z_1)^{Z_2} \right).$$

After some computations one can verify that  $\mathcal{G}_U(Z)$  actually has same distribution  $U$  as  $Z$ , but  $\mathcal{G}_U(Z)$  is obviously not  $Z$ .

**Remark 11.** In the generator  $Q$ -geometry  $\mathcal{G}_F(z_1, z_2)$  the parameter  $z_1$  is the "probability below the curve" and the parameter  $z_2$  is the "probability along the curve". They are of equal importance to locate  $X = \mathcal{G}_F(Z)$  in  $\mathcal{R}_X$ .

The main property making the coordinate system of  $\mathcal{G}$  universal is that all couplings  $(\mathcal{G}_F(Z), \mathcal{G}_G(Z))$  are simultaneously ordered in the sense of (4). In the spirit of Section 4, they affect the same  $Q$ -ranks to both marginals.

## 2.4 The $Q$ -transform $\tau_{FG}$

**Proposition 12.** For  $(F, G) \in \mathbb{F} \times \mathbb{F}$ , the map  $\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1}$  satisfies  $\tau_{FG} \in \mathcal{T}_1(F, G)$ .

By combining Theorem 8, Definition 9 and Proposition 12 we have proved Theorem 1. Clearly,  $\tau_{FG}$  is the unique one to one mapping between  $\mathcal{R}_X$  and  $\mathcal{R}_Y$  preserving the Kendall quantile ordering of  $F$  and  $G$ .

## 3 Probabilistic properties of the transport map $\tau_{FG}$

### 3.1 $Q$ -curves optimal transport

In this section we generalize the usual formula of the optimal transport between two distributions on  $\mathbb{R}$  to the case of two uniform distributions  $U_x$  and  $U_y$  on any smooth curves  $x$  and  $y$  in  $\mathbb{R}^d$  that are globally coordinate-wise co-monotonic. The above  $Q$ -curves are a special case, for  $d = 2$ .

**Theorem 13.** Fix  $d > 1$ . Let  $x(t)$  and  $y(t)$  be two  $C^1$  curves in  $\mathbb{R}^d$  parametrized by  $t \in (0, 1)$ . For  $1 \leq i \leq d$ , denote  $x_i(t)$  and  $y_i(t)$  their coordinates and assume that for any  $(t_1, t_2) \in (0, 1)^2$  their derivatives satisfy  $x'_i(t_1)y'_i(t_2) > 0$ . Let  $c$  be any cost of the form

$$c(x, y) = \sum_{i=1}^d c_i(x_i - y_i)$$

where the function  $c_i$  are  $C^1$  on  $\mathbb{R}$ ,  $C^2$  on  $\mathbb{R}^*$ , strictly convex, non negative, null at 0 and satisfy, for  $1 \leq i \leq d$  and all  $x_i, x'_i, y_i, y'_i$  in  $\mathbb{R}$ ,

$$-\int_{x_i}^{x'_i} \int_{y_i}^{y'_i} c''_i(x - y) dx dy = c_i(x'_i - y'_i) - c_i(x'_i - y_i) - c_i(x_i - y'_i) + c_i(x_i - y_i). \quad (18)$$

Consider two uniform r.v. on  $(0, 1)$ ,  $U$  and  $V$ . Then the  $c$ -optimal transport map between  $X = x(U)$  and  $Y = y(V)$  is given by  $(x(U), y(U))$ .

*Proof.* We mimic the proof on the real line, coordinate by coordinate. Let us estimate  $\mathbb{E}(c(x(U) - y(V)))$  whatever the copula  $\pi$ , that is the joint distribution of  $(U, V)$ . First observe that, for  $1 \leq i \leq d$ ,

$$\begin{aligned} c_i(x_i(u) - y_i(v)) &= c_i(x_i(u) - y_i(0)) + c_i(x_i(0) - y_i(v)) - c_i(x_i(0) - y_i(0)) \\ &\quad - \int_0^u \int_0^v c''_i(x_i(s) - y_i(t)) x'_i(s) y'_i(t) ds dt. \end{aligned}$$

Since  $\mathbb{E}(c_i(x_i(U) - y_i(0)))$  and  $\mathbb{E}(c_i(x_i(0) - y_i(V)))$  only depend on the known marginals  $x_i(U)$  and  $y_i(V)$  we only need to evaluate  $\mathbb{E}(-\int_0^U \int_0^V c''_i(x_i(s) - y_i(t)) x'_i(s) y'_i(t) ds dt)$ . By denoting  $\Pi$  the distribution function of  $\pi$  it holds, applying Fubini's theorem,

$$\begin{aligned} &\mathbb{E} \left( - \int_0^U \int_0^V c''_i(x_i(s) - y_i(t)) x'_i(s) y'_i(t) ds dt \right) \\ &= - \int_0^1 \int_0^1 \int_s^1 \int_t^1 \pi(du, dv) c''_i(x_i(s) - y_i(t)) x'_i(s) y'_i(t) ds dt \\ &= - \int_0^1 \int_0^1 c''_i(x_i(s) - y_i(t)) x'_i(s) y'_i(t) (1 - s - t + \Pi(s, t)) ds dt. \end{aligned}$$



Here  $c_i''(x_i(s) - y_i(t))x_i'(s)y_i'(t) \geq 0$  by hypothesis, and it is well known that the maximum of  $\Pi(s, t)$  is achieved for  $\Pi^+(s, t) = s \wedge t$ , that is the distribution of  $(U, U)$ .  $\square$

Write  $p_x = x \circ U$  and  $p_y = y \circ U$  the distribution on the parametrized curves  $x$  and  $y$  associated to a uniform parameter. Theorem 13 yields

$$\begin{aligned} W_c(p_x, p_y) &= \min_{X \sim p_x, Y \sim p_y} \mathbb{E}(c(X, Y)) \\ &= \int_0^1 c(x(u), y(u)) du = \sum_{i=1}^d \int_0^1 c_i(x_i(u) - y_i(u)) du. \end{aligned}$$

**Example 14.** All the  $p$ -norms,  $p > 1$ , satisfy (18). One can also use various weighted sums of power functions, namely, for  $p_i > 1$  and  $a_i > 0$ ,

$$c(x) = \sum_{i=1}^d a_i |x_i|^{p_i}.$$

**Remark 15.** Actually the condition (18) can be weakened but this is not the point here. For sake of simplicity we use coordinate-wise costs, however the result is valid for costs and curves satisfying, for all  $(u, v) \in (0, 1)^2$ ,

$$\int_0^u \int_0^v x'(s)^T c''(x(s) - y(t)) y'(t) ds dt + c(x(u) - y(v)) = \phi(x, u) + \psi(y, v)$$

and  $x'(s)^T c''(x(s) - y(t)) y'(t) \geq 0$  for all  $(s, t) \in \mathbb{U}$ , with  $c''$  the Hessian matrix of  $c$  and  $x'(s)^T$  the transposed gradient of  $x$  at time  $t$ .

Let denote  $\tilde{F}_\alpha$  the probability measure on  $Q_F(\alpha)$  of (6) putting measure  $F_\alpha(t_1) - F_\alpha(t_0)$  from (13) to the  $Q$ -curve arc joining  $Q_F(\alpha, t_0)$  to  $Q_F(\alpha, t_1)$ , for any  $0 < t_0 < t_1 < T_F(\alpha)$  of (10). Consider again  $\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1}$ .

**Corollary 16.** Let  $c$  satisfy (1) and (18) for  $i = 1, 2$ . For any  $\alpha \in (0, 1)$ ,  $(F, G) \in \mathbb{F} \times \mathbb{F}$ ,  $\tau_{FG}$   $c$ -optimally transports  $\tilde{F}_\alpha$  onto  $\tilde{G}_{K_G^{-1} \circ K_F(\alpha)}$ .

*Proof.* We can straightforwardly apply Theorem 13 for  $d = 2$  to the  $Q$ -curves parametrized by  $u \in (0, 1)$ , namely  $x(u) := x_F(\alpha, u) = Q_F(\alpha, F_\alpha^{-1}(u))$  and  $y(u) := x_G(\beta, u) = Q_G(\beta, G_\beta^{-1}(u))$ . Clearly, whatever  $(\alpha, \beta) \in (0, 1)^2$  and  $(F, G) \in \mathbb{F} \times \mathbb{F}$ , the two components of  $\nabla F$  and  $\nabla G$  have always the same sign, hence the assumptions of Theorem 13 are satisfied.  $\square$

### 3.2 R-curves optimal transport

Let show that  $\tau_{FG}$  satisfies a property similar to Corollary 16 for the following curves.

**Definition 17.** Given  $z_2 \in (0, 1)$ , the  $z_2$ -th  $R$ -curve of  $F \in \mathbb{F}$  is the parametrized curve  $R_F(z_2) : \alpha \in (0, 1) \rightarrow x_F(\alpha, z_2)$ .

For instance in the uniform case  $U$  of Example 10 for  $(z_2, u) \in (0, 1)^2$   $R_U(z_2)$  is the (power) curve  $(u^{1-z_2}, u^{z_2})$ .

Let denote  $\tilde{K}_{F, z_2}$  the probability measure on  $R_F(z_2)$  putting measure  $K_F(\alpha_1) - K_F(\alpha_0)$  to the  $R$ -curve arc joining  $x_F(\alpha_0, z_2)$  to  $x_F(\alpha_1, z_2)$  along  $R_F(z_2)$ , for any  $0 < \alpha_0 < \alpha_1 < 1$ . For  $G \in \mathbb{F}$  consider  $R_G(z_2)$  and the probability measure  $\tilde{K}_{G, z_2}$ . Observe that  $R_G(z_2)$  is also the image curve

$$\tau_{FG}(R_F(z_2)) : \alpha \in (0, 1) \rightarrow x_G(K_G^{-1} \circ K_F(\alpha), z_2),$$

however with a different parametrization. By Definitions 2 and 6,

$$\begin{aligned} F((0, \alpha) \times (1 - x_F((0, \alpha), z_2))) &= z_2 K_F(\alpha), \\ G((0, \beta) \times (1 - x_G((0, \beta), z_2))) &= z_2 K_G(\beta), \end{aligned}$$

and when  $\beta = K_G^{-1} \circ K_F(\alpha)$  it holds  $K_G(\beta) = K_F(\alpha)$ . Which simply means that the  $R$ -curves  $R_F(z_2)$  divide the open square in two open subsets of probability  $z_2$  and  $1 - z_2$  and of probability  $z_2 K_F(\alpha)$  and  $(1 - z_2) K_F(\alpha)$  when stopped at "mass-time"  $\alpha$  (the same holds for  $R_G(z_2)$  with  $K_G$ ).

As observed above,  $\tau_{FG}$  maps  $R_F(z_2)$  onto  $R_G(z_2)$  and simply consists in transporting  $K_F$  onto  $K_G$  by  $\alpha \rightarrow K_G^{-1} \circ K_F(\alpha)$ , which actually is on  $\mathbb{R}$  the optimal transport of  $K_F$  to  $K_G$ . The map  $\tau_{FG}$  always preserves  $R$ -curves and also optimally transports them if the supports are equal.

**Proposition 18.** *Let  $c$  satisfy (1). For any  $\alpha \in (0, 1)$  and  $(F, G) \in \mathbb{F} \times \mathbb{F}$  such that  $\mathcal{R}_X = \mathcal{R}_Y$ ,  $\tau_{FG}$   $c$ -optimally transports  $\tilde{K}_{F, z_2}$  onto  $\tilde{K}_{G, z_2}$ .*

*Proof.* As a matter of fact there is only two  $\mathcal{C}^1$  smooth transports and the restriction of  $\tau_{FG}$  is optimal. Observe that if a transport of  $\tilde{K}_F$  on  $R_F(z_2)$  to  $\tilde{K}_G$  on  $R_G(z_2)$  is  $\mathcal{C}^1$  and given by the change of index  $l(\alpha)$  then

$$k_F(\alpha) = g(x_G(l(\alpha), z_2)) \left| \frac{dx_G(l(\alpha), z_2)}{d\alpha} l'(\alpha) \right| = k_G(l(\alpha)) |l'(\alpha)|.$$

Hence  $l'(\alpha) = \pm \frac{k_F(\alpha)}{k_G(l(\alpha))}$ . The choice "+" implies  $l(\alpha) = K_G^{-1} \circ K_F(\alpha)$  since  $l(0) = 0$  and  $l(1) = 1$ . Now, the curves  $R_F(z_2)$  and  $R_G(z_2)$  having the same limiting points  $\lim_{\alpha \rightarrow 1} x_F(\alpha, z_2) = (x_1^+, x_2^+)$  and  $\lim_{\alpha \rightarrow 0} x_F(\alpha, z_2) = (x_1^-, x_2^-)$ , the choice "-" is obviously not better.  $\square$

**Remark 19.** *In other words, the parametrization of the curves  $R_F(z_2)$  and  $R_G(z'_2)$  by  $\alpha$  gives the  $\mathcal{C}^1$  optimal transport maps of  $\tilde{K}_{F, z_2}$  onto  $\tilde{K}_{G, z'_2}$ . This remains true when  $\mathcal{R}_X \neq \mathcal{R}_Y$ .*

At this stage we have defined the  $Q$ -geometry characterizing  $F$  as the system of  $Q$ -curves and  $R$ -curves driven by the new coordinates  $(\alpha, z_2)$  or  $(z_1, z_2)$ . These curves are optimally sent by  $\tau_{FG}$  on the analog  $Q$ -geometry of  $G$  with respect to a family of costs including all the  $p$ -norm costs,  $p > 1$ .

### 3.3 An example

We now provide an example where  $\tau_{FG}$  is the optimal transportation map. It is known ([5]) that if  $F$  and  $G$  have the same copula the optimal transport map is the product of the optimal transport maps coordinate by coordinate.

**Proposition 20.** *If  $(F, G) \in \mathbb{F} \times \mathbb{F}$  that share the same copula then  $\tau_{FG}$  is the unique optimal transport map of  $F$  onto  $G$ .*

*Proof.* Write the common copula  $C_F = C_G$ . The transport map is

$$\tau_{FG}(Q_F(K_F^{-1}(z_1), F_{K_F^{-1}(z_1)}^{-1}(z_2))) = Q_G(K_G^{-1}(z_1), G_{K_G^{-1}(z_1)}^{-1}(z_2)).$$

Write  $\bar{F} = (F_1, F_2)$  and  $\bar{G}^{-1} = (G_1^{-1}, G_2^{-1})$  for convenience. Let show that for any fixed  $z_1$  the two following curves are the same,

$$\begin{aligned} z_2 &\rightarrow Q_G(K_G^{-1}(z_1), G_{K_G^{-1}(z_1)}^{-1}(z_2)), \\ z_2 &\rightarrow \bar{G}^{-1} \circ \bar{F}(Q_F(K_F^{-1}(z_1), F_{K_F^{-1}(z_1)}^{-1}(z_2))). \end{aligned}$$

By Proposition 3 we have  $K_F = K_G$  and, for  $z_2 = 0$  it holds

$$Q_G(K_G^{-1}(z_1), G_{K_G^{-1}(z_1)}^{-1}(0)) = (G_1^{-1}(K_G^{-1}(z_1)), 1) = \bar{G}^{-1} \circ \bar{F}(F_1^{-1}(K_F^{-1}(z_1)), 1)$$

hence they start from the same point. It is then sufficient to verify that they are driven by the same differential equation with respect to  $z_2$ . Put  $\alpha = K_F^{-1}(z_1) = K_G^{-1}(z_1)$ . Let use the shortcut  $g_i(\cdot)$  for the value of the density  $g_i$  at the coordinate  $i$  of  $Q_G(\alpha, G_\alpha^{-1}(z_2))$  and mutatis mutandis for  $f_i$  and  $Q_F$  and denote  $\cdot *$  the coordinate-wise product. Then

$$\begin{aligned} &\frac{dQ_G(\alpha, G_\alpha^{-1}(z_2))}{dz_2} \\ &= \frac{1}{g_\alpha(G_\alpha^{-1}(z_2))} (g_2(\cdot), g_1(\cdot))^T \cdot * \bar{\nabla} C_G(\cdot, \cdot) \\ &= \frac{k_G(\alpha)}{c_G(\bar{G}(\cdot, \cdot)) g_1(\cdot) g_2(\cdot)} (g_2(\cdot), g_1(\cdot))^T \cdot * \bar{\nabla} C_G(\cdot, \cdot) \end{aligned}$$

with  $c_F = c_G$  the copula density, and

$$\begin{aligned} & \frac{d\bar{G}^{-1} \circ \bar{F}(Q_F(\alpha, F_\alpha^{-1}(z_2)))}{dz_2} \\ &= \left( \frac{f_1(\cdot)}{g_1(\cdot)}, \frac{f_2(\cdot)}{g_2(\cdot)} \right)^T \cdot * \frac{1}{f_\alpha(F_\alpha^{-1}(z_2))} (f_2(\cdot), f_1(\cdot))^T \cdot * \bar{\nabla} C_F(\cdot, \cdot) \\ &= \left( \frac{f_1(\cdot)}{g_1(\cdot)}, \frac{f_2(\cdot)}{g_2(\cdot)} \right)^T \cdot * \frac{k_F(\alpha)}{c_F(\bar{F}(\cdot, \cdot)) f_1(\cdot) f_2(\cdot)} (f_2(\cdot), f_1(\cdot))^T \cdot * \bar{\nabla} C_F(\cdot, \cdot). \end{aligned}$$

Inspecting each component of the two formulas reveals the same differential equation. Thus  $\tau_{FG}$  is the optimal transport in this special case.  $\square$

### 3.4 Extension to $\bar{\mathbb{F}}$

First we consider the extension to whole  $\mathbb{R}^2$ . Let  $\bar{\mathcal{F}}$  be the set of the  $\mathcal{C}^2$  distribution functions  $F$  on  $\mathbb{R}^2$  with  $\mathcal{C}^1$  positive density  $f$ . Assume that for non negative continuous functions  $A_1, A_2$  we have

$$c_i(x - y) \leq A_i(x) + A_i(y), \quad i = 1, 2. \quad (19)$$

Consider  $\bar{\mathcal{F}}_1 = \{F \in \bar{\mathcal{F}} : \mathbb{E}(A_1(X_1)) < \infty, \mathbb{E}(A_2(X_2)) < \infty\}$ . We first need to extend Definition 5 to  $F \in \bar{\mathcal{F}}$  then to define  $\tau_{FG}$  for  $F, G \in \bar{\mathcal{F}}_1$  to have  $\mathbb{E}(c(X, \tau_{FG}(X))) < \infty$  by (19). The definitions of the  $Q$ -sets (6),  $Q$ -curves (7) and Kendall distribution (8) remain the same, however the  $Q$ -curves and the  $Q$ -sets are no more bounded. The choice of each  $x_\alpha$  below is left arbitrary since it leads to the same time-mass  $Q$ -geometry of  $F$ .

**Definition 21.** For any  $\alpha \in (0, 1)$ , any  $x_\alpha \in Q_F(\alpha)$  define the  $\alpha$ -th  $Q$ -curve  $Q_F(\alpha, x_\alpha, \cdot)$ , indexed by time  $t \in \mathbb{R}$ , to be the solution of the ordinary differential equation

$$\frac{dQ_F(\alpha, x_\alpha, t)}{dt} = \bar{\nabla} F(Q_F(\alpha, x_\alpha, t)), \quad Q_F(\alpha, x_\alpha, 0) = x_\alpha. \quad (20)$$

We have  $Q_F(\alpha) = \{Q_F(\alpha, x_\alpha, t), t \in \mathbb{R}\}$ . Define  $F_\alpha$  to be the time distribution on  $\mathbb{R}$  with density

$$f_\alpha(t) = \frac{f(Q_F(\alpha, x_\alpha, t))}{k_F(\alpha)},$$

where

$$k_F(\alpha) = \int_{\mathbb{R}} f(Q_F(\alpha, x_\alpha, t)) dt.$$

We have  $k_F(\alpha) < \infty$  and  $k_F(\alpha)$  does not depend on  $x_\alpha \in Q_F(\alpha)$ . Indeed, for any  $x_1 \in Q_F(\alpha)$  there exists a unique  $t_1 \in \mathbb{R}$  depending on  $x_\alpha$  such that  $x_1 = Q_F(\alpha, x_\alpha, t_1)$  and hence  $Q_F(\alpha, x_\alpha, t - t_1) = Q_F(\alpha, x_1, t)$ . The time-mass at  $x$ , is  $t_F(x)$  such that  $x = Q_F(F(x), x_{F(x)}, t_F(x))$ . The local *c.d.f.* along the  $Q_F$ -curves and  $Q_G$ -curves are

$$F_\alpha(t) = \int_{-\infty}^t f_\alpha(s) ds, \quad G_\beta(t) = \int_{-\infty}^t g_\beta(s) ds, \quad \text{for } t \in \mathbb{R},$$

and their inverse functions are defined by, with obvious notation  $y_\beta$ ,

$$\begin{aligned} F_\alpha^{-1}(u) &= Q_F(\alpha, x_\alpha, s) \text{ such that } F_\alpha(s) = u, & \text{for } u \in (0, 1), \\ G_\beta^{-1}(v) &= Q_G(\beta, y_\beta, t) \text{ such that } G_\beta(t) = v, & \text{for } v \in (0, 1). \end{aligned}$$

Finally we write  $x_F(\alpha, u) = F_\alpha^{-1}(u)$  and  $x_G(\beta, v) = G_\beta^{-1}(v)$  the local mass parametrizations on  $(0, 1)$  of the  $Q$ -curves  $Q_F(\alpha)$  and  $Q_G(\beta)$ .

The following proposition is then straightforward.

**Proposition 22.** If  $F \in \bar{\mathcal{F}}$  then  $K_F$  has continuous positive density  $k_F$  on  $(0, 1)$ , the quantile curves  $\alpha \rightarrow Q_F(\alpha)$  uniquely determine  $F$  and the quantile sets  $\alpha \rightarrow \mathbb{Q}_F(\alpha)$  uniquely determine  $F$ .

The definition of our universal generator on  $\mathbb{R}^2$  follows from the fact that the *r.v.*  $x_F(K_F^{-1}(Z_1), Z_2)$  has the distribution  $F$ . When  $F, G \in \bar{\mathcal{F}}$ , we again define  $\mathcal{G}_F(Z) = x_F(K_F^{-1}(Z_1), Z_2)$  and  $\tau_{FG} = \mathcal{G}_G \circ \mathcal{G}_F^{-1}$ . Then the method and results of the previous Sections 2, 3.1, 3.2, 3.3 remain valid with slight modifications. Obviously by changing a few details this extends to the remainder of distributions in  $\bar{\mathbb{F}}$ , those supported by unbounded rectangles of any shape – that is, intersections of parallel or orthogonal half-spaces.

## 4 Statistical aspects of the Kendall ordering

### 4.1 Kendall quantiles and ranks

The quantiles and ranks built from the universal generator map  $\mathcal{G}_F$  are bivariate statistics that provide a stochastic comparison of distributions within  $\overline{\mathbb{F}}$ , through the rank square  $\mathbb{U}$ .

**Definition 23.** For  $z \in \mathbb{U}$ , the  $z$ -th Kendall quantile point of  $F \in \overline{\mathbb{F}}$  is

$$\mathcal{G}_F(z) = x_F(K_F^{-1}(z_1), z_2) = (x_1, x_2) \in Q_F(K_F^{-1}(z_1)).$$

For  $x \in \mathcal{R}_X$ , the  $x$ -th Kendall rank of  $F$  is

$$\mathcal{G}_F^{-1}(x) = (K_F(F(x)), F_{F(x)}(t_F(x))) = (z_1, z_2) \in \mathbb{U}.$$

Next we give an alternative expression for  $\mathcal{G}_F^{-1}(x)$ .

**Proposition 24.** The  $F$ -rank point  $z = \mathcal{G}_F^{-1}(x)$  of  $x \in \mathcal{R}_X$  satisfies, for  $\mathcal{R}_X^x = \{x' \in \mathcal{R}_X : x'_2 > x_2\}$ ,

$$z_2 = \lim_{\varepsilon \rightarrow 0} \frac{F((\mathbb{Q}_F(F(x) + \varepsilon) \setminus \mathbb{Q}_F(F(x))) \cap \mathcal{R}_X^x)}{F(\mathbb{Q}_F(F(x) + \varepsilon) \setminus \mathbb{Q}_F(F(x)))}.$$

Moreover, if  $X$  has distribution  $F$  then  $\mathcal{G}_F^{-1}(X)$  has distribution  $U$ . For  $z \in \mathbb{U}$  the  $z$ -quantile point  $\mathcal{G}_F(z)$  satisfies (4).

*Proof.* From  $F(\mathbb{Q}_F(F(x) + \varepsilon) \setminus \mathbb{Q}_F(F(x))) = K_F(F(x) + \varepsilon) - K_F(x)$  we see by dividing by  $\varepsilon$  that the denominator tends to  $k_F(F(x))$  since  $k_F(\alpha) = \partial K_F(\alpha) / \partial \alpha$  by the proof of Proposition 4. Next observe that Definition 9 also applies to the positive measure with distribution  $F^x(x') := F(x')$ , for  $x' \in \mathcal{R}_X^x$ . Clearly (9) has solution  $Q_{F^x}(F(x), t) = Q_F(F(x), t) \in \mathcal{R}_X^x$  for  $0 < t < t_F(x) = t_{F^x}(x)$  and  $K_{F^x}(F(x)) = F(\mathbb{Q}_F(F(x)) \cap \mathcal{R}_X^x)$  has derivative  $k_{F^x}(F(x)) = \int_0^{t_F(x)} f(Q_F(F(x), t)) dt$ , according to the arguments of the proof of Proposition 4. As a consequence, the numerator is  $K_{F^x}(F(x) + \varepsilon) - K_{F^x}(F(x))$  and the ratio converges to  $k_{F^x}(F(x)) / k_F(F(x)) = F_{F(x)}(t_F(x))$  which establishes the formula for  $z_2$ .

Secondly, if  $X$  has distribution  $F$  then  $\mathcal{G}_F^{-1}(X)$  has distribution  $U$  as  $\mathcal{G}_F$  is a one to one map and by the representation Theorem 8.

Finally, it holds  $\mathbb{P}(\mathcal{G}_F^{-1}(X) \leq z) = \mathbb{P}(K_F(F(X)) \leq z_1) = z_1$  by (12). By applying Theorem 8 with  $X = x_F(K_F^{-1}(Z_1), Z_2)$  we get, for  $\alpha_1 = K_F^{-1}(z_1)$ ,

$$\begin{aligned} & \mathbb{P}(X_2 \geq \langle \mathcal{G}_F(z), e_2 \rangle \mid \mathcal{G}_F^{-1}(X) \in \{z_1\} \times (0, 1)) \\ &= \mathbb{P}(\langle X, e_2 \rangle \geq \langle x_F(K_F^{-1}(z_1), z_2), e_2 \rangle \mid K_F(F(X)) = z_1) \\ &= \mathbb{P}(\langle x_F(K_F^{-1}(Z_1), Z_2), e_2 \rangle \geq \langle x_F(\alpha_1, z_2), e_2 \rangle \mid F(X) = \alpha_1) \\ &= \mathbb{P}(\langle x_F(\alpha_1, Z_2), e_2 \rangle \geq \langle x_F(\alpha_1, z_2), e_2 \rangle \mid Z_1 = z_1) \\ &= \mathbb{P}(Z_2 \leq z_2 \mid Z_1 = z_1) \\ &= z_2. \end{aligned}$$

Therefore (4) holds true.  $\square$

For  $F, G \in \overline{\mathbb{F}}$  we have  $X(\omega) \leq Y(\omega)$  in the Kendall  $Q$ -geometry sense if  $\mathcal{G}_F^{-1}(X(\omega)) \leq \mathcal{G}_G^{-1}(Y(\omega))$ , that is  $\tau_{FG}(X(\omega)) \leq Y(\omega)$ .

### 4.2 One-sided Kendall risk bands

In this subsection and the next one let assume that the coordinates  $(X_1, X_2)$  of  $X$  are meaningful statistical quantitative variables. It may happen in applications that having one at least of these quantities too small characterizes a sub-population at risk. We can then propose the following one sided risk band.

**Definition 25.** The Kendall one sided risk band of level  $z_1 \in (0, 1)$  is the set  $\mathbb{Q}_F(K_F^{-1}(z_1)) = \{x : K_F(F(x)) \leq z_1\}$ .

This is the band before the level curve of the distribution  $F$  of  $X$  at level  $K_F^{-1}(z_1)$ . Considering a similar statistical quantitative variable  $Y$  with distribution  $G$ , we obviously have from Corollary 16 the property that  $\tau_{FG}$  optimally matches the two sub-populations at risk of  $X$  and  $Y$  among the transports maps that preserve the Kendall ordering.

### 4.3 Bivariate Kendall tau

Given  $X = (X_1, X_2)$  a popular measure of rank correlation between  $X_1$  and  $X_2$  is the univariate [7] tau

$$\tau(X_1, X_2) = \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0) - \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) \leq 0)$$

where  $X$  and  $X'$  are independent with distribution  $F$ . If  $F \in \overline{\mathbb{F}}$  this reduces to a product of  $\{1, -1\}$ -valued signs comparison, or equivalently to relative rank comparisons, which is non parametric and robust. Clearly  $\tau(X_1, X_2) = 0$  if  $X_1$  and  $X_2$  are independent, and  $\tau(X_1, X_2) = 1$  (resp.  $-1$ ) if, and only if,  $F$  is degenerated with  $X_2 = F_2^{-1} \circ F_1(X_1)$  (resp.  $X_2 = F_2^{-1} \circ (1 - F_1)(X_1)$ ).

The generator provides a bivariate extension. From  $\mathcal{G}_F^{-1} = (z_1^F, z_2^F)$  define the bivariate signs  $s^F(X, X') \in \{1, -1\}^2$  with coordinates

$$s_j^F(X, X') = 1_{\{z_j^F(X) > z_j^F(X')\}} - 1_{\{z_j^F(X) \leq z_j^F(X')\}}, \quad j = 1, 2.$$

**Definition 26.** Let  $(X, Y) \in \mathbb{R}^2 \times \mathbb{R}^2$  have distribution  $H$  with marginals  $F$  and  $G$ . Given two independent versions  $(X, Y)$  and  $(X', Y')$  with law  $H$ , define the bivariate Kendall correlation to be

$$k(X, Y) = (k_1(X, Y), k_2(X, Y))$$

where, for  $j = 1, 2$ ,

$$k_j(X, Y) = \mathbb{P}(s_j^F(X, X')s_j^G(Y, Y') = 1) - \mathbb{P}(s_j^F(X, X')s_j^G(Y, Y') = -1).$$

If  $X$  and  $Y$  are independent then  $k(X, Y) = (0, 0)$ . The most extreme correlations  $\|k(X, Y)\|_1 = 2$  are achieved in degenerated cases  $Y = \varphi(X)$  that we restrict to  $\varphi \in \tau_1(F, G)$ . We have  $k(X, Y) = (1, 1)$  if, and only if,  $Y = \tau_{FG}(X)$ . Likewise  $(1, -1)$ ,  $(-1, 1)$  and  $(-1, -1)$  are uniquely obtained by  $X = \mathcal{G}_F(Z)$  and, respectively,  $Y = \mathcal{G}_G(Z_1, 1 - Z_2)$ ,  $Y = \mathcal{G}_G(1 - Z_1, Z_2)$  and  $Y = \mathcal{G}_G(1 - Z_1, 1 - Z_2)$ .

As illustrated by the following example,  $k_j(X, Y) = \tau(z_j^F(X), z_j^G(Y))$  provides a different insight on  $H$  than the purely marginal  $\tau(X_j, Y_j)$ . Indeed, all the  $Q$ -geometries of  $F$  and  $G$  are involved in each  $k_j$ , through  $H$ .

**Example 27.** Let  $X_j$  and  $Y_j$  be scores at exam  $j = 1, 2$  – written and oral – in two different lectures. A high  $z_1^F(X)$  characterizes a good student in the first lecture, and a high  $z_2^F(X)$  indicates a student that performs better at written than at oral in that lecture. Then  $k_1(X, Y)$  measures the rank concordance between lectures and  $k_2(X, Y)$  between the evaluation type. Separating these two meaningful effects is indeed not possible by using the pairwise marginal Kendall tau  $\tau(X_1, Y_1)$ ,  $\tau(X_2, Y_2)$ ,  $\tau(X_1, X_2)$  and  $\tau(Y_1, Y_2)$ .

## 5 Empirical transportation maps

In this section we deal with two i.i.d. samples drawn from two smooth distributions  $F$  and  $G$ , with size  $n$  and  $m$  respectively. Write  $F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x}$  the empirical distribution function induced by the sample  $(X_1, \dots, X_n)$  from  $F$ . Let define the  $Q$ -curves,  $Q$ -sets and Kendall distribution with respect to  $F_n$  mutatis mutandis, simply adding the subscript  $n$ : for  $0 < \alpha < 1$ ,

$$\begin{aligned} Q_{F,n}(\alpha) &= \{x \in \mathbb{R}^2 : F_n(x) = \alpha\}, \\ \mathbb{Q}_{F,n}(\alpha) &= \{x \in \mathbb{R}^2 : F_n(x) \leq \alpha\}, \\ K_{F,n}(\alpha) &= F_n(\mathbb{Q}_{F,n}(\alpha)) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{F_n(X_i) \leq \alpha}. \end{aligned}$$

In the sequel  $\alpha = k/n$ ,  $1 \leq k \leq n - 1$ , so that  $Q_{F,n}(\alpha)$  is not empty and  $K_{F,n}(\alpha) = \frac{1}{n} \text{card}\{i : \sum_{j=1}^n \mathbf{1}_{X_j \leq X_i} \leq k\}$ .

### 5.1 The empirical $Q$ -curves algorithm

Assuming  $F \in \mathbb{F}$ , with probability one the first and second coordinates of the sample are all distinct. Thus the empirical  $Q$ -curves  $Q_{F,n}(\alpha)$  are straightforwardly obtained by computing the  $F_n(X_i)$  to identify  $\mathbb{Q}_{F,n}(\alpha)$  then working on the two sets  $H_n(\alpha)$  and  $V_n(\alpha)$  of the first and second coordinates of the  $X_i \in \mathbb{Q}_{F,n}(\alpha)$  – each of these marginal sets being sorted in increasing order, separately. Denote  $F_{i,n}$  the marginal empirical *c.d.f.*, for  $i = 1, 2$ , and  $F_{i,n}^{-1}$  their inverse. Let us start from the "highest" point of  $Q_{F,n}(\alpha)$  defined to be  $(F_{1,n}^{-1}(\alpha), \max(V_n(\alpha)))$ . Next draw an horizontal line up to the point  $x$  having immediately higher first coordinate in  $H_n(\alpha)$  then draw from  $x$  a vertical line up to the

point  $y$  having immediately lower second coordinate in  $V_n(\alpha)$ . Only the upper vertex  $x$  is excluded from the stepwise function  $Q_{F,n}(\alpha)$ . Indeed  $x$  is necessary not an  $X_i$  and  $x \in Q_{F,n}(\alpha + 1/n)$ . Continue the  $nK_{F,n}(\alpha)$  steps " $\rightarrow x \downarrow y$ " until reaching the point  $(\max(H_n(\alpha), F_{2,n}^{-1}(\alpha)), F_{2,n}^{-1}(\alpha))$  as an  $x$  or an  $y$ . It is worth to remark that only vertices  $y$  can be  $X_i$ 's and rather few points of  $Q_{F,n}(\alpha)$  belong to the sample – sometimes none. Moreover, the  $Q_{F,n}(\alpha)$  are equal for  $k/n \leq \alpha < (k+1)/n$ . Figure 1 shows 50 empirical  $Q$ -curves for a  $n = 10^5$  sample of a standard bivariate Gaussian  $\mathcal{N}(0, I_2)$ . For  $\alpha = (2k-1)/100$ ,  $k = 1, \dots, 50$ , the strips between the curves  $Q_{F,n}(\alpha)$  have different empirical probabilities. At the next step they are made equal and equi-subdivided.

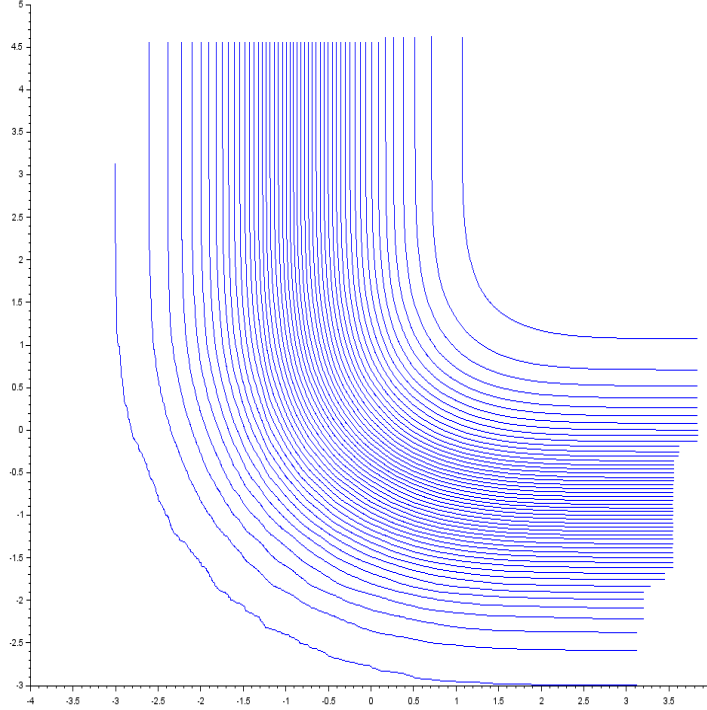


Figure 1: For  $10^5$  sample of  $\mathcal{N}(0, I)$ , 50 empirical  $Q$ -curves  $1/100, 3/100, \dots, 99/100$ .

## 5.2 The empirical Kendall quantiles and ranks algorithm

For  $z \in \mathbb{U}$  we estimate the Kendall quantile point  $\mathcal{G}_F(z) = x_F(K_F^{-1}(z_1), z_2)$  by considering the above empirical  $Q$ -curve  $Q_{F,n}$  of order  $K_{F,n}^{-1}(z_1)$ . It is not so obvious how to define an empirical conditional distribution along the stepwise  $Q_{F,n}(K_{F,n}^{-1}(z_1))$  that is an analog of  $F_\alpha$  along  $Q_F(K_F^{-1}(z_1))$ . To get an easily computed, mathematically tractable, approximation we first select a subset of the  $Q$ -curves orders  $\alpha$  to define a partition of the sample in strips between the selected  $Q_{F,n}(\alpha)$ . Hence we consider  $z_1 = (2i-1)/2p$  for  $i = 1, \dots, p$  and  $z_2 = (2j-1)/2q$  for  $j = 1, \dots, q$ . Taking  $p = q \ll n$  yields a regular grid in  $\mathbb{U}$ . Here the number  $p$  of strips have to be "reasonably" chosen, actually for a further asymptotic study of these empirical curves,  $p$  will depend on  $n$ . For sake of simplicity, assume that  $n/p \in \mathbb{N}$  and  $n/q \in \mathbb{N}$ .

Define the quantiles  $\alpha_1, \dots, \alpha_p$  of  $K_{F,n}$  verifying  $K_{F,n}(\alpha_i) = i/p$ . Recall that they can always be taken of the form  $k/n$ . A non trivial nonparametric geometrical aspect comes from the fact that  $\alpha_i$  are indirect empirical quantiles of  $F(X)$  since  $F(X_i)$  are not observed. The  $Q$ -quantile points we are looking for will be on the "median" curves between  $Q_{F,n}(\alpha_i)$  and  $Q_{F,n}(\alpha_{i+1})$  defined to be  $Q_{F,n}(a_i)$  with  $K_{F,n}(a_i) = (2i+1)/p$ . Let approximate the conditional quantiles  $x_F(K_F^{-1}(z_1), z_2)$  on the curve  $Q_F(\alpha)$  by a random point on  $Q_{F,n}(a_i)$  in the following way. Denote  $S(i)$  the

random strip between  $Q_{F,n}(\alpha_i)$  and  $Q_{F,n}(\alpha_{i+1})$  and consider the (outside) virtual point  $s(i)$  with coordinates the maxima of the first and second coordinates of the sample points in  $S(i)$ . An efficient way to order the points in  $S(i)$  is to sort them in decreasing order of the angle made by the horizontal axis from  $s(i)$  and the lines joining  $s(i)$  to the sample points – see Figure 2. By absolute continuity this ordering is *a.s.* strict. Recalling that all the strips have the same number  $n/p$  of sample points, we approximate the quantiles of the conditionnal distribution on  $Q_{F,n}(\alpha_i)$  by intersecting with the  $z_2$ -th empirical quantile of the angles. If we are looking for  $q$  quantile  $Q$ -curve points, we denote them  $x(i, j)$  for  $z_2 = (2j - 1)/q, j = 1, \dots, q$  as on Figure 2. Again for an asymptotic study  $q$  should depend on  $n$ .

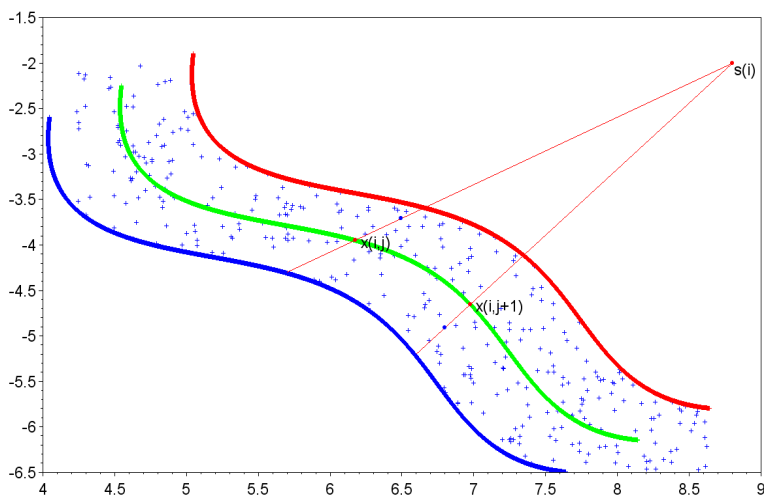


Figure 2: Sample in the strip  $S(i)$  between (blue)  $Q_{F,n}(\alpha_i)$  and (red)  $Q_{F,n}(\alpha_{i+1})$ , the two (red) points  $x(i, j)$ ,  $x(i, j + 1)$  on the (green) median curve  $Q_{F,n}(\alpha_i)$  correspond to quantiles of the angles with  $s(i)$ , ordered top-down, reached by the two (blue) points on the (red) lines.

Figure 3 shows the previous 50 empirical level curves and 50 empirical quantiles curves for the same sample of  $\mathcal{N}(0, I_2)$ . One can notice that these  $R$ -curves of conditionnal quantiles are much less stable than the level  $Q$ -curves. This is due to the relatively low number of sample points in each strip and the wellknown difficulty to estimate quantiles even on the real line.

### 5.3 Transportation maps

To obtain an approximation of the transport map  $\tau_{FG}$  of (5) we compute by the previous algorithm the level curves and conditional quantile curves on two samples of  $F$  and  $G$  with the same  $p$  and  $q$  – the sample sizes  $n$  and  $m$  need not to be equal. If  $x(i, j)$  and  $y(i, j)$  are the respective conditional quantiles, that we hereafter call the grids, then the approximation  $\tau_{n,m}$  of  $\tau_{FG}$  on the grids is

$$\tau_{n,m}(x(i, j)) = y(i, j). \quad (21)$$

Thus we have built an empirical skeleton of the theoretical transportation map by matching these two random grids. One can extend this map to any point  $x$  of the convex hull of the  $F$  sample by simple interpolation.

Alternatively we can use the above algorithms starting from  $\alpha = F_n(x)$  then compute the empirical quantile of the angle reaching  $x$  – interpolated in the strip – and procede in the same way along the  $K_{G,n}$  curve of index  $K_{G,n}^{-1}(K_{F,n}(\alpha))$  with the same – interpolated in the strip – quantile angle. Notice that neither  $x$  nor  $\tau_{n,m}(x)$  are sample points.

## 6 Some numerical examples

We only consider the quadratic cost. The optimal transportation cost is the Wasserstein square distance  $W_2^2$  between probability measures. We restrict ourselves to a few numerical study on  $\overline{\mathbb{F}}$ , with  $n = m = 10^5$ .

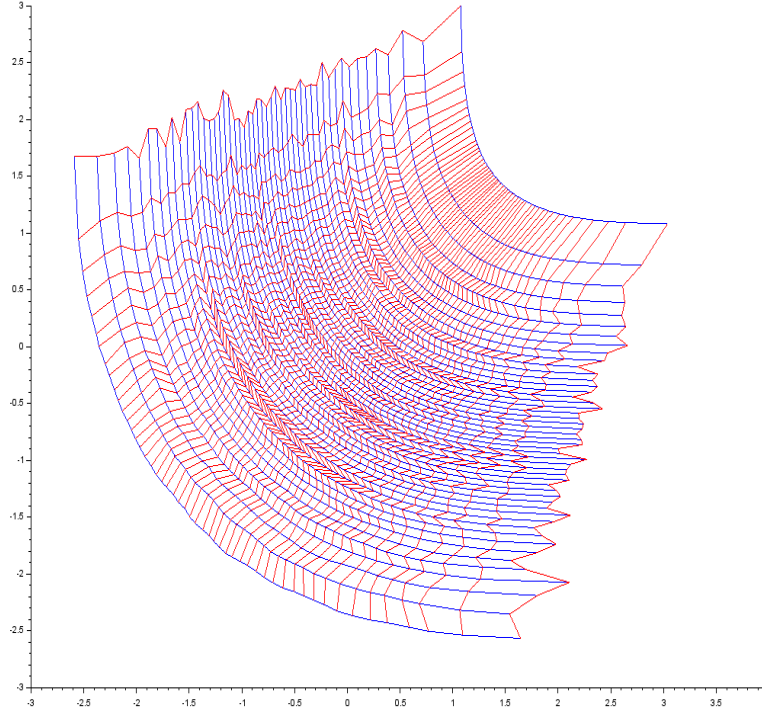


Figure 3: 50 level  $Q$ -curves (blue) and quantile  $R$ -curves (red) of a  $10^5$  sample of  $\mathcal{N}(0, I_2)$ .

## 6.1 Cartesian coordinates

Actually just a few optimal transportation maps are explicitly known. Let begin with some of these usual examples.

First the trivial case  $F = G$  allows to "quantify" the numerical error due to our approximating method. The theoretical  $\tau_{FF}$  is the identity map, and we compute the mean square distance error with the empirical map on the grids. In the previous Gaussian case  $\mathcal{N}(0, I_2)$  with two samples of size  $10^5$  and grids  $50 \times 50$  the mean square error is .002 – see Figure 4.

The next example illustrates the case where the two distributions share the same copula. By Proposition 20 the optimal transport is the product of the optimal marginal transports. In our numerical example we simulate a  $10^5$  sample of

a  $\mathcal{N}(0, \Sigma)$ ,  $\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$  and a  $10^5$  sample of a  $r.v.$  with the same Gaussian copula and marginals having respective

densities  $a|x|e^{-ax^2}$  – using  $a = .05$  – and  $\frac{1}{8}|x|e^{-x/2}$  – a symetrized  $\chi^2(4)$ . As expected our empirical map is close to the optimal one that we estimate on the grid  $x(i, j)$  by the product of empirical marginal quantiles. Figure 5 shows the transport segments from  $x(i, j)$  to the red points  $y(i, j)$ . Figure 6 draws the segments from  $y(i, j)$  to the image of  $x(i, j)$  by the product transport. The mean quadratic error between the two maps is .027, what is negligible before the true cost 17.94 of the optimal product transportation map on the starting grid.

Remind that  $\tau_{FG}$  and  $\tau_{n,m}$  depend on the choice of the orthonormal basis  $b_0$ . In order to look at the effect of this choice we simulate two  $10^5$  samples of centered mixtures of Gaussian distributions – with 4 and 3 components respectively – for which both the optimal transportation map and the Wasserstein distance  $W_2^2$  are unknown. Figures 7 and 8 show subsamples. Then we consider the rotations of  $b_0$  with angle  $2k\pi/360$ ,  $0 \leq k \leq 359$ , and compute  $\tau_{n,n}$  on the  $50 \times 50$  grids. We plot the 360 values of the empirical cost at Figure 9. The variation of the transport cost is quite large. We draw the best transportation map at Figure 10.



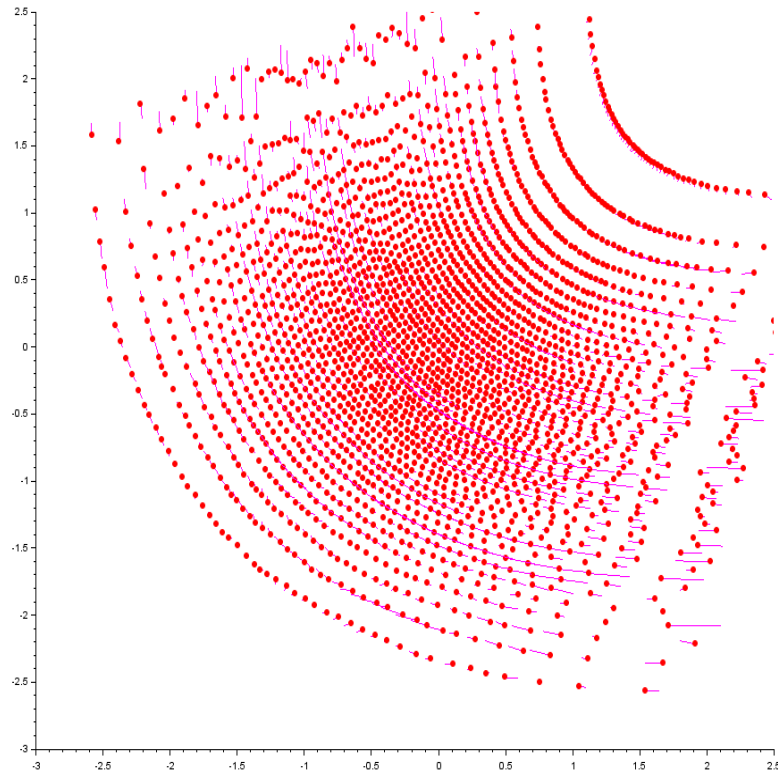


Figure 4: Transportation map with grids  $50 \times 50$ , two  $10^5$  samples of  $\mathcal{N}(0, I_2)$ : (red) points are the images of the starting grid and the (cyan) lines show the move from initial points.

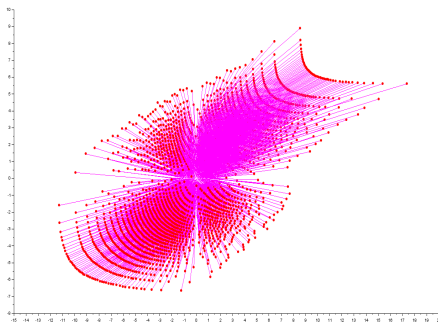


Figure 5: Example with the same copula: transportation map  $x \rightarrow \tau_{n,n}(x)$  on the  $Q$ -quantile grid  $50 \times 50$ ,  $n = m = 10^5$ .

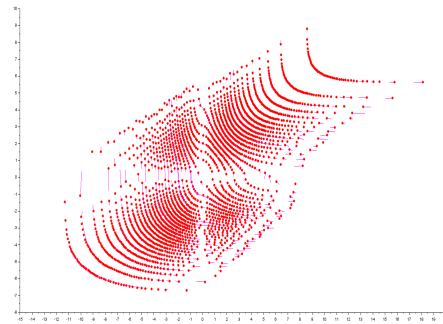


Figure 6: Small errors between images of the initial grid of Figure 5 by  $\tau_{n,n}$  – the (red) points – and the product empirical map.

To end this subsection we go back to the same copula case. We keep the previous sample of the 4 Gaussians mixture and draw a second sample of this distribution then transform its marginals onto the marginals of the 3 Gaussians mixtures.

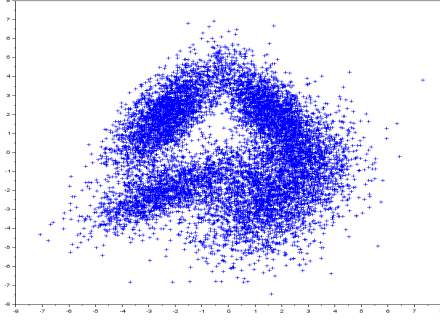


Figure 7: Mixture of four Gaussian distributions, sub-sample  $10^4$ .

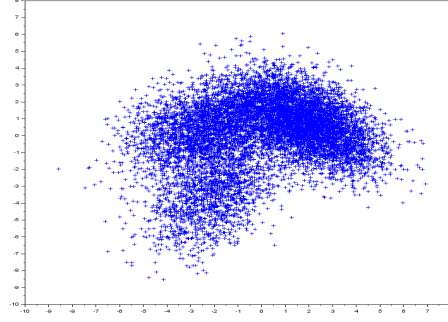


Figure 8: Mixture of three Gaussian distributions, sub-sample  $10^4$ .

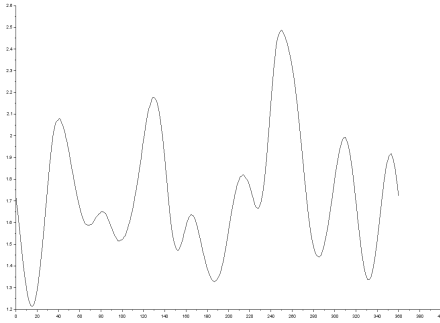


Figure 9: Values of the  $W_2$  cost versus angle of rotation for the Gaussian mixtures of Figures 7 and 8.

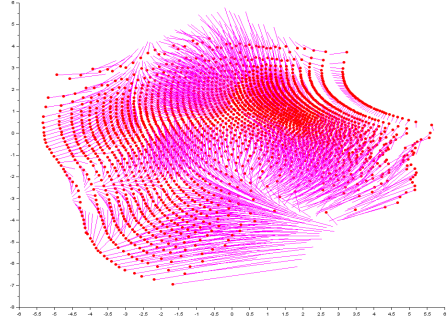


Figure 10: The best transportation map for Gaussian mixtures, with a  $W_2$  cost 1.214. It shows the relocation and continuous splitting of the four components.

We observe that  $\tau_{n,n}$  is close to the product of the marginal transportation maps up to a mean square error less than .0065, to be compared to the empirical optimal quadratic cost .545.

## 6.2 Polar coordinates

As a matter of fact we may use other systems of coordinates, for instance the polar coordinates  $(\rho, \theta)$ . We still assume the  $r.v$  centered. Let consider two radial distributions with the same angle distribution, moreover independent of the radius distribution. In this case the optimal transport consists only in optimally transporting the radius. Observe that the quadratic cost in cartesian coordinates differs from the quadratic cost in polar coordinates. However, as the angles share the same distribution, the two optimal transport maps coincide.

We simulate radius with respective densities  $2a\rho e^{-a\rho^2}$  – with  $a = .05$  – and  $\frac{1}{4}\rho e^{-\rho/2}$  – a  $\chi^2(4)$  – and independent angles sharing the density  $\frac{1}{4}\sin \theta/2$ . Figures 11, 12 show the obtained samples. Figure 13 illustrates that  $\tau_{n,n}$  looks radial, that seems confirmed by Figure 14 showing that the images by optimal and the empirical transports of three thin centered ring almost coincide. The empirical transport cost is .68 and the optimal one on the grid is .646 and the actual optimal cost is .743. Note that estimating the optimal cost with marginal and 50 exact quantiles is 0.643.

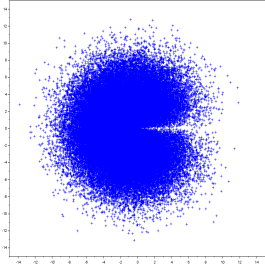


Figure 11: "Pseudo-radial" example, first sample.

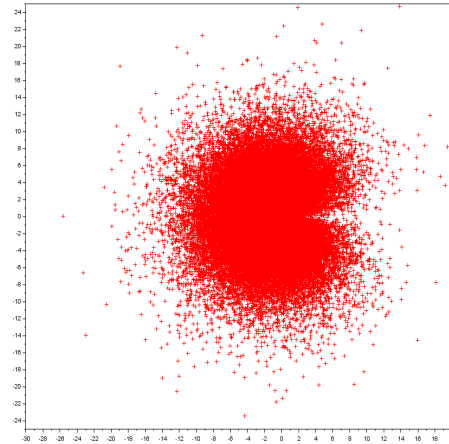


Figure 12: "Pseudo-radial" example, second sample, same scale.

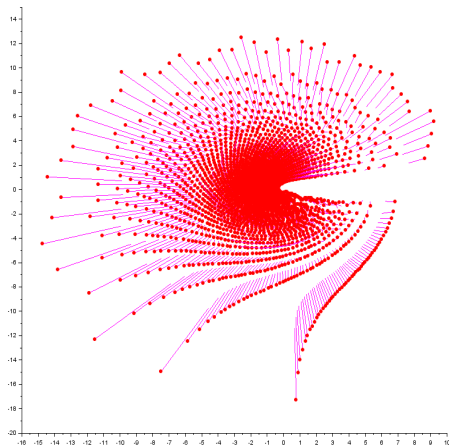


Figure 13: "Radial" example, transportation map  $\tau_{nn}$  on an initial Kendall quantile grid  $50 \times 50$ , samples sizes  $10^5$ .

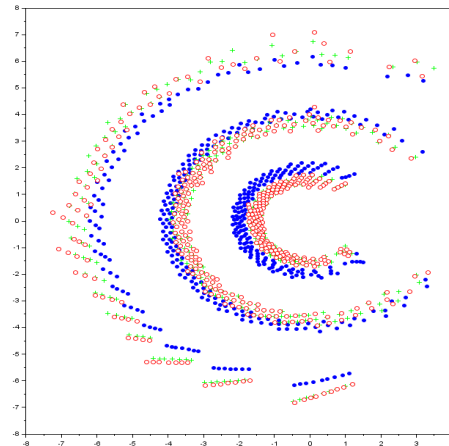


Figure 14: (blue)  $\bullet$  are sent to (red)  $\circ$  by  $\tau_{n,n}$  map and to (green)  $+$  by optimal map.

## 7 Conclusion and further statistical applications

### 7.1 Perspectives

In this article we build a family (5) of theoretical transportation maps and their numerical counterpart (21) through empirical transportation maps.

On a practical side these maps are easily built from any pair of samples of two unknown probability distributions. The computation cost is low for samples of size up to  $10^5$ . In simple cases these maps give an actual approximation of the optimal transport. By construction they are basis dependent, but an inspection of all the orthonormal basis of the plane allows to select a  $\tau_{n,m}$  performing better.

From a mathematical statistics viewpoint  $\tau_{FG}$  are a new tool to compare or transport probability distributions. Various tests can be built using pseudo Wasserstein distance costs. More importantly, we have designed (5) and (21) in a non parametric way exclusively based on the *d.f.* and empirical *d.f.* respectively. Therefore our next achievable goal is to derive *C.L.T.* type theorems for the empirical transport cost and – more difficult – on the approximation of  $\tau_{FG}$  by  $\tau_{n,m}$  on  $\mathcal{R}_X$ .

Another direction to be worked out is to overcome the question of  $b_0$  by defining a transport similar to  $\tau_{FG}$  but derived from – and instead of  $F$  – a real valued function whose level curves characterize the distribution and are independent of the basis chosen for the computations.

From a statistical application viewpoint  $\tau_{FG}$  can be used in various ways. First, beyond the bivariate Kendall tau of Definition 26, one can build goodness-of-fit tests based on large samples with a reasonable computing time, see (22) below. When the marginals of the two distributions and their ordering make sense, working in this particular basis  $b_0$  is meaningful since the transportation map preserve the marginals order. Second, in the next subsections we show that once  $\mathcal{G}_F$  is built, entirely new statistical objects describing  $F$  come out as free byproducts. They are based on the easily estimated Kendall ordering of the universal generator, hence on level sets of  $F$  instead of level sets of  $f$  as in classical approaches.

## 7.2 Kendall rank distance and rank paths

Any distance  $d$  on the rank square  $\mathbb{U}$  induces a rank distance  $d(\mathcal{G}_F(x), \mathcal{G}_G(y))$  between  $x, y \in \mathcal{R}_X$  inside the distribution  $F$ . Likewise one can define  $d(\mathcal{G}_F(x), \mathcal{G}_G(y))$  if  $F, G \in \overline{\mathbb{F}}$ . In order to remove the dependency on the orientation  $b_0$ , let simply use the rotations  $r_\theta$  by angle  $\theta \in [0, 2\pi)$ . Write  $F_\theta = F \circ r_{-\theta}$  and  $G_\theta = G \circ r_{-\theta}$ .

**Definition 28.** For  $F, G$  in  $\overline{\mathbb{F}}$  let the minimal (resp. maximal) rank distance between  $x \in \mathcal{R}_X$  and  $y \in \mathcal{R}_Y$  be  $r_{FG}^-(x, y) = \min_{\theta \in [0, 2\pi)} d(\mathcal{G}_{F_\theta}(x), \mathcal{G}_{G_\theta}(y))$  (resp.  $r_{FG}^+(x, y) = \max_{\theta \in [0, 2\pi)} d(\mathcal{G}_{F_\theta}(x), \mathcal{G}_{G_\theta}(y))$ ) and the mean rank distance be  $\bar{r}_{FG}(x, y) = \int_0^{2\pi} d(\mathcal{G}_{F_\theta}(x), \mathcal{G}_{G_\theta}(y)) d\theta$ .

Hence  $\bar{r}_{FG}(x, x)$  quantifies the different positioning of  $x$  in the planar distributions  $F$  and  $G$ . Assume that  $F = G$  and  $d$  is the  $L_2$  distance, in order to treat equally the  $Q$  and  $R$  curves. The rank range  $[r_{FF}^-(x, y), r_{FF}^+(x, y)]$  characterizes the proximity of  $x$  and  $y$  inside  $F$  and a rank path can be used to join  $x$  to  $y$ ,

$$t \in [0, 1] \rightarrow \mathcal{G}_{F_\theta}^{-1}(t\mathcal{G}_{F_\theta}(x) + (1-t)\mathcal{G}_{F_\theta}(y)).$$

If  $x$  and  $y$  share the same  $R$  or  $Q$  curve, the rank path is a segment along that curve. A rank array – several or all  $\theta$  – may characterize the stability around the mean rank path – averaging in  $\theta$ .

## 7.3 Generator depth, local depth and contours

A depth value aims to quantify whether a point  $x$  is close to the main mass concentrations or not. Let define the  $\mathcal{G}_F$ -depth relatively to a central point  $y \in \mathcal{R}_X$ , typically the median point  $\mathcal{G}_F^{-1}(1/2, 1/2)$  or a symmetry point, to be  $1/d(\mathcal{G}_F(x), \mathcal{G}_F(y))$ . This puts uniformly bounded small values at boundaries and infinite value at  $y$  – take the inverse ratio to revert the scoring. This definition easily extends to a depth value relatively to a set  $\mathcal{Y}$ . As generally  $F$  is viewed as a measure in the euclidean plane we give a basis free definition.

**Definition 29.** Let  $F \in \overline{\mathbb{F}}$ . The  $\mathcal{G}$ -depth relative to  $y \in \mathcal{R}_X$  is

$$x \in \mathcal{R}_X \rightarrow D_F(x|y) = \int_0^{2\pi} \frac{1}{d(\mathcal{G}_{F_\theta}(x), \mathcal{G}_{F_\theta}(y))} d\theta.$$

The  $\mathcal{G}$ -local depth relative to  $\mathcal{Y} \subset \mathcal{R}_X$  is

$$x \in \mathcal{R}_X \rightarrow D_F(x|\mathcal{Y}) = \inf_{y \in \mathcal{Y}} D_F(x|y).$$

If  $\text{card}(\mathcal{Y}) < \infty$  the  $\mathcal{G}$ -attractor function of  $\mathcal{Y}$  is

$$x \in \mathcal{R}_X \rightarrow A_F(x|\mathcal{Y}) = \underset{y \in \mathcal{Y}}{\text{Argmin}} D_F(x|y).$$

In a descriptive statistics perspective, like clustering or mode estimation,  $\mathcal{Y}$  are mass concentration centroids, typically modes of a multimodal density  $f$  or sub-population reference points in a mixture of sub-populations.

The  $\mathcal{G}$ -depth sets  $\{x : D_F(x|y) \geq \gamma\}$  and  $\mathcal{D}_\gamma(\mathcal{Y}) = \{x : D_F(x|\mathcal{Y}) \geq \gamma\}$  are strictly increasing for inclusion as  $\gamma$  increases. We can deduce from Definition 29 confident regions around specific points  $\mathcal{Y}$  to be central in the region.

The  $\mathcal{G}$ -contours  $\mathcal{C}_\gamma(\mathcal{Y}) = \{x : D_F(x|\mathcal{Y}) = \gamma\}$  are closed, continuous, nested curves describing the mass distribution in a similar way as density level sets. They fit what is expected from probability contours – see Figure 15

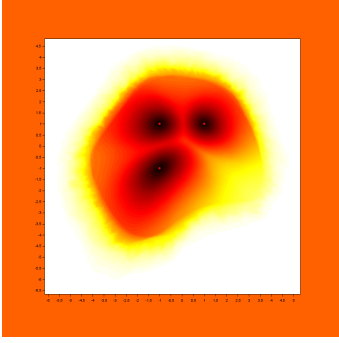


Figure 15: Depth contours obtained for the Gaussian mixture of Figure 8 by using 128 angles and Definition 29

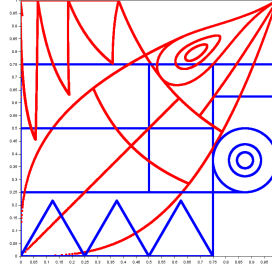


Figure 16: The curved (red) partition is the  $\mathcal{G}_U^{-1}$  image of the line-circle (blue) partition of ranks in the Kendall  $Q$ -geometry of the uniform distribution  $U$  as computed at Example 10. The surface of ceils is preserved.

#### 7.4 Generator minimal and mean $W_2$ transportation costs

Assuming that  $c$  is the quadratic cost, the choice of the orthogonal basis is unimportant. Observe that the *r.v.*  $Y_\theta = r_\theta^{-1}(\tau_{F_\theta G_\theta}(r_\theta(X)))$  has distribution  $G$ . In a goodness-of-fit test context, a robust decision could then be based on

$$\min_{\theta \in [0, 2\pi)} \int_{\mathbb{U}} \|\mathcal{G}_{F_\theta}^{-1}(z) - \mathcal{G}_{G_\theta}^{-1}(z)\|_2^2 dz = \min_{\theta \in [0, 2\pi)} \mathbb{E}(\|X - Y_\theta\|_2^2). \quad (22)$$

As a matter of fact, changing drastically a few sample points won't change the estimators of the Kendall  $Q$ -geometry at Section 5. Incidentally (22) is an explicit upper bound for the optimal transport cost  $W_2(F, G)$ .

#### 7.5 Kendall partitions

Another nice feature of  $\mathcal{G}_F$  is to provide tessellations of  $\mathcal{R}_X$  in ceils with desired  $F$ -probability. Start from any partition  $A_i$  of the unit square  $\mathbb{U}$  – for instance in squares, triangles or hexagons with same surface  $|A_i|$  or not, or a mixed configuration. The quantile ceils  $\mathcal{G}_F^{-1}(A_i)$  have probabilities  $|A_i|$  and sometimes surprising characteristic shapes adapted to the Kendall  $Q$ -geometry characterizing  $F$  – see Figure 16. The empirical version could be used to build balanced random forests.

## References

- [1] L. Ambrosio, N. Gigli, and G. Savaré. Gradient flows in metric spaces and in the space of probability measures. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.
- [2] S. Balakrishnan, T. Manole, J. Niles-Weed, and L. Wasserman. Plugin estimation of smooth optimal transport maps. <https://doi.org/10.48550/arxiv.2107.12364>, 2021.
- [3] B. Bercu, J. Bigot, S. Gadat, and E. Siviero. A stochastic gauss-newton algorithm for regularized semi-discrete optimal transport. Information and Inference: A Journal of the IMA, to appear.
- [4] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. Communications on pure and applied mathematics, 44:375–417, 1991.
- [5] J.A. Cuesta Albertos, L. Ruschendorf, and A. Tuero-Diaz. On the regularity of solutions of optimal transportation problems. J. Multivariate Anal., 46:335–361, 1993.
- [6] G. De Philippis and A. Figalli. The monge-ampère equation and its link to optimal transportation. Bull. Amer. Math. Soc., 51:527–580, 2014.
- [7] M.G. Kendall. A new measure of rank correlation. Biometrika, page 81–93, 1938.
- [8] G. Peyré and M. Cuturi. Computational optimal transport. <https://doi.org/10.48550/arXiv.1803.00567>, 2018.

- [9] A.-A. Pooladian and J. Niles-Weed. Entropic estimation of optimal transport maps. <https://doi.org/10.48550/arxiv.2109.12004>, 2021.
- [10] S.T. Rachev and L. Rüschendorf. Mass transportation problems. Vol. I: Theory. Probability and its Applications (New York). Springer-Verlag, New York, 1998.
- [11] C. Villani. Optimal transport, Old and new, volume 338 of Fundamental Principles of Mathematical Sciences. Springer-Verlag, Berlin, 2009.